ОБРОБЛЕННЯ СИГНАЛІВ І ЗОБРАЖЕНЬ ТА РОЗПІЗНАВАННЯ ОБРАЗІВ

Signal/Image Processing and Pattern Recognition

ПРАЦІ ✵ PROCEEDINGS

ШОСТА ВСЕУКРАЇНСЬКА МІЖНАРОДНА КОНФЕРЕНЦІЯ
8 - 12 жовтня 2002 року ✵ Київ, Україна

THE SIXTH ALL-UKRAINIAN INTERNATIONAL CONFERENCE
October 8 - 12, 2002 ✵ Kyjiv, Ukraine

Українська асоціація з оброблення інформації та розпізнавання образів
Ukrainian Association for Information Processing and Pattern Recognition
Національна академія наук України та Міністерство освіти і науки України -
Міжнародний науково-навчальний центр інформаційних технологій та систем
National Academy of Sciences of Ukraine & Education and Science Ministry of Ukraine -
International Research and Training Center for Information Technologies and Systems
Національна академія наук України - Інститут кібернетики
National Academy of Sciences of Ukraine - Institute of Cybernetics
Академія інженерних наук України - Відділення інформатизації та автоматизації
Academy of Engineering Sciences - Department of Informatization and Automatization
Українська інженерна спілка - Секція інформатики
Ukrainian Engineering Association - Section of Informatics

ОБРОБЛЕННЯ СИГНАЛІВ І ЗОБРАЖЕНЬ ТА РОЗПІЗНАВАННЯ ОБРАЗІВ

Signal/Image Processing and Pattern Recognition

ПРАЦІ ✵ PROCEEDINGS

ШОСТА ВСЕУКРАЇНСЬКА МІЖНАРОДНА КОНФЕРЕНЦІЯ
8 —12 жовтня 2002 року, Київ, Україна

THE SIXTH ALL-UKRAINIAN INTERNATIONAL CONFERENCE
October 8—12, 2002 Kyjiv, Ukraine

УкрОБРАЗ'2002 ✵ UkrOBRAZ'2002

Київ 2002

Статті публікуються без змін, зі збереженням авторського оригіналу

The papers are published without any change in the text that have been presented by authors

Веб-сторінка конференції: http://www.uasoiro.org.ua/ukrobraz2002/

The conference homepage: http://www.uasoiro.org.ua/ukrobraz2002/

ISBN 966-02-2652-7

Видання Української асоціації з оброблення інформації та розпізнавання образів:
40, проспект Академіка Глушкова, Київ 03680, Україна

Наклад 200 примірників

ЗМІСТ
CONTENTS

I. *Проблеми створення образного комп'ютера*
Problems of Pattern Computer Designing

II. *Теоретичні питання оброблення та розпізнавання сигналів і зображень*
Theoretical Problems of Signal Processing and Recognition

III. Усномовні інформаційні технології
Speech Information Technologies

IV. Технічна діагностика об'єктів і машин за їх сигналами та полями
Automatic Diagnostic of Objects and Machines on Its Signals and Fields

V. Автоматичне оброблення та розпізнавання зображень
Automatic Image Processing and Recognition

VI. Цифрова фільтрація, кодування та відновлення сигналів і зображень
Digital Filtering, Analysis, Coding and Reconstruction of Signals and Images

VII. Нові інформаційні технології та системи, що ґрунтуються на обробленні сигналів і зображень
Advanced Information Technologies Based on Signal/Image Processing and Pattern Recognition

ГЕНЕРАТИВНА МОДЕЛЬ ОБРАЗНОГО КОМП'ЮТЕРА[1]

Тарас Вінцюк

Міжнародний науково-навчальний центр ЮНЕСКО інформаційних технологій та систем
40 проспект Академіка Глушкова, Київ 03680, Україна
Тел.: +380 44 266-4356 Факс: +380 44 266-4356
vintsiuk@uasoiro.org.ua

ABSTRACT

Taras K. Vintsiuk. Generative Model for Pattern Computer. Pattern computer (PatCom) is such intellectual cybernetic system that provides a functional simulation of intelligent, mainly subconscious activity of all living and a human being particularly. In PatCom this activity is related to image, sound and other patterns perception, scene analysis, action and movement planning, generalisation of observations, discovering of regularities, prediction, decision making etc. Pattern computer operates with patterns and other complex notions. It actualises both pattern and logical reasoning.

Pattern computer is a parallel system. It has several information perception channels (acoustic, visual, scential) that is multimodal perception, pattern operation system, improved human-machine interface. Unlike usual computer, which is based on a rapid arithmetic-logical processor and a large RAM space, PatCom is grounded on external world models including physical, geometry, acustical, language, linguistic, semantical, canonical forms etc models.

Here a so-called generative model for how to create a pattern computer is debated. This model is based on both the generative grammar hierarchy for multimodal prototype scene composition and the comparison of them with a scene to be percepted. So a multimodal information synthesis is used as a feedback in the pattern analysis and understanding. Examples are given in reference to dictation and spoken translation machine creating and scene analysis.

ВСТУП

Людство потребує створення принципово нових комп'ютерів, які здатні сприймати та розуміти звуки, зображення, людську мову, рукописні тексти, креслення, просторові та звукові сцени, інші образи, описувати та озвучувати зображення, перекладати з однієї мови на іншу тощо.

Розроблення таких комп'ютерів, які виконують не тільки обчислення, але й моделюють образне сприйняття світу та образне прийняття рішень відносять до проривних напрямів у науково-технологічному поступі.

Кабінет Міністрів України своєю Постановою від 8.11.2000 № 1652 схвалив Державну науково-технічну програму "Образний комп'ютер".

Програма має на меті створення принципово нових інформаційних технологій та систем — образних комп'ютерів.

ЩО ТАКЕ ОБРАЗНИЙ КОМП'ЮТЕР

Образний комп'ютер — це така кібернетична система, в якій виконується функційне моделювання інтелектуальної, головно підсвідомої, діяльності людини та всього живого, що пов'язана зі сприйняттям зорових, слухових та інших образів, аналізом сцен та складних ситуацій, плануванням дій та рухів, узагальненням спостережень, встановленням закономірностей, прогнозуванням, прийняттям рішень тощо. ОК оперує образами та іншими складними поняттями, реалізує як образне, так і логічне мислення.

Образний комп'ютер є паралельною мультимодальною системою, яка має у своєму складі декілька каналів сприйняття інформації (слухової, зорової, текстової, смакової, нюхової тощо), образну операційну систему, моделі зовнішнього світу (в тому числі акустичну, оптичну, геометричну, лінгвістичну, семантичну, канонічних форм тощо), розвинений інтерфейс з людиною, засоби взаємодії з існуючими комп'ютерними та телекомунікаційними мережами.

Образна операційна система "синхронізує" оброблення інформації, що надходить різними каналами її сприйняття, та, оперуючи моделями зовнішнього світу, виконує комплексну семантичну інтерпретацію всієї отриманої інформації.

Отже, змістовно визначені два базових поняття: образний комп'ютер (ОК) та образна операційна система (ООС).

Підкреслимо, що на відміну від звичайного комп'ютера у центрі образного комп'ютера знаходяться моделі зовнішнього світу з усіма його об'єктами, явищами та їх проявами.

СТРУКТУРА ОБРАЗНОГО КОМП'ЮТЕРА

На рис. 1 подано схематичне зображення ОК. Наведені чотири канали сприйняття інформації: слуховий, зоровий, нюховий та тактильний. В кожному каналі виділено персептор, аналізатор, розпізнавач та інтерпретатор.

Пунктирними лініями зображається ієрархія рівнів оброблення інформації. Найнижчий рівень

[1] **Робота виконана в рамках Контракту № ОК_2002_1_МІПІ_ЦЕНТР за ДНТП "Образний комп'ютер"**

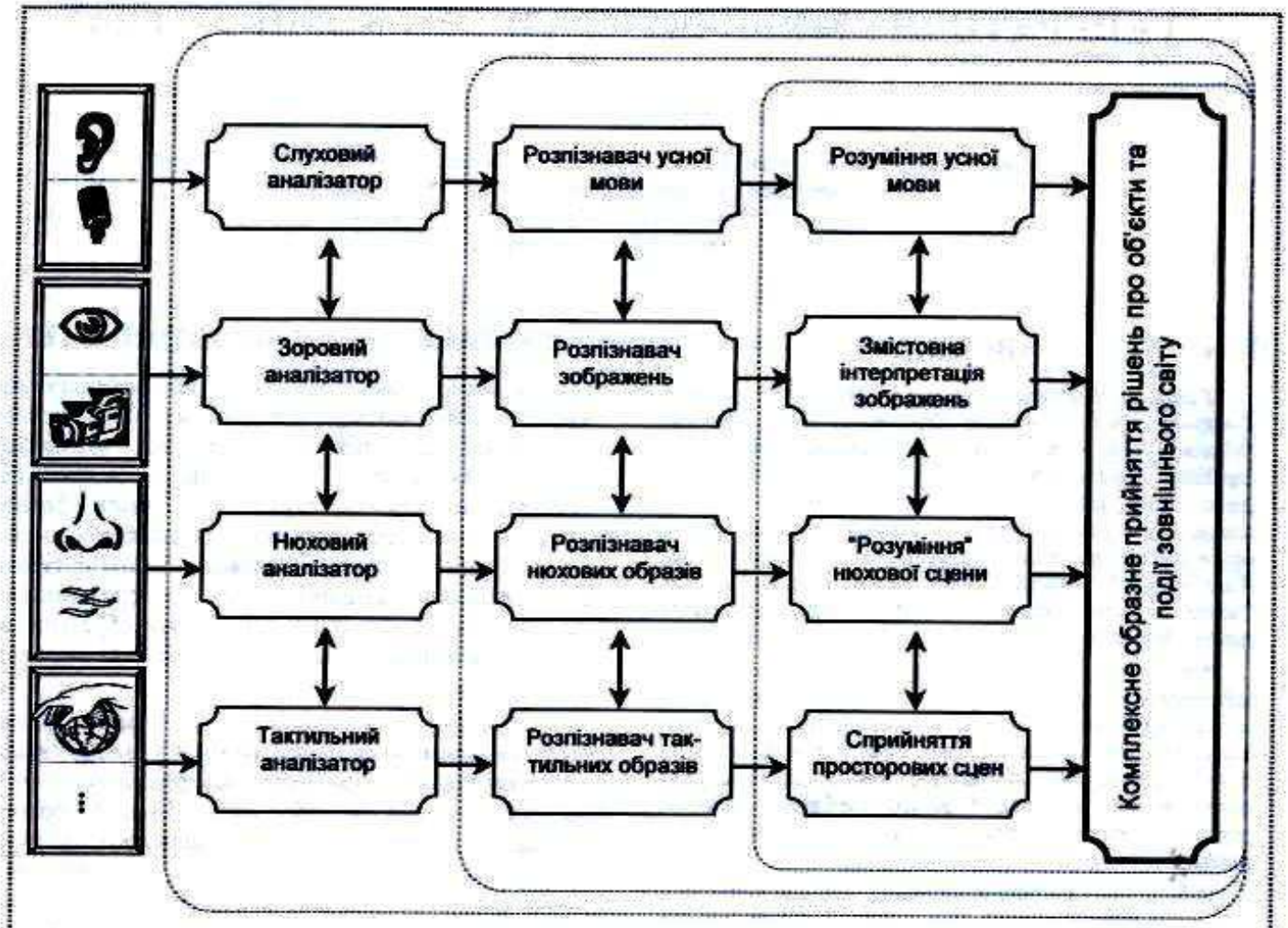


Рис. 1. Схематичне зображення структури образного комп'ютера. Зворотні зв'язки не показані.

стосується аналізу, найвищий — змістовної інтерпретації та розуміння.

Образна операційна система виконує комплексну, узгоджену за всіма каналами, інтерпретацію та образне прийняття рішень про об'єкти та події зовнішнього світу. Ієрархії в ООС відображено вкладеними пунктирними фігурами.

Взаємодії між каналами показані вертикальними стрілками. Зворотні ж зв'язки в каналах та між каналами не відмічені.

КОНЦЕПЦІЯ ГЕНЕРАТИВНОЇ МОДЕЛІ

Генеративна модель образного комп'ютера ґрунтується на конструктивній реалізації загальновизнаних принципів оброблення інформації в живій природі, техніці та суспільстві, таких як: аналіз через синтез, індуктивне та дедуктивне виведення, генерація та направлений перебір варіантів, зворотній зв'язок, використання апріорної інформації, адаптація, навчання та самонавчання, розпізнавання образів, побудова моделей зовнішнього світу та розумової діяльності.

Відомо, що кожне вимовляння одного й того ж слова чи написання однієї й тієї ж літери навіть однією й тією ж людиною, скажімо, з інтервалом в одну секунду часу, завжди передаються різними ("двічі в одну й ту ж воду ввійти неможливо"), але чимось схожими сигналами або зображеннями.

Отож, універсальним алгоритмом розпізнавання образів міг би бути такий. Спочатку запам'ятати всі можливі сигнали чи зображення, які передають один і той самий образ (клас, слово, букву тощо), а потім при автоматичному розпізнаванні порівнювати пред'явлений для оброблення сигнал чи зображення з усіма раніше запам'ятованими сигналами чи зображеннями й віднести розпізнаваний сигнал (зображення) до того образу (класу), з чим раніше запам'ятованим сигналом (зображенням) він збігся. На жаль, такий алгоритм розпізнавання хоч і є "сильним", але він не є конструктивним: не знайдеться такий звичайний комп'ютер, ні тепер, ні в майбутньому, який був би в змозі запам'ятати всі можливі, розмаїті сигнали (зображення) та виконувати в реальному часі необхідні порівняння.

Труднощі, що виникають при реалізації цього універсального алгоритму розпізнавання в

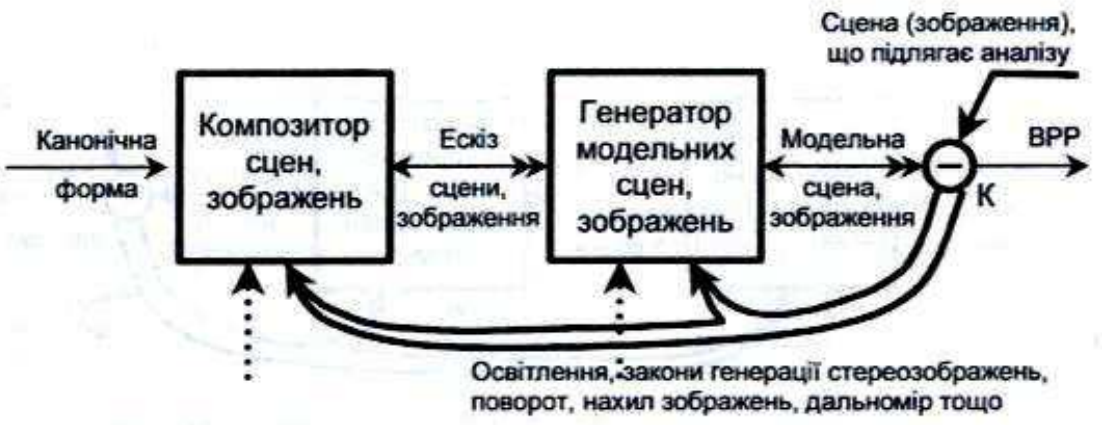


Рис. 2. Генеративна модель змістовної інтерпретації сцен.

генеративній моделі можуть долатися таким чином: 1) треба навчитись запам'ятовувати, структуризувати всі можливі сигнали (зображення) якимсь економним способом (адже вони пов'язані сильними залежностями, бо є схожими!); 2) а порівняння сигналів (зображень) треба виконувати направленим перебором варіантів.

СПЕЦИФІКАЦІЯ ОБРАЗІВ

Відомо, що за допомогою диференційних чи різницевих рівнянь або рівнянь з частинними похідними чи приростами можна генерувати величезну кількість близьких, "схожих" розв'язків. Для економного ж опису сигналів і зображень з подальшим їх розпізнаванням були запропоновані ієрархічно організовані одно- та двовимірні стохастичні породжувальні граматики (генеративні моделі), які використані в якості зворотного зв'язку для направленого перебору та порівняння сигналів і зображень, що стало ефективною конструктивною реалізацією ідеї аналізу сигналів (зображень) через їх синтез [1—5].

ГЕНЕРАТИВНІ МОДЕЛІ ДЛЯ ОКРЕМИХ КАНАЛІВ ОК

На рис. 2 схематично подана трирівнева генеративна модель розуміння зорових сцен (зображень). Модель зовнішнього світу (ЗС-модель, див. далі) породжує (генерує) канонічну форму сцени. "Композитор" (генератор) сцен (ГС) за канонічною формою синтезує ескіз сцени. Генератор модельних сцен і зображень (ГМСЗ) за ескізом сцени, моделюючи освітлення, розміри об'єктів, їх взаємне розташування, повороти, нахил, проектування на площину тощо, породжує модельну сцену або зображення.

Компаратор (К) порівнює модельну сцену з пред'явленою для розпізнавання. Результат порівняння використовується як зворотній зв'язок, під впливом якого направленим перебором відсікаються неперспективні варіанти та відшукується найкраща модельна сцена, яка й аналізується і на підставі якої формується результат розуміння сцени.

На рис. 3 схематично подана трирівнева генеративна модель розуміння мовного сигналу. За канонічною формою передаваного смислу, що надходить з ЗС-моделі, генератор семантично еквівалентних речень (ГСЕР) породжує всі можливі речення, що передають один й той самий смисл, визначений вхідною канонічною формою. Генератор модельних сигналів зв'язного мовлення (ГМСЗМ) ставить у відповідність принятому орфографічному текстові всі можливі модельні сигнали злитого мовлення, які відображають розмаїті мовні сигнали, що відрізняються темпом, інтенсивністю, інтонацією, моделюють індивідуальні особливості мовлення тощо.

В компараторі (К) модельні сигнали порівнюються з мовним сигналом, що пред'являється для аналізу. Результат поточного порівняння використовується у зворотньому зв'язку для направленого відбору та пошуку найкращого модельного сигналу. Останній аналізується, тобто вказується, якому реченню (орфографічному текстові) він відповідає (текст можна надрукувати — автоматична машинка, що друкує під диктування) і/або отримати відповідну цьому текстові канонічну форму, що передається (автоматичне розуміння мовного сигналу — цей результат можна використати для виконання дії, щоб задовольнити інтереси людини, яка говорить).

Рис. 4 ілюструє функціонування генеративних моделей ОК, наведених на рис. 2 та рис. 3.

Відображено, що кожній предметній області відповідає скінченна множина канонічних форм смислів або сцен, що передаються; що кожній канонічній формі відповідає своя скінченна множина всіх можливих речень або ескізів сцен; що кожному реченню або ескізові сцени відповідає своя скінченна

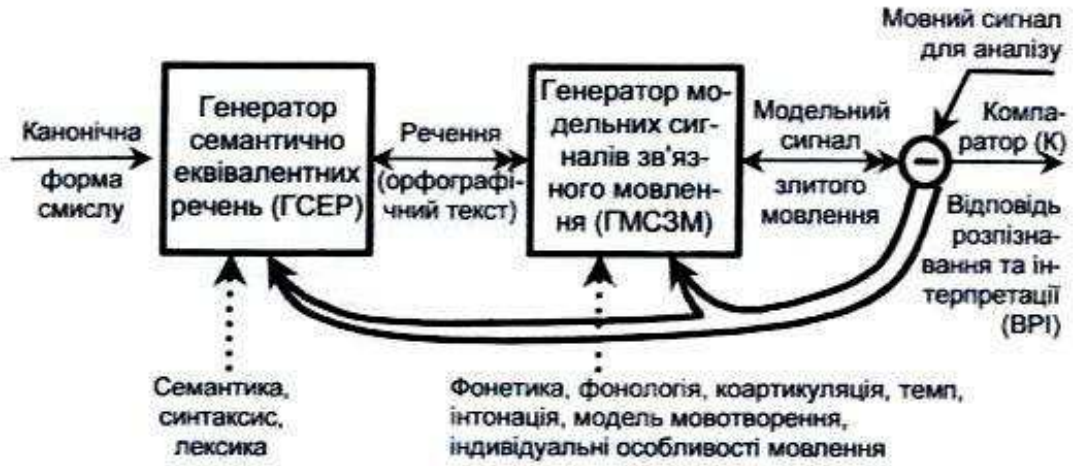


Рис. 3. Генеративна модель розуміння мовних сигналів.

множина всіх можливих модельних сигналів або зображень. На цьому ж рис. 4 пунктирними стрілками також показано як формуються відповіді розпізнавання та інтерпретації (ВРІ).

ГЕНЕРАТИВНА МОДЕЛЬ ПУД-МАШИНИ

Узагальнена структура диктувальної (Д) машини та машини усного (У) перекладу (П) — ПУД-машини, що ґрунтується на генеративній моделі, подана на рис. 5. Генеративна модель ПУД-машини утворюється об'єднанням генеративних моделей розуміння мовного сигналу для окремих мов. На рис. 5 зображені дві мови: канал природньої мови 1 та канал природньої мови 2.

Диктувальна машина. Проблема автоматичного редагування та друкування текстів під диктування на природній мові, наприклад 1, вирішується так. Для усномовного сигналу, що пред'явлений для оброблення, спершу знаходимо (шляхом направленого синтезу та відбору) такий модельний сигнал зв'язного мовлення, який є в певному сенсі найбільш схожим на розпізнаваний

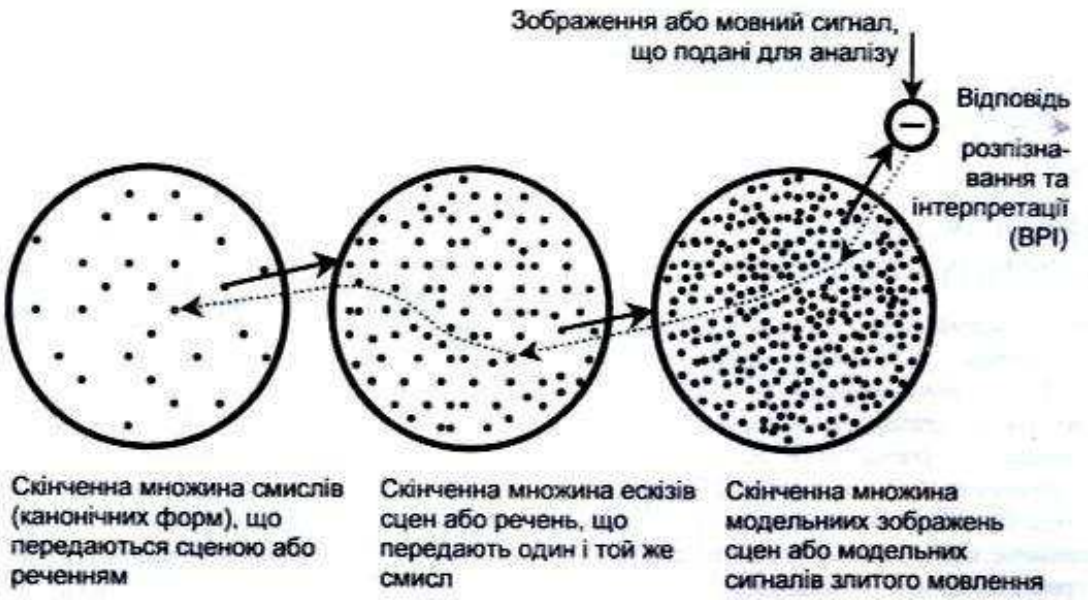


Рис. 4. Ілюстрація функціонування генеративних моделей ОК і формування відповіді розпізнавання та змістовної інтерпретації (ВРІ). Зворотніми стрілками показано порядок формування ВРІ

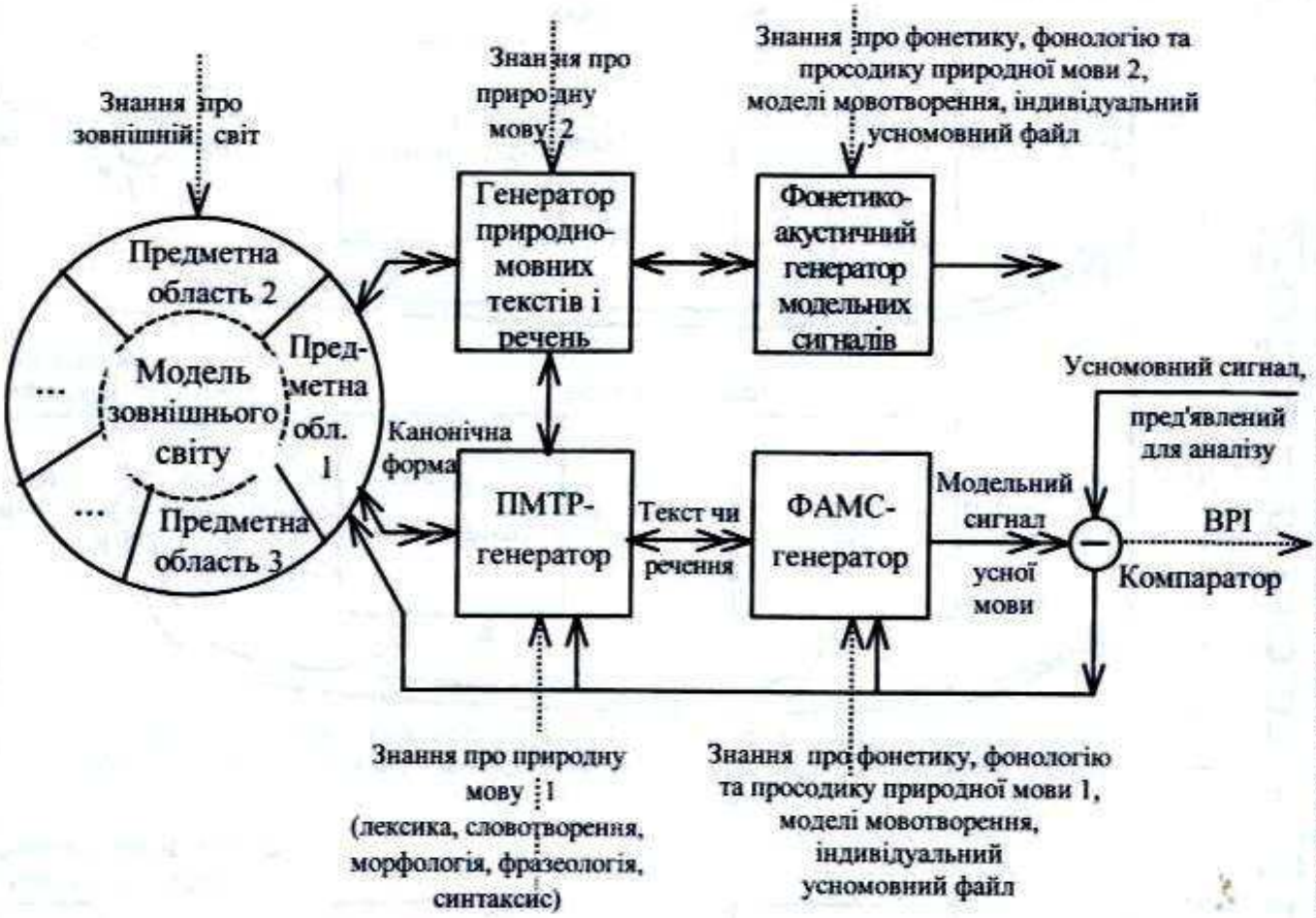


Рис. 5. Структура диктувальної машини та машини для усного перекладу.

сигнал. Потім цей знайдений, найбільш схожий модельний сигнал аналізується на предмет виявлення переданих цим сигналом послідовності слів та смислу. Оскільки виявлені послідовність слів і канонічна форма смислу є граматично та семантично допустимими, то сама послідовність слів може бути надрукована. Це і буде диктувальна машина.

Машина усного перекладу. Якщо ж скористатись канонічною формою смислу, що був переданий мовним сигналом на природній мові 1, і звернутись до ЗС-моделі та каналу природньої мови 2, то цій канонічній формі може бути поставлений у відповідність текст на природній мові 2, з тим же смислом, що і на природній мові 1 (перекладально-диктувальна машина), а сам текст на природній мові 2 може бути озвучений модельним сигналом природньої мови 2 – машина усного перекладу.

Моделі зовнішнього світу в ПУД-машині (ЗС-модель) відводиться найголовніша роль. ЗС-модель можна розглядати як об'єднання ЗС-підмоделей тематичних (предметних) областей з використанням спільної частини, що виражає загальні лінгвістичні властивості зовнішнього світу. ЗС-модель є спільною для всіх природних та штучних мов і, власне, мало від них залежить. Описується ЗС-модель спеціальною математичною мовою, наприклад мовою канонічних форм. Опускаючи тут подробиці задання ЗС-моделі, підкреслимо лише, що ЗС-модель генерує канонічні форми смислів, які можуть передаватись в процесі усномовної комунікації. Очевидно, що для кожної предметної області може бути вказана скінченна множина канонічних форм, можливих при діалозі. Наприклад, якщо йдеться про усномовний калькулятор на чотири арифметичні дії, то в цьому випадку допустимі канонічні форми мають просту структуру: операція та перший і другий операнди.

ГЕНЕРАТИВНА МОДЕЛЬ ОБРАЗНОГО КОМП'ЮТЕРА

Доповнимо генеративну модель ПУД-машини (див. рис. 5), зокрема, приєднаємо до моделі зовнішнього світу канал генерації зображень та просторових сцен, що складається з композитора зображень і сцен та генератора модельних зображень

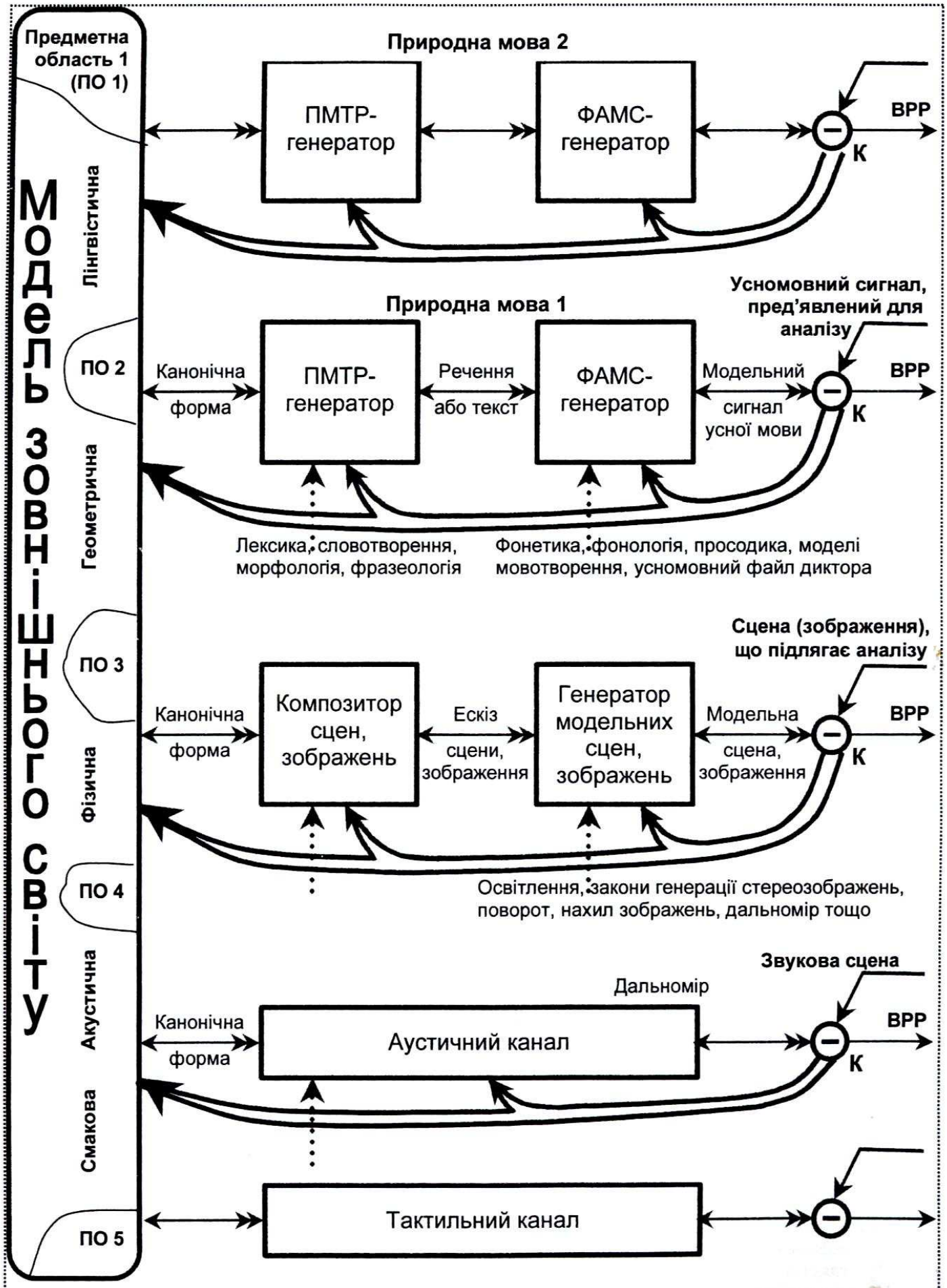


Рис. 6. Генеративна модель образного комп'ютера.

і сцен. Так само приєднаємо канал сприйняття звукових (неусномовних) сцен — акустичний канал, тактильний, нюховий та смаковий канали, які також організуємо за принципом генеративної моделі.

В результаті отримаємо генеративну модель образного комп'ютера, яка не тільки є диктувальною машиною та машиною усного перекладу, але й сприймає зображення та аналізує просторові сцени, дає їм усномовну інтерпретацію та робить їх текстовий опис, будує зображення за усномовним повідомленням, озвучує тексти, сприймає акустичні образи та звукові сцени, аналізує тактильну, нюхову та смакову інформацію.

Генеративна модель ОК схематично зображена на рис. 6. Взаємодія між каналами сприйняття інформації (на рис. 6 ця взаємодія не показана) та образне прийняття рішень організується образним операційним середовищем (ООС), в яке власне і "погружається" генеративна модель ОК.

На рис. 6 образне операційне середовище зображається полем прямокутної рамки цього рисунку.

РЕАЛІЗАЦІЯ ГЕНЕРАТИВНОЇ МОДЕЛІ

Виникає ряд принципових питань щодо реалізації генеративної моделі образного комп'ютера.

Розглянемо ці питання на прикладі реалізації диктувальної машини та машини усного перекладу — ПУД-машини, що є чи не однією з найскладніших підсистем у складі ОК.

Якщо говорити про ПУД-машину як тільки автономну підсистему, що не взаємодіє з іншими каналами ОК, то досить обмежитись тільки такою моделлю зовнішнього світу або мовою канонічних форм передаваного смислу, які є зорієнтованими на лінгвістику — на передачу смислу мовами (орфографічними текстами або усномовними сигналами).

Для спрощення викладок, розіб'ємо за передаваним смислом всі можливі в мові речення на предметні області, а всі можливі речення предметної області — на категорії або типи смислу.

Так, для інформаційно-довідникової служби аеропорту можна виділити такі категорії смислу: питання, які стосуються прибуття літаків; питання стосовно відльоту літаків; довідки про наявність квитків; довідки про маршрут; питання про розташування служб аеропорту тощо. Кожній категорії смислу відповідає скінченна множина типів речень, які передають цей смисл. Тип речень – це конструкція, що економно описує множини речень, які можна отримати із якогось одного речення шляхом незалежних допустимих замін та допустимих перестановок окремих слів та словосполучень. Основним елементом типу речень є підсловник. Підсловники іменуються по їх відношенню до предметної області.

Типи речень зручно задавати списковими структурами, наприклад, *LISP*-структурами. Отже, типи смислів специфікуються списком (переліком) своїх *LISP*-структур.

Іншим еквівалентним способом специфікації всіх можливих речень предметної области є орієнтована семантична мережа (ОСМ). ОСМ може бути побудована, зокрема, на основі категорій смислу та типів речень. Так, для кожного типу речень спочатку будується своя часткова ОСМ, а потім ОСМ для окремих категорій смислу та повна ОСМ для всієї предметної області утворюються шляхом простого об'єднання всіх часткових ОСМ. Звичайно, важливо мінімізувати кількість станів в ОСМ. Будь-яке речення може бути перевірене на допустимість в ОСМ. Якщо воно є допустимим, то переданий цим реченням смисл або канонічна форма смислу можуть бути найдені також за допомогою ОСМ, наприклад, шляхом аналізу типу речень та категорії смислу, яким це речення належить.

Не менш важливе питання — це питання породження, перебору та порівняння всіх можливих модельних сигналів з розпізнаваним. Але тут і далі матимемо на увазі лише такі процедури економного опису процесів генерації модельних сигналів та направленого перебору варіантів, коли результати поточного порівняння спостережуваного та модельних сигналів використовуються для звуження, при подальшій генерації, підмножини модельних сигналів, відповідно текстів і канонічних форм, що претендують у відповідь розпізнаванння, так, щоб гарантувати не втратити оптимальне рішення. Саме ці згадані дві суперечливі вимоги: економна специфікація породжувальних процедур та направлений перебір варіантів, – задовольняються ІКДП(*HCDP*)–технологією автоматичного розпізнавання, розуміння та синтезу усномовних сигналів [1–3, 5, 7–8]. Перша вимога задовольняється використанням ієрархічно (І (*H – Hierarchy*)) структурованих автоматних породжувальних граматик, які синтезують (композиція (К (*C – Composition*))) модельні сигнали, а друга – заснована на динамічному програмуванні (ДП (*DP – Dynamic Programming*)).

Один із можливих способів використання ОСМ в ПУД-машині полягає в акустичній деталізації ОСМ. Для цього кожне слово-місце в ОСМ заміщується його графом (як в *HCDP*- чи *HMM*[9]-технології). Тоді цей граф генеруватиме всі можливі сигнали цього слова в контексті злитого мовлення, а сама ОСМ стає усномовною, тобто такою, яка генерує модельні усномовні сигнали для вибраної предметної области. Тепер вирішення проблеми ПУД-машини для цієї предметної области полягає у знаходженні найбільш схожого модельного сигналу та у його аналізі на предмет виявлення послідовності слів та канонічної форми смислу, що передаються пред'явленим усномов ним сигналом [1–3, 5, 7–8].

Другою в ієрархії частиною ПУД-машини є генератор природномовних речень або текстів (ПМТР-генератор). Цей блок ґрунтується на знаннях про лексику, словотворення, морфологію, фразеологію та синтаксис кожної конкретної природної мови. Отримуючи від ЗС-моделі канонічну форму передаваного смислу, ПМТР-генератор породжує всі допустимі для даної мови речення і тексти, котрі виражають один і той же смисл, що визначений принятою канонічною формою. Можна сказати, що ПМТР-генератор є семантико-граматичною текстовою моделлю конкретної природної мови.

На третьому рівні ієрархії знаходиться фонетико-акустичний генератор модельних усномовних сигналів (ФАМС-генератор). Цей блок бере до уваги всі знання про фонетичні та фонологічні особливості певної природної мови, всі знання про мовотворення, включаючи моделі мовного тракту та джерел його збурення, рівно ж наші уявлення про такі явища як коартикуляція звуків, їх редукція, нелінійні зміни темпу та інтенсивності вимовляння, просодичні властивості вимовляння тощо. В цей же блок вводяться дані про індивідуальні особливості голосу у вигляді так званих індивідуальних усномовних файлів (ІУМФ) диктора [1–3, 5, 7–8].

Отже, ФАМС-моделі не тільки виражають фонетичні властивості (фонетика, фонологія, просодика) певної конкретної мови, але й включають моделі мовотворення, які є однаковими для всіх мов.

Отримавши від "свого", за мовою, ПМТР-блоку текст чи речення, ФАМС-генератор породжує всі можливі модельні сигнали зв'зного мовлення, що відповідають цьому текстові чи реченню та моделюють голос певного диктора, ІУМФ котрого є в бібліотеці дикторів ПУД-машини. Ці модельні сигнали відрізняються нелінійно змінюваним темпом та інтенсивністю вимовляння, просодичними характеристиками тощо [1–3, 5, 7–8].

ПРИКІНЦЕВІ ПОЛОЖЕННЯ

Генеративна модель розпізнавання усномовних сигналів (інша назва – ІКДП-технологія) була запропонована автором у 1966 році. Перша публікація з'явилась у 1968 році [2]. Ці та подальші публікації автора були визнані піонерними в світі (див., наприклад, [9]). З 1975 року генеративна модель розпізнавання мовних сигналів в аналогічній інтерпретації від імені інших авторів поширюється також під назвою HMM-модель [10].

В кінці 60-х автор спробував узагальнити генеративну модель і на розпізнавання зображень, використовуючи прийоми двовимірних генеративних процесів та відповідного їм деякого розширеного динамічного програмування [3]. Пізніше більш глибокі теоретичні розробки стосовно двовимірних породжувальних граматик та їх використання в обробленні зображень були виконані М.І.Шлезінгером [4, 6].

В 1997 році з'явилась ідея об'єднання досліджень з автоматичного розпізнавання образів (сприйняття слухової, зорової, акустичної, смакової, тактильної тощо інформації). Ця ідея почала реалізовуватись у 2000 році після схвалення ДНТП "Образний комп'ютер".

ЛІТЕРАТУРА

[1] Тарас Вінцюк, "Образний комп'ютер", *Зб. наук. праць "Сучасні проблеми в комп'ютерних науках"*, Вид-во Нац. ун-ту "Львівська політехніка", Львів, 2000, с. 5-14.

[2] Т.К. Винцюк, "Распознавание слов устной речи методами динамического программирования", *Кибернетика*, 1968, № 1, с. 81-88.

[3] Т.К. Винцюк, "Распознавание рукописных знаков методами динамического программирования", *Сб. «Кибернетика и вычислительная техника», Вып. 3, «Распознавание образов»*, Киев: "Наукова думка", 1969, с. 52-77.

[4] М.И.Шлезингер, "Синтактический анализ двумерных зрительных сигналов в условиях помех", *Кибернетика*, 1976, № 4, с. 113-130.

[5] Т.К. Винцюк, "Анализ, распознавание и интерпретация речевых сигналов", Киев: "Наукова думка", 1987, 264 с.

[6] М.И.Шлезингер, "Математические средства обработки изображений", Киев: "Наукова думка", 1989, 198 с.

[7] T.K. Vintsiuk, "HCDP-Technique for Automatic Analysis, Recognition and Understanding of Speech Signals", *Proc. First Intern. Conf. on Information Technology for Image Analysis and Pattern Recognition*, L'viv, 1990, Vol 1, pp 108 - 112.

[8] Taras K.Vintsiuk, "Two Approaches to Create a Dictation/ Translation Machine", *Proc. Second Intern. Workshop "Speech and Computer"*, Cluj-Napoca, 1997, pp 1-6.

[9] S.E.Levinson, "Structural Methods in Automatic Speech Recognition", *Proc. of the IEEE, Vol. 73, No. 11*, Nov. 1985, pp 1625-1650.

[10] F.Jelinek, "Continuous Speech Recognition by Statistical Methods", *Proc. IEEE, Vol. 64*, Apr. 1976, pp 532-556.

Analysis of optimal labelling problems and their application to image segmentation and binocular stereovision

Schlesinger M.I.[1], Flach B.[2]

[1]IRTC ITS, 40, prospect Academika Glusckova, 03680, Kyiv, Ukraine;
tel.: 266 62 08, E-mail: schles@image.kiev.ua

[2]Technische Universität Dresden, Fakultät Informatik, Institut für künstliche Intelligenz,
D-01062 Dresden, Deutschland
Email: bflach@ics.inf.tu-dresden.de

ABSTRACT

The approach to optimal labelling is described. Algorithms developed within the proposed approach are of polynomial complexity and defined on the whole class of labelling problems. Some labelling problems are processed so that no labelling is given and such result shall be interpreted as an answer "I do not know". However, if the algorithm produces a labelling it can be only optimal. Such feature of proposed algorithms distinguishes them essentially from known algorithms, which are either defined for some subclass of labelling problems or fulfil local improvements of labelling and do not provide the globally optimal solution.

1. FORMULATION OF THE OPTIMAL LABELLING PROBLEM

Let T be a finite set of pixels, $\Im \subset T \times T$ be a subset of pixel pairs referred to as neighbours, K be a finite set of labels. A function of the form $\bar{k} : T \to K$ will be called a labelling, the set of all possible labellings will be denoted K^T. The label of the pixel t will be denoted $k(t)$.

Let for every pair $tt' \in \Im$ of neighbours a function $g_{tt'} : K \times K \to R$ be defined as well as a function $q_t : K \to R$ for every pixel $t \in T$. The quality of the labelling $\bar{k} \in K^T$ is defined as

$$G(\bar{k}) = \sum_{tt' \in \Im} g_{tt'}(k(t), k(t')) + \sum_{t \in T} q_t(k(t)). \quad (1)$$

The optimal labelling problem consists in constructing an algorithm which gets input data

$$z = \langle T, \Im, K, (g_{tt'} \mid tt' \in \Im), (q_t \mid t \in T) \rangle$$

and finds the best labelling

$$\bar{k}^* = \arg\max_{\bar{k} \in K^T} \left[\sum_{tt' \in \Im} g_{tt'}(k(t), k(t')) + \sum_{t \in T} q_t(k(t)) \right]. \quad (2)$$

Hereinafter we will sometimes omit the second sum in the expression (2), assuming that all numbers $q_t(k)$, $t \in T$, $k \in K$, are zeros. This does not constrict the class of problems under consideration if each pixel has at least one neighbour. Indeed, for each labelling $\bar{k} \in K^T$ the equality

$$\sum_{tt' \in \Im} g_{tt'}(k(t), k(t')) + \sum_{t \in T} q_t(k(t)) = G(\bar{k}) = \sum_{tt' \in \Im} g^*_{tt'}(k(t), k(t')) \quad (3)$$

is valid, where

$$g^*_{tt'}(k, k') = g_{tt'}(k, k') + \frac{q_t(k)}{|N(t)|} + \frac{q_{t'}(k')}{|N(t')|} \quad (4)$$

and $N(t)$ is the set of neighbours of the pixel t.

The set of problems of the form (2) is *NP*-complete. The known approaches to the problem can be divided into two groups. In the works of the first group some polynomially solvable subclass of labelling problems is specified with constrains either on the functions $g_{tt'} : K \times K \to R$ [1, 2, 3, 6] or the neighbourhood $\Im$ [5, 7]. Algorithms of the second group fulfil local step-by-step improvements of current labelling. Such algorithms are defined on the whole set of labelling problems, but some problems are not solved correctly: the algorithm may specify a locally unimprovable labelling which is not optimal. We propose another approach that distinguishes from the well-known ones as it was quoted in the Abstract.

2. DESCRIPTION OF THE APPROACH

2.1. Definition of trivial problems

Let $z = \langle T, \Im, K, (g_{tt'} \mid tt' \in \Im) \rangle$ be input data for a labelling problem

$$\bar{k}^{*} = \underset{\bar{k}\in K^{T}}{\operatorname{argmax}} \sum_{tt'\in\Im} g_{tt'}\left(k(t),k(t')\right).$$

Let the functions $\bar{g}_{tt'} : K\times K \to \{0,1\}$ be such that

$$\bar{g}_{tt'}(k,k') = 1 \ ,$$

$$\text{if } g_{tt'}(k,k') = \max_{k,k'} g_{tt'}(k,k') \ ,$$

$$\bar{g}_{tt'}(k,k') = 0 \ ,$$

$$\text{if } g_{tt'}(k,k') < \max_{k,k'} g_{tt'}(k,k') \ .$$

The problem $z = \left\langle T,\Im,K,\left(g_{tt'} \mid tt'\in\Im\right)\right\rangle$ is called trivial if there exists a labelling $\bar{k}^{\circ}$ such that

$$\underset{tt'\in\Im}{\&}\ \bar{g}_{tt'}\left(k^{\circ}(t),k^{\circ}(t')\right) = 1.$$

Obviously, the labelling k° is optimal in this case. Indeed, for each labelling $\bar{k}'\in K^{T}$ the following chain of equalities and inequality

$$\begin{aligned} G\left(\bar{k}'\right) &= \sum_{tt'\in\Im} g_{tt'}\left(k'(t),k'(t')\right) \le \\ &\le \sum_{tt'\in\Im} \max_{k,k'} g_{tt'}(k,k') = \\ &= \sum_{tt'\in\Im} g_{tt'}\left(k^{\circ}(t),k^{\circ}(t')\right) = G\left(\bar{k}^{\circ}\right) \end{aligned} \tag{5}$$

is valid. The number $\sum_{tt'\in\Im} \max_{k,k'} g_{tt'}(k,k')$ in the chain does not depend on the labelling, it depends only on the problem z. This characteristic of the problem will be called the problem potential and denoted $\Phi(z)$,

$$\Phi(z) = \sum_{tt'\in\Im} \max_{k,k'} g_{tt'}(k,k'). \tag{6}$$

2.2. Equivalent problems

Two problems

$$z_1 = \left\langle T,\Im,K,\left(g^{1}_{tt'} \mid tt'\in\Im\right)\right\rangle$$

and

$$z_2 = \left\langle T,\Im,K,\left(g^{2}_{tt'} \mid tt'\in\Im\right)\right\rangle$$

are called equivalent if for each labelling $\bar{k}\in K^{T}$ the equality

$$\sum_{tt'\in\Im} g^{1}_{tt'}\left(k(t),k(t')\right) = \sum_{tt'\in\Im} g^{2}_{tt'}\left(k(t),k(t')\right) \tag{7}$$

is valid. This definition is not constructive because recognition of equivalency requires checking the $|K|^{|T|}$ equalities (7). The following theorem defines the equivalency in the constructive way.

Theorem 1. If the neighbourhood $\Im$ forms a connected graph, then the problems

$$\left\langle T,\Im,K,\left(g^{1}_{tt'} \mid tt'\in\Im\right)\right\rangle \ ,$$

$$\left\langle T,\Im,K,\left(g^{2}_{tt'} \mid tt'\in\Im\right)\right\rangle$$

are equivalent if and only if there exists an array of numbers $\varphi_{tt'}(k),\ t\in T,\ t'\in N(t),\ k\in K$, which satisfy the equalities

$$\left.\begin{aligned} &\varphi_{tt'}(k) + \varphi_{t't}(k') = g^{2}_{tt'}(k,k') - g^{1}_{tt'}(k,k'), \\ &\qquad t\in T,\ t'\in N(t),\ k\in K,\ k'\in K; \\ &\sum_{t'\in N(t)} \varphi_{tt'}(k) = 0,\ t\in T,\ k\in K. \end{aligned}\right\} \tag{8}$$

■

2.3. Transformation of the non-trivial problem into trivial

The proposed approach consists in searching for a trivial equivalent for the problem under solution. Certainly, it can be done only if such a trivial equivalent exists. Implementation of this idea is based on the following considerations.

Let Z be an equivalency class, z being a problem of this class; let $\bar{k}$ be some labelling. The problem potential $\Phi(z)$ does not depend on the labelling and the labelling quality $G\left(\bar{k}\right)$ does not depend on the problem of the class Z. The inequality

$$\Phi(z) \ge G\left(\bar{k}\right)$$

is evident (see (3)). If z^{*} is a trivial problem and $\bar{k}^{*}$ is an optimal labelling then the inequality

$$\Phi\left(z^{*}\right) = G\left(\bar{k}^{*}\right)$$

is evident too (see (3)). It means that the following theorem is valid.

Theorem 2. If z^* is a trivial problem then

$$z^* = \arg\min_{z \in Z} \Phi(z),$$

where Z is a set of problems which are equivalent to z^*. ■

The following theorem, which is inverse in certain sense, is valid too.

Theorem 3. If a class of equivalent problems includes at least one trivial problem then any problem

$$z^* = \arg\min_{z \in Z} \Phi(z)$$

is trivial. ■

The following theorem shows the constructive way for searching for a problem with the minimal potential.

Theorem 4. For each problem z' there exists a problem

$$z'' = \arg\min_{z \in Z} \Phi(z),$$

where Z is the class of problems, equivalent to z'. This problem is a solution of the following linear programming problem:

minimise $$\sum_{tt' \in \Im} h(tt') \tag{9}$$

under conditions

$$\left.\begin{aligned} & h(tt') \geq g''_{tt'}(k,k'),\ tt' \in \Im,\ k \in K,\ k' \in K, \\ & \varphi_{tt'}(k) + \varphi_{t't}(k') = g''_{tt'}(k,k') - g'_{tt'}(k,k'), \\ & \qquad t \in T,\ t' \in N(t),\ k \in K,\ k' \in K, \\ & \sum_{t' \in N(t)} \varphi_{tt'}(k) = 0,\ t \in T,\ k \in K. \end{aligned}\right\} \tag{10}$$

■

The fulfilled analysis allows to formulate the following approach to optimal labelling searching.

1. Find a problem equivalent to the initial one, which minimises the problem potential.
2. Check whether the found problem is trivial or not.
3. If YES, declare any solution to the trivial problem to be a solution to the initial problem.
4. If NO, choose no solution and interpret such an outcome as "I DO NOT KNOW" answer.

We will show how some image segmentation and binocular stereovision problems are reduced to optimal labelling searching. Moreover, owing to the peculiarities of these problems the do-not-know answer is impossible. It means that problems of such class admit an exact solution.

3. IMAGE SEGMENTATION

Let T be a set of pixels, $\Im$ be a neighbourhood, X be a set of signal values observed in each pixel, K be a set of segment names. An image is a function $\bar{x}: T \to K$ and a segmentation is a function $\bar{k}: T \to K$.

A priori quality is defined for each segmentation; it is based on the intuitive idea that two neighbouring pixels most likely belong to the same segment. This can be expressed so that the a priori quality of the segmentation $\bar{k}$ is the sum

$$\sum_{tt' \in \Im} g_{tt'}(k(t), k(t')),$$

where $$\left.\begin{aligned} & g_{tt'}(k,k') = \alpha > 0, \text{ if } k = k', \\ & g_{tt'}(k,k') = 0, \quad \text{ if } k \neq k'. \end{aligned}\right\} \tag{11}$$

For each pair $\bar{x}, \bar{k}$ (image - segmentation) a similarity measure is defined

$$\sum_{t \in T} q_t(k(t), \bar{x}). \tag{12}$$

The numbers $q_t(k, \bar{x})$, $k \in K$, $\bar{x} \in X^T$, in the sum (12) express to what extent a decision for the segment k in the pixel t goes with the image $\bar{x}$ under observation. We do not here go into problems on how these numbers must be reasonably chosen, since a possibility of constructive problem solution is not determined by the form of the functions q_t but the form of the local qualities $g_{tt'}$ given by the expression (11).

The segmentation problem is formulated as searching for a function $\bar{k}: T \to K$, which maximises the sum of its a priori quality and its similarity to the image $\bar{x}$.

4. BINOCULAR STEREOVISION

Let T be a finite set of points on a 2-D-plane. A surface is a function $\bar{k}: T \to K$, where $k(t)$ is a height of the surface above the point $t \in T$. Let a subset of admissible surfaces is chosen in the following way: for each pair $tt' \in \Im$ of neighbouring points a number $\Delta_{tt'}$ is specified, and a surface $\bar{k}$ is regarded as admissible if heights in neighbouring points differ by no more than $\Delta_{tt'}$. It means that admissibility of the surface $\bar{k}$ is defined by the sum

$$\sum_{tt' \in \Im} g_{tt'}(k(t), k(t')),$$

where $$\left.\begin{aligned} g_{tt'}(k,k') &= 0, \quad \text{if } |k-k'| \le \Delta_{tt'}, \\ &= -\infty, \text{ if } |k-k'| > \Delta_{tt'}. \end{aligned}\right\} \quad (13)$$

A surface is admissible if the sum (13) is equal to zero and inadmissible otherwise.

The surface $\bar{k} : T \to K$ is not directly observable. Instead, there are two images $\bar{x}_1$ and $\bar{x}_2$ under observation, which form a stereopair. On the basis of these images the numbers $q_t(k, \bar{x}_1, \bar{x}_2)$ are calculated, which specify, how well a decision that the height of the surface above the point t is k goes with the observable images $\bar{x}_1$ and $\bar{x}_2$. Degree of conformity of the surface $\bar{k}$ with the stereopair $\bar{x}_1$ and $\bar{x}_2$ is defined as the sum

$$\sum_{t \in T} q_t(k(t), \bar{x}_1, \bar{x}_2).$$

We do not here consider how the numbers $q_t(k, \bar{x}_1, \bar{x}_2)$ must be calculated, because it will be shown below that a possibility of the constructive problem solution is determined only by the form (13) of the functions $g_{tt'}$. The binocular stereovision problem consists in searching for such an admissible surface which shows the best conformity with the observations $\bar{x}_1$ and $\bar{x}_2$.

5. MONOTONOUS LABELLING PROBLEMS

The labelling problem is called monotonous if the set K is ordered and the numbers $g_{tt'}(k,k')$ satisfy inequalities

$$g_{tt'}(k_1,k_2') + g_{tt'}(k_2,k_1') \le g_{tt'}(k_1,k_1') + g_{tt'}(k_2,k_2') \quad (14)$$

for any $tt' \in \Im$ and $k_1 < k_2$, $k_1' < k_2'$.

Image segmentation into two segments and binocular stereovision problem in the above stated formulations are monotonous labelling problems.

It is worth mentioning that in monotonous problems only the form of functions $g_{tt'}$ is restricted and by no means the form of functions q_t. It was noted above that the numbers $q_t(k)$ can be set to zeros by means of changing the numbers $g_{tt'}(k,k')$ without loss of generality (see (4)). It is essential that after changing the numbers $g_{tt'}(k,k')$ by the formula (4) the problem remains monotonous. Moreover, any equivalent transformation of the problem preserves monotonicity in the sense of definition (7).

A remarkable attribute of monotonous problems consists in validity of the following theorem.

Theorem 5. For any monotonous problem there exists an equivalent trivial problem. ■

CONCLUSION

The approach to construction of optimal labelling algorithms is described, which are not defined on a subclass of labelling problems, but on the class of all possible problems, which is well known to be *NP*-complete. After processing some input problems such an algorithm may not output a labelling but a special answer, which must be understood as a denial of solution of exactly this problem, as an answer "*I do not know*". Essential advantage of such algorithms is, however, that the situation is excluded when an algorithm outputs a non-optimal labelling. Construction of such optimal labelling algorithms is feasible, this being the main scientific result of the given research. Besides the main result it is essential that the certain problem subclass is described, for which the answer "*I do not know*" cannot be given. Some image segmentation and binocular stereovision problems are included in this class, and their exact solutions can be thus obtained.

REFERENCES

[1] Boykov, Yu., Veksler, O., and Zabih, R., Fast Approximate Energy Minimization via Graph Cuts, in: IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 23, No. 11, November 2001.

[2] Ishikawa H., Geiger D., Segmentation by Grouping Junctions, in: IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Santa Barbara, CA, June 1998.

[3] Kolmogorov, V., and Zabih, R., What Energy Functions Can Be Minimized via Graph Cuts?, in:A. Heyden et al. (ed.), Computer Vision - ECCV2002, LNCS 2352, Copenhagen, Denmark, May 2002.

[4] Schlesinger M.I. Sintaksicheskij analiz dvumernyh zritel'nyh signalov v uslovijah pomeh, in Russian, (Syntactic analysis of noisy images). Kybernetika, 4, 1976.

[5] Schlesinger, M.I., Matematiceskie sredstva obrabotki izobrazenij, in Russian (Mathematical tools for image processing). Naukova Dumka, Kiev, 1989.

[6] Schlesinger, M.I., and Flach, B., Some solvable subclasses of structural recognition problems, in: T. Svoboda (ed.), Proceedings of the Czech Pattern Recognition Workshop 2000, Praga 2000.

[7] Schlesinger, M.I., and Hlavac, V., Ten Lectures on Statistical and Structural Pattern Recognition, Computational Imaging and Vision, Kluwer Academic Publishers, Dordrecht / Boston / London, 2002.

СПРИЙНЯТТЯ І АНАЛІЗ АКУСТИЧНИХ СИГНАЛІВ, РОЗПІЗНАВАННЯ ТИПУ ДЖЕРЕЛА ЗВУКУ, ОЦІНЮВАННЯ НАПРЯМКУ ТА ВІДСТАНІ ДО ДЖЕРЕЛА

Ярема Зєлик, Михайло Личак, Олександр Лук'янчук
Інститут космічних досліджень НАН та НКА України (ІКД НАНУ-НКАУ)
Проспект Академіка Глушкова, 40, Київ 03680, Україна
Тел.: +380 44 2661291 Факс: +380 44 2664124
adapt@space.is.kiev.ua

ABSTRACT

Ya. Zyelyk, M. Lychak, A. Luckhianchuk. Perception and analysis of acoustic signals, recognition of sound source type and estimation of a direction and distance to a source. The initial results of researches obtained within the framework of State scientific-technological program "Pattern Computer" are considered. The realistic mathematical models of an acoustic component of the external world are offered. The typical sources of an acoustic noise in aerospace engineering are analyzed. The structure of a receiving acoustic antenna and choice of input-output device of acoustic signals in the computer is justified. The functional purpose of the designed software in MATLAB is described. The results of inputting, analysis and representation in the forms of vector time series, matrices of spectral densities and matrices of spectrograms (matrices of patterns!) of real acoustic sources signals in the room are represented. The technique of estimation of a direction and distance to a sound source is offered and the examples of its application are considered.

ВСТУП

Акустичне поле, збуджуване природними і штучними джерелами, є важливим компонентом зовнішнього світу. Згідно з концепцією Образного комп'ютера (ОК) вимоги щодо підсистеми ОК, пов'язаної з врахуванням та представленням у внутрішніх моделях ОК цього компоненту формулюються таким чином. Сприйняття та аналіз акустичних сигналів, розпізнавання типу джерела звуку, оцінювання напрямку та відстані до джерела. Виконання цих вимог у створюваній в ІКД НАНУ-НКАУ експериментальній системі є предметом цієї статті. У ній відображені результати досліджень, проведених в ІКД НАНУ-НКАУ за проектом № 15 "Сприйняття та розпізнавання просторових звукових образів джерел акустичних сигналів, імітація (генерація) та активна компенсація акустичних полів", що виконувався в рамках Державної науково-технічної програми "Образний комп'ютер". Автори мають професійний досвід створення та широкого впровадження на протязі понад 20 років цифрових систем реального часу для відтворення акустичних полів.

1. КОНСТРУКТИВНІ МАТЕМАТИЧНІ МОДЕЛІ АКУСТИЧНОГО КОМПОНЕНТУ ЗОВНІШНЬОГО СВІТУ

Акустичне поле - це функція координат простору і часу, що характеризує стан суцільного середовища (далі - газу), у якому звук поширюється за допомогою хвиль. При невеликих амплітудах та значних довжинах хвиль поле звуку може бути вичерпним чином описане однією скалярною функцією - тиском , яка задовольняє хвильовому рівнянню. Для отримання розв'язку цього рівняння, який описує конкретну реалізацію хвильового поля, задаються початкові і граничні (разом крайові) умови, і мають справу з крайовими задачами розсіяння звуку. У класичній акустиці для хвильового рівняння ставляться прямі та обернені крайові задачі розсіяння та випромінювання. Найбільш загальна крайова задача характеризується так званими імпедансними граничними умовами. Акустичний імпеданс є величиною, пропорціональною відношенню акустичного тиску до його похідної за нормаллю у деякій точці поверхні розсіяння (випромінювання). Імпеданс є об'єктивною характеристикою поверхні тільки для вузького класу крайових задач. Задачі такого класу зводяться до простих схем розсіяння (випромінювання) плоскої хвилі на плоскій поверхні, сферичних хвиль точкового джерела на площині, циліндрі, кулі, тощо. В реальних граничних задачах в загальному випадку акустичний імпеданс поверхні, крім залежності від частоти і акустичних властивостей матеріалу поверхні, залежить ще від координат точки на ній (форми поверхні), кута падіння (при плоскій хвилі), а при довільній формі хвильового фронту є функцією його форми. Таким чином, взагалі акустичний імпеданс як неперервна функція наперед не може бути заданий без проведення дискретних вимірювань. Неможливість задання акустичного імпедансу поверхні особливо проявляється у закритих приміщеннях через багатократні випадкові відбивання хвиль від стінок і те, що ми не в стані передбачити результуючий хвильовий фронт. Отже, при класичному підході хвильові задачі розсіяння і випромінювання в акустиці в загальному випадку практично не можуть бути поставлені через неможливість апріорного

задання граничних умов як неперервних функцій. Тому в загальному випадку хвильові акустичні моделі не можуть бути конструктивними моделями зовнішнього світу при їх представленні у внутрішніх моделях ОК.

Конструктивним шляхом наближеного вирішення хвильових задач акустики в загальному випадку, у тому числі в закритих приміщеннях, може бути відновлення граничних умов і (або) всього шуканого розв'язку (інтерполяція акустичних полів) за даними дискретних вимірювань. Авторами запропонований новий оригінальний підхід до розв'язання задачі відновлення акустичних полів у закритих приміщеннях за даними дискретних вимірювань, який полягає в апроксимації поля відрізком потрійного ряду за системою відомих функцій, коефіцієнти якого знаходяться в результаті вирішення трикритеріальної екстремальної задачі, один з критеріїв якої враховує саме акустичну природу поля і геометрію довільних обмежуючих поверхонь. При розв'язанні екстремальної задачі одержана система лінійних алгебраїчних рівнянь для знаходження коефіцієнтів ряду з матрицею істотно меншої розмірності, ніж розмірність матриці системи, до якої приводило б використання для розв'язування відповідної крайової задачі (з залученням даних вимірювань) скінченнорізницевих методів. До того ж конкретна скінченнорізницева схема справедлива для єдиної реалізації поля, а для інших реалізацій необхідно було б будувати та розв'язувати інші скінченнорізницеві схеми. При відновленні поля запропонованим підходом для іншої реалізації поля слід по-новому вирахувати тільки елементи вектора у правій частині рівнянь, що служить для знаходження коефіцієнтів розкладу поля.

Таким чином, конструктивними математичними моделями акустичного компоненту зовнішнього світу можуть бути ті, що базуються на дискретному у просторі представленні поля звуку акустичними сигналами мікрофонів та інтерполяції поля з врахуванням саме акустичної його природи розглянутим вище методом за даними дискретних спостережень.

2. ВИПАДКОВІ АКУСТИЧНІ ПОЛЯ

Реальні акустичні поля найчастіше розглядаються як випадкові і можуть бути описані потенціалом швидкостей $\Phi(\vec{r},t)$ як випадковою функцією координат і часу. При статистичному підході до задач акустики мова йде про відшукання загальних властивостей ансамблю реалізацій поля, який має місце при статистично заданих умовах. Хвильові рівняння і крайові умови набувають стохастичного характеру, і функції, оператори та параметри, що присутні в них, є випадковими, заданими своїми багатовимірними розподілами ймовірностей. Повний статистичний опис поля може бути здійснений також за допомогою моментів, які є функціями координат і часу, що вираховуються через багатовимірні функції розподілу шляхом відповідних усереднень за ансамблем реалізацій поля. Для випадкових ергодичних полів вирахування моментів може бути здійснене без знання багатовимірних функцій розподілу випадкового поля (які до того ж ми не в стані оцінити на практиці) шляхом усереднення у часовій або частотній (просторовій чи просторово-частотній області) алгоритмічно-програмними чи апаратними засобами. У більшості розглядуваних задач відтворення і відновлення полів хвильове поле формується в результаті суперпозиції значного числа незалежних парціальних полів. Тому функції розподілу поля можна вважати нормальними, а його опис на рівні перших двох моментів - статистично повним.

Випадкове стаціонарне ергодичне акустичне поле потенціалів швидкостей (тисків) характеризується функцією просторово-часової кореляції

$$R(\vec{r}_1,\vec{r}_2,\tau)=\overline{\Phi(\vec{r}_1,t)\cdot\Phi^*(\vec{r}_2,t+\tau)}, \qquad (1)$$

де $\vec{r}_1$ і $\vec{r}_2$ радіус-вектори двох довільних точок простору, τ - часовий зсув, риска зверху - усереднення за часом. Для випадкового стаціонарного в широкому сенсі акустичного поля функція (1) є комплексною аналітичною за часовим аргументом τ і припускає такий спектральний розклад:

$$R(\vec{r}_1,\vec{r}_2,\tau)=\int_0^{\infty}S(\vec{r}_1,\vec{r}_2,\omega)\,e^{-j\omega\tau}d\omega \qquad (2)$$

$$S(\vec{r}_1,\vec{r}_2,\omega)=\frac{1}{2\pi}\int_{-\infty}^{\infty}R(\vec{r}_1,\vec{r}_2,\tau)e^{j\omega\tau}d\tau. \qquad (3)$$

Таким чином, присутня в спектральному розкладі (2), (3) функція спектральної густини $S(\vec{r}_1,\vec{r}_2,\omega)$ рівною мірою, як і функція просторово-часової кореляції, є вичерпною характеристикою випадкового стаціонарного акустичного поля з нормальними густинами розподілу ймовірностей.

На практиці технічно можливе оцінювання функції просторово-часової кореляції при розгортанні у часі для низки фіксованих точок простору, де знаходяться мікрофони. Таким чином, в контрольних точках акустичне поле можна характеризувати векторним випадковим стаціонарним процесом. Результати оцінювання кореляційних властивостей випадкового стаціонарного акустичного поля в m точках можна представити матрицею просторово-часової кореляції $\mathbf{R}=\left[R_{ij}\right](i=\overline{1,m};j=\overline{1,m})$ або матрицею спектральних густин спостережуваного в цих точках векторного випадкового процесу $\mathbf{S}=[S_{ij}](i=\overline{1,m};j=\overline{1,m})$, де R_{ij} і S_{ij} - оцінені значення функцій (2) і (3) відповідно в точках з координатами $\vec{r}_i$ і $\vec{r}_j$. Діагональні елементи матриці $\mathbf{S}$ (власні спектри) являють собою спектральну густину потужності поля в кожній точці простору, а недіагональні (взаємні спектри)

характеризують просторову кореляцію поля на частоті ω.

3. ПРИЙМАЛЬНА АКУСТИЧНА АНТЕНА ТА ПРИСТРІЙ ВВЕДЕННЯ-ВИВЕДЕННЯ АКУСТИЧНИХ СИГНАЛІВ

Сенсорними елементами приймальної акустичної антени є конденсаторні мікрофони типу МК 102 (фірми RFT) з чутливістю 50 мВ/па і практично круговою діаграмою спрямованості разом з узгоджуючими підсилювачами Mv 102. Конструкція акустичної антени визначається прийнятим методом оцінювання напрямку та відстані до джерела звуку за допомогою ненапрямлених мікрофонів. В такому випадку необхідно мати мінімум 4 мікрофони, 3 з яких розташовані на одній лінії, а четвертий – не перпендикулярній лінії (всі в одній площині) у тривимірному просторі.

Пристроєм введення-виведення акустичних сигналів в комп'ютер звукова карта DELTA 44 лінії M Audio (корпорація MIDIMAN), яка забезпечує синхронне введення-виведення звукових сигналів чотирма незалежними каналах з частотою перетворення 96000 відліків за сек. і розрядністю 24 біти.

4. ПРОГРАМНЕ ЗАБЕЗПЕЧЕННЯ

Програмне забезпечення системи сприйняття, аналізу акустичних сигналів та розпізнавання типу джерела звуку розробляється у потужному програмному середовищі MATLAB, який є і мовою програмування водночас. Ліцензійна мережева версія системи MATLAB на 5 клієнтів придбана ІКД НАНУ-НКАУ в авторизованого реселлера HUMUSOFT s.r.o. (Prague, Czech Republic) корпорації The MathWorks, Inc. (Natick, USA). Маючи додаткові колекції проблемно-орієнтованих функцій MATLAB Application Toolboxes, і зокрема, Data Acquisition Toolbox та Signal Processing Toolbox ця система дозволяє при наявності комп'ютера з відповідним характеристиками та звуковою картою здійснювати введення, аналіз, оброблення та синтез звукових сигналів у реальному масштабі часу. Відлагодивши програмно-алгоритмічне забезпечення підсистеми сприйняття, введення та розпізнавання в MATLAB, можна за допомогою Matlab Compiler трансформувати створені програмні модулі у мову *C* і створити виконувані модулі, незалежні від наявності на деякому іншому комп'ютері MATLAB. При виконанні трансформованих модулів будуть реалізуватися зручні, дружні і багаті графічні інтерфейси, створені засобами MATLAB. Програмне забезпечення складається з основних модулів такого призначення:

- синхронне введення 4-ох звукових сигналів і запис їх в файл;
- циклічна видача на гучномовці записаного векторного звукового сигналу;
- візуалізація часових реалізацій сигналу, спектральних густин;
- проектування і візуалізація цифрових фільтрів;
- оцінювання власних і взаємних спектрів, функції когерентності оцінювання спектрограм;
- оцінювання спектрограм, як часово-частотних образів акустичних сигналів;
- знаходження часу різниці ходу сигналів до мікрофонів для стаціонарних сигналів на основі аналізу фази функції когерентності, максимуму кореляційної функції;
- знаходження моментів часу початку дії акустичних сигналів на мікрофони і вирахування часу різниці ходу сигналів;
- розв'язання геометричної задачі локалізації одного джерела у тривимірному просторі за відомими значеннями часу різниці ходу сигналів до мікрофонів.

5. ОЦІНЮВАННЯ НАПРЯМКУ ТА ВІДСТАНІ ДО ДЖЕРЕЛА

Для розв'язання задачі локалізації джерела звуку у просторі перш за все слід знати час затримок акустичних сигналів, що сприймаються мікрофонами приймальної акустичної антени. В залежності від того, чи є спостережувані за допомогою мікрофонів сигнали стаціонарними, чи нестаціонарними, для визначення часу затримок сигналів при надходженні до мікрофонів, використовуються різні методи. Для стаціонарних акустичних сигналів, незначно спотворених майже некорельованими сигналами, час затримки надходження сигналу до більш віддаленого мікрофона порівняно з менш віддаленим можна визначити або за аргументом максимуму їх взаємної кореляційної функції або за фазою функції когерентності. Однак на практиці при локалізації джерела звуку в достатньо малих приміщеннях навіть для стаціонарних сигналів мають місце значні труднощі. Вони спричинені розподіленістю джерела у просторі, значною реверберація від різноманітних відбиваючих поверхонь, що породжує значні корельовані шуми, які спотворюють корисні акустичні сигнали. В цьому випадку ефективним виявляється метод локалізації джерела за оцінюванням різниці моментів часу появи імпульсу акустичного сигналу на відповідних мікрофонах приймальної антени після того, як джерело тільки почало випромінювати звук (після "мовчання"). При реалізації такого методу маємо справу з ефектом променевого поширення прямих акустичних хвиль в першу чергу саме від джерела до мікрофонів найкоротшим шляхом, а не від різноманітних відбиваючих поверхонь. Внесок же відбитих внаслідок реверберації хвиль проявляється у прийнятих за допомогою мікрофонів сигналах явно пізніше, ніж мають місце моменти появи імпульсів інформативних сигналів на виходах мікрофонів. Для точного визначення моментів появи імпульсів

акустичних сигналів на мікрофонах розв'язується задача апроксимації фронту імпульсного сигналу лінійною залежністю чи кубічним сплайном за критерієм мінімуму середньоквадратичної похибки на деякому часовому інтервалі, що містить момент початку імпульсу. Сам момент початку оцінюється як значення часового аргументу, при якому відхилення значення апроксимуючої функції від значення сигналу є мінімальним.

На рис.1. зображені часові реалізації акустичних сигналів 4-ох мікрофонів приймальної антени, введених в комп'ютер і візуалізованих. Сигнали збуджені випромінюванням гучномовця, заживленого акустичним сигналом шуму літака АН 70 від звукової карти іншого незалежного комп'ютера.

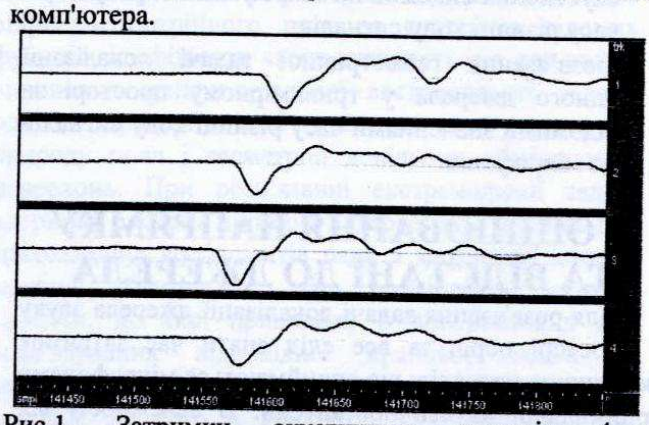


Рис.1. Затримки акустичних сигналів 4-ох мікрофонів після вмикання джерела звуку

На рис. 2 схематично зображена геометрія задачі локалізації в площині одного джерела звуку, що знаходиться в точці S за результатами спостереження акустичних сигналів двома мікрофонами, розміщеними відповідно в точках D_1 і D_2.

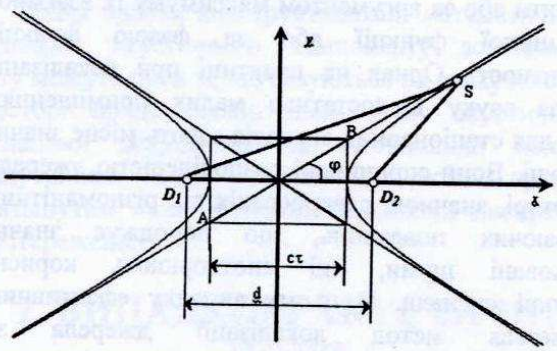


Рис. 2 Локалізація джерела звуку у площині

За відомою швидкістю поширення звуку c та затримкою τ між двома сигналами мікрофонів вираховується величина

$$\Delta d = SD_1 - SD_2 = c\tau \qquad (4)$$

Рівняння (4) задає множину точок, що лежать на гіперболічній поверхні обертання відносно осі OX

Кут напрямку поширення звуку від джерела у фіксованій площині знаходиться як

$$\cos\varphi = c\tau / d.$$

Використовуючи 4 мікрофони і зафіксувавши положення приймальної антени відносно осей прямокутної системи координат, знаходимо координати точки локалізації джерела як точки перетину трьох відповідних гіперболоїдів обертання.

6. ОСНОВНІ РЕЗУЛЬТАТИ ПРЕДСТАВЛЕННЯ АКУСТИЧНИХ СИГНАЛІВ ЗА ДОПОМОГОЮ ОБРАЗІВ

На рис. 3 представлені оцінки 1-ї власної (з 4-ох) та модулів 3-ох взаємних (з 10) спектральних густин 4-ох акустичних сигналів мікрофонів при генеруванні 1-им гучномовцем сигналу шуму літака АН 70.

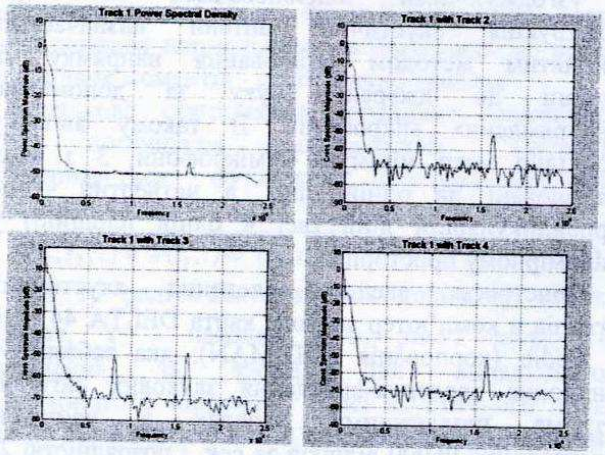


Рис. 3. Спектральні густини 4-ох акустичних сигналів

Спектрограма, відображена на рис. 4, є не графіком, а образом акустичного сигналу 1-го мікрофона, відфільтрованого в діапазоні 20 Hz – 20 KHz, і відображає часово-частотну залежність миттєвого спектру сигналу.

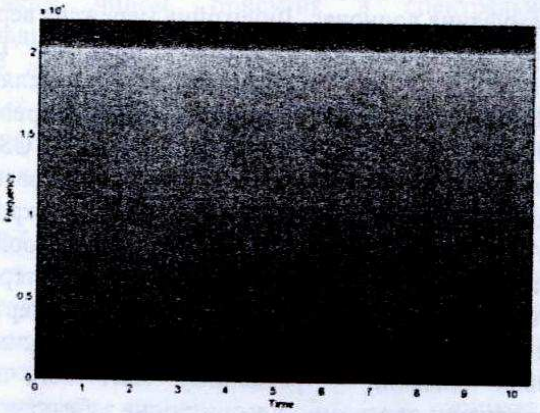


Рис. 4. Спектрограма сигналу 1-го мікрофона

Для 4-ох акустичних сигналів, прийнятих мікрофонами антени, будується 10 подібних спектрограм (трикутна матриця образів сигналів) – 4 діагональні елементи – образи власних миттєвих спектрів сигналів, а 6 наддіагональних елементів – образи модулів взаємних спектрів сигналів.

ВИСНОВКИ

Розроблена методика оцінювання напрямку та відстані до джерела звуку. Задача розпізнавання типу джерела звуку зведена до задачі розпізнавання матриці образів (спектрограм) акустичних сигналів, прийнятих акустичною антеною.

РОЗРОБКА КОМП'ЮТЕРНИХ ТЕХНОЛОГІЙ МОДЕЛЮВАННЯ ТА КЕРУВАННЯ ВІЗУАЛЬНИМИ ОБРАЗАМИ ЛЮДСЬКОГО ОБЛИЧЧЯ ПРИ СИНТЕЗІ МОВЛЕННЯ[1]

Юрій Крак, Тарас Вінцюк, Микола Кириченко, Федір Гаращенко, Олександр Бармак

Київський національний університет імені Тараса Шевченка
64 вул. Володимирська, Київ 03022
Електронна пошта: krak@unicyb.kiev.ua, vintsiuk@uasoiro.org.ua, kir@dept115.icyb.kiev.ua, garash@unicyb.kiev.ua, barmak@svitonline.com

Yuriy Krak, Taras Vintsyuk, Mykola Kirichenko, Fedir Garashchenko, Olexander Barmak, **Development of a computer technologies for modeling and control of visual images of human face under speech synthesis.** Technologies for modeling of processes animation of arbitrary text of are proposed. Mathematical model of human head with possibilities of facial expression during conversation and synchronization with speech are created.

Вступ

Розвиток сучасних комп'ютерних технологій моделювання і обробки аудіо і відео інформації спрямований на створення систем візуалізації процесу промовляння за допомогою комп'ютерних засобів [1-8]. Важливість використання таких технологій підтверджується їх включенням в MPEG-4 стандарт [9]. Дані технології мають значні прикладні застосування в інтелектуалізації роботи з комп'ютером, кінематографії, телебаченні, телефонії, передачі інформації, тощо. Схема системи озвучення текстів з моделюванням голови людини наведена на рис. 1.

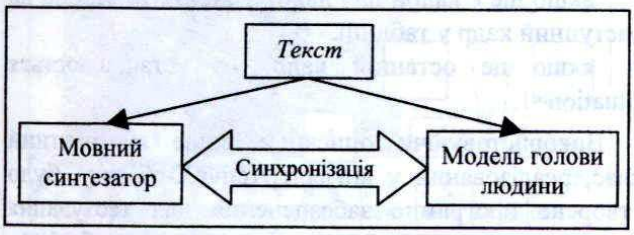


Рис.1.

Для реалізації такої системи потрібно створити програмне забезпечення для побудови об'ємної моделі голови людини і алгоритми, що реалізують динамічну генерацію візуальних образів людського обличчя при синтезі мовлення. Тобто потрібно вміти моделювати процес зміни образів обличчя людини, який синхронний звукам (фонемам), генерованим мовним синтезатором. Припускається, що є мовний синтезатор[10,11], який вхідний орфографічний текст перетворює у розмічений фонетичний текст (транскрипцію), тобто набір фонем з тривалістю кожної фонеми та генерує звук.

В даній роботі пропонується візуалізація процесу промовляння двома методами:

1. Як послідовність змін кадрів зображення конкретного людського обличчя, яке промовляє фонетично розмічений текст – 2D-технологія.
2. Як анімацію (плавний перехід від однієї моделі до іншої) послідовностей об'ємних 3D-моделей людського обличчя, яке промовляє фонетично розмічений текст – 3D-технологія.

Дослідження цих методів, не дивлячись на їх принципову відмінність, показали подібність їх алгоритмічної реалізації. Ця подібність спонукала до створення абстрактного класу, у якому реалізовано перший метод. Важливо відмітити, що після побудови та тестування об'єкта, створеного на основі цього абстрактного класу перехід до другого методу полягає лише у переписуванні відповідних методів об'єкту-нащадку [12-14]. Розглянемо детально два запропонованих методи.

1. 2D-технологія

Алгоритм реалізації першого методу можна описати наступною послідовністю кроків:

1. За допомогою відеокамери знімається людське обличчя, яке промовляє довільний текст. При цьому дотримуються наступні обмеження:
   - 1.1. Камера закріплена на штативі, та сфокусована на голову.
   - 1.2. Фон зйомки має бути однорідним.
   - 1.3. Текст має промовлятися у одному (середньому) темпі.
   - 1.4. При промовлянні тексту голова не повинна рухатися.
   - 1.5. Текст має містити набір фонем за допомогою яких можливо збудувати довільний інший текст (так звана навчальна вибірка).
2. Отримане у п.1 аналогове зображення з допомогою стандартних засобів перетворюється у цифрове. При цьому дотримуються наступні обмеження:
   - 2.1. Має бути файл формату AVI.
   - 2.2. Файл має містити 30 кадрів на одну секунду.
   - 2.3. Розмір кадрів 320x240.

[1] **Робота виконана в рамках ДНТП України "Образний комп'ютер"**

3. З отриманим у п.2 AVI-файлом робляться наступні дії:

3.1. За допомогою програми відео монтажу (ADOBE PREMIER, AVI Constructor, тощо) вибираються з AVI-файлу послідовності кадрів на яких зображено процес промовляння конкретних фонем.

3.2. Вибрані послідовності кадрів запам'ятовуються у BMP-файлах.

3.3. Послідовності BMP-файлів, на яких зображене промовляння конкретної фонеми (трифону), записуються у директорії, назви яких містять назви трифонів.

4. Інформація, створена у п.3, переноситься, за допомогою спеціального програмного забезпечення у реляційну базу даних. У цій базі даних кожній фонемі (трифону) співставленні послідовності кадрів, на яких зображено процес промовляння. База даних (реалізована у ***Paradox***) складається з двох таблиць (таблиці по полю ***Id_Phonemes*** зв'язані як ***Master → Detail***) (Рис. 2).

5. Використовуючи створену у п.4 базу даних, моделюється процес промовляння, як послідовність фонем (трифонів) із транскрипції та відповідні їм послідовності кадрів (із бази даних), які міняються відповідно до тривалостей звучання конкретних фонем.

6. Описаним у пунктах 1-4 способом створюються бази даних з різними дикторами. Користувач має можливість, вказуючи шлях до конкретної бази даних, вибирати диктора для промовляння тексту.

Таблиця ***Phonemes*** (***Master***-таблиця):

№	*Field Name*	*Type*	*Size*	*Key*
1	Id_Phonemes	+		*
2	Phoneme	A	1	***Secondary index***
3	BeforePhoneme	A	1	
4	AfterPhoneme	A	1	
5	Duration	I		

Таблиця ***Pictures*** (***Detail***-таблиця):

№	*Field Name*	*Type*	*Size*	*Key*
1.	Id_Picture	+		*
2.	Id_Phoneme	I		
3.	Item	I		
4.	Picture	G		

Рис. 2

Використовуючи парадигму об'єктно-орієнтованого проектування [13], створюється абстрактний клас TTalkingFace, який і буде моделювати процес промовляння (п.5). До основних полів класу відносяться поля для роботи з базою даних [15-17]:

- база даних (DB);
- джерела даних (DataSourse);
- реляційні таблиці (Table);
- візуальний проглядач графічної інформації (DBImage) .

Основною подією, за якою працюватиме алгоритм візуалізації буде подія від таймеру, тобто подія, яка сигналізуватиме про закінчення заданого інтервалу часу (тривалість звучання конкретної фонеми).

Опишемо основні методи абстрактного класу TtalkingFace.

Метод TimerTimer полягає в наступному:

- Вимикається таймер.
- Запам'ятовується початковий час. В залежності від значення поля situation (1 чи 2) викликаються, відповідно методи Situation1 або Situation2.
- Запам'ятовується кінцевий час.
- Розраховується інтервал для "спання" таймеру на основі витраченого часу (кінцевий час мінус початковий час).
- Вмикається таймер.

Метод Situation1 реалізує алгоритм:

- збільшується лічильник поточної фонеми.
- якщо дійшли до останньої фонеми, то – кінець.
- будується трифон (до поточної фонеми приєднується попередня та наступна фонеми).
- шукається в базі даних рядок з поточним трифоном.
- визначається кількість кадрів для візуалізації фонеми (із бази даних).
- відповідно до вхідної тривалості фонеми та кількості кадрів – розраховується керуючий вектор тривалостей візуалізації кадрів.
- Встановлюється situation=2.

Метод Situation2: використовуючи керуючий вектор тривалостей візуалізації кадрів:

- якщо ще є кадри для даної фонеми, то стаємо на наступний кадр у таблиці.
- якщо це останній кадр, то встановлюється situation=1.

Використовуючи описаний вище абстрактний клас, реалізований у вигляді ActiveX-об'єкту, було створене програмне забезпечення для тестування запропонованого алгоритму. Для тестування бралось моделюватися промовляння фрази: "Добрий день, Україно!". Для створення бази даних, на відео була записана навчальна вибірка фраз, з яких можна було б вибирати фонеми для моделювання. Вибірка складалася з наступних слів:

Слово для відеозапису	*Трифон для бази даних*		
	перед	*фонема*	*після*
Дорога	#	д	о
Здоба	д	о	б
Обрій	о	б	р
...			
Гаї	а	й	і
Кіно	і	н	о
Кіно	н	о	#

Трифони, з відповідними відеокадрами промовляння фонем, були занесені у базу даних

(Рис.3). У відповідній програмі (з реалізованим об'єктом TalkingFace (Рис.4), був змодельований процес візуалізації промовляння фрази "Добрий день, Україно!".

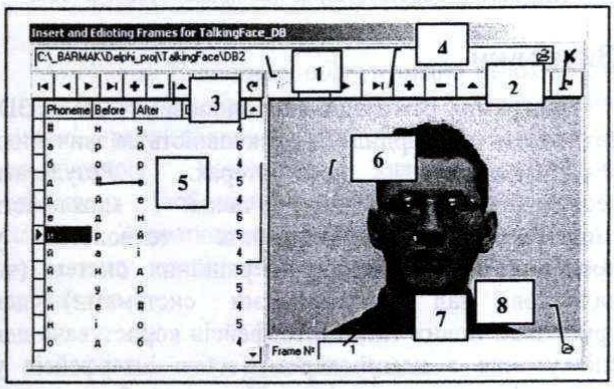


Рис. 3

1. *Вибір бази даних з образами дикторів;*
2. *Вихід з програми;*
3. *Навігатор для роботи з таблицею **Phonemes;***
4. *Навігатор для роботи з таблицею **Pictures;***
5. *Таблиця **Phonemes;***
6. *Поле **Picture** із таблиці **Pictures;***
7. *Поле **Item** із таблиці **Pictures;***
8. *Виведений на кнопку метод **LoadFromFile** для завантаження кадру з BMP-файлу у поле **Picture** таблиці **Pictures;***

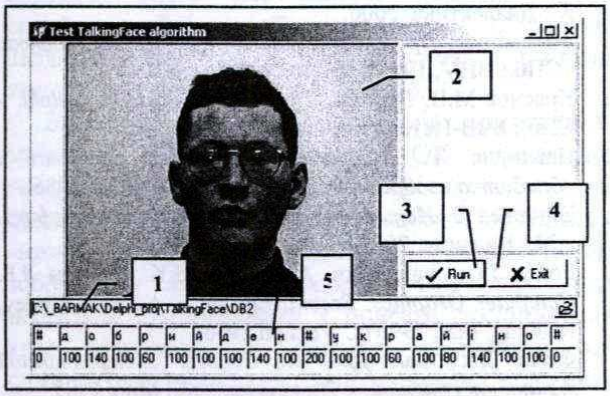


Рис.4

1. *Вибір шляху до бази даних з записами фонем та візем.*
2. *ActiveX об'єкт **TalkingFace**.*
3. *Запуск візуалізації промовляння фрази з тривалостями.*
4. *Завершення роботи з програмою.*
5. *Панель для завдання фраз та тривалостей фонем.*

2. 3D-технологія

Використовуючи отримані вище результати, створимо технологію та алгоритмічну реалізацію візуалізації промовляння на основі 3D-моделей. Створивши абстрактний клас, у якому реалізований 2D спосіб, та, на його основі, об'єкт, розроблено метод, за допомогою якого перехід до 3D способу буде полягати лише у переписуванні відповідних методів об'єкту-нащадку. Виходячи з цього розглянемо дії, необхідні для реалізації методів 3D-анімації візуалізації промовляння:

1. У пакеті програм трьохмірного моделювання 3D Studio MAX [2] створимо 3D-модель людської голови з морферами, які дозволятимуть керувати мімікою обличчя:
   - 1.1. Беручи за зразок кадри з бази даних на яких зображено процес промовляння трифонів, методом клонування та морфінгу створюється набір 3D-моделей, які їм відповідають.
   - 1.2. Нанесемо на отриманні моделі у якості текстур відповідні їм кадри з бази даних 2D-технології.
   - 1.3. Зробимо експорт отриманих моделей у файли з форматом ASE (ASCII scene export).
   - 1.4. Послідовності ASE-файлів та відповідних їм BMP-файлів (з текстурами) запишемо у директорії, назви яких містять назви трифонів.

2. На основі об'єктів програми 2D-технології розроблено нове програмне забезпечення для роботи з базою даних, яка зберігатиме трифони та відповідні їм моделі. Інформація, створена у попередньому пункті (1.4), переноситься, з допомогою цього програмного забезпечення у реляційну базу даних. У цій базі даних кожній фонемі (трифону) співставленні послідовності 3D-моделей, які зображають процес промовляння. На відміну від бази даних для 2D-технології, у новій базі даних будуть зберігатися не графічні образи, а інформація для інтерактивної побудови 3D-моделі: масив координат вершин трикутників, масив нормалей у кожній вершині, масив текстурних координат для кожної вершини трикутника, масив пікселів самої текстури. Структура таблиці ***Models*** (яка відповідає таблиці ***Pictures*** у 2D-технології) представлена на рис. 5.

Таблиця ***Models*** (***Detail***-таблиця):

№	*Field Name*	*Type*	*Size*	*Key*
1	Id_Model	+		*
2	Id_Phoneme	I		
3	Item	I		
4	Count_Triangles	I		
5	Vertexes	B		
6	Normals	B		
7	Textures	B		
8	Pictures	B		

Рис. 5

3. Використовуючи створену у п.2 базу даних та абстрактний клас TTalkingFace моделюється процес промовляння, як послідовність фонем (трифонів) із транскрипції та відповідні їм послідовності моделей (із бази даних), які міняються відповідно до тривалостей звучання конкретних фонем. У об'єкті, створеному на базі абстрактного класу TTalkingFace

переписується метод роботи з базою даних. Замість візуалізації кадру по події наступного запису у базі даних, створюється метод, який буде по цій події робити рендерінг 3D-моделі, тобто 2D-візуалізацію абстрактної сцени, існуючої у вигляді масивів вершин, масивів нормалей у цих вершинах та масиву текстурних координат.

4. Описаним у пунктах 1-2 способом створюються бази даних з різними моделями-дикторами. Користувач має можливість, вказуючи шлях до конкретної бази даних, вибирати певного диктора для промовляння тексту.

Використовуючи абстрактний клас TTalkingFace з методом для рендерінгу 3D-моделей було розроблене програмне забезпечення для реалізації промовляння по 3D-технології (п. 3). Для моделювання бралася таж сама тестова фраза "Добрий день, Україно!", що і в 2D-технології. Була створена база даних з трифонами і відповідними 3D-моделями для можливості промовляння цієї фрази (Рис.6).

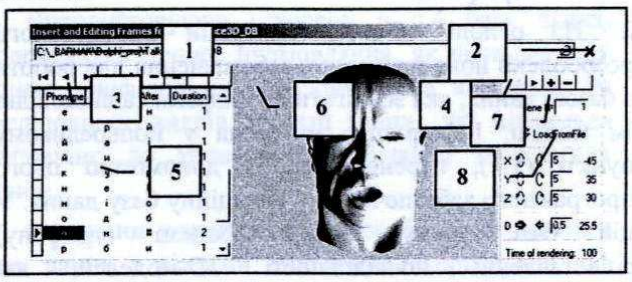


Рис. 6

1. *Вибір бази даних з образами дикторів;*
2. *Вихід з програми;*
3. *Навігатор для роботи з таблицею **Phonemes;***
4. *Навігатор для роботи з таблицею **Models;***
5. *Таблиця **Phonemes;***
6. *3D-двигун;*
7. *Поле **Item** із таблиці **Models;***
8. *Виведений на кнопку метод **LoadFromFile** для завантаження моделі з ASC-файлу.*

У відповідній програмі був змодельований процес візуалізації промовляння фрази "Добрий день, Україно!" (див.Рис.7).

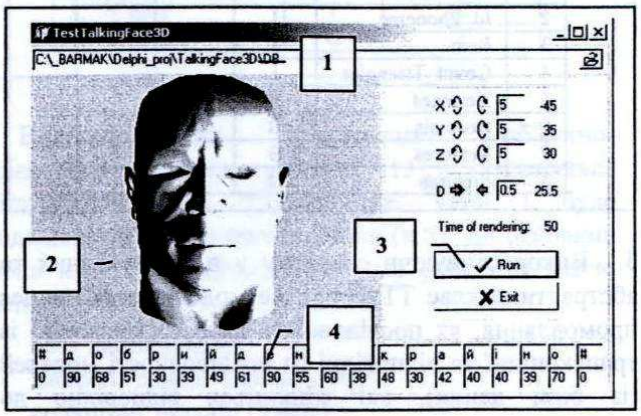


Рис. 7

1. *Вибір шляху до бази даних з записами моделей.*
2. *Панель для рендерінгу.*
3. *Запуск візуалізації промовляння фрази з тривалостями*
4. *Панель для завдання фраз та тривалостей фонем.*

Висновки

Програмна реалізація запропонованих 2D та 3D-технологій підтвердила їх ефективність на звичайних мультимедійних комп'ютерах. Результати тестування показали також можливість імплементації запропонованих технологій у програмне забезпечення операційних систем (чи надбудов над операційними системами) для організації нових типів інтерфейсів користувача для спілкування з комп'ютером (тобто інтерфейсів у яких спілкування користувача з комп'ютером та комп'ютера з користувачем відбувається з допомогою звичайної мови). Подальший розвиток таких технологій полягає в розробці нових математичних методів для побудови об'ємних зображень голови людини і створення якісних синтезаторів української мови.

Література

1. Флеминг Б., Доббс Д. *Методы анимации лица. Мимика и артикуляция*. Пер. с англ. – М.: ДКМ Пресс, 2002 .
2. Мердок К. *3D Studio MAX R3. Библия пользователя.* – К.: Диалектика, 2000.
3. Тихомиров. *Программирование трехмерной графики.* –СПб.: BHV, 1998
4. Краснов М.В. *OpenGL. Графика в проектах Delphi.* – СПб.: БЧВ-Петербург, 2001.
5. Павлидис Т. *Алгоритмы машинной графики и обработка изображений.* – М.: Радио и связь, 1986.
6. Эйнджел Э. *Интерактивная компьютерная графика.* – М.: Вильямс, 2001.
7. Foley J.D., van Dam A., Feiner S.K., Hughes J.F. *Computer Graphics, Second Edition.* – Addison-Wesley, Reading, MA, 1990 (C Version 1996).
8. Фоли Дж., ван Дэм А. *Основы интерактивной машинной графики: в 2-х книгах.* – М.: Мир, 1985
9. Pandzic I.S., Ostermann J., Millen D. *User evaluation: Synthetic talking faces for interactive services*. The Visual Computer. 15, 1999. – pp. 330-340.
10. Т.К. Винцюк. *Анализ, распознавание и смысловая интерпретация речевых сигналов.* – Киев: Наукова думка, 1987.
11. Дж.Л.Фланаган. *Анализ, синтез и восприятие речи.* Пер. с англ. М.:Связь, 1968
12. Каханер Д., Моулер К., Нэш С. *Численные методы и программное обеспечение.* – М.: Мир, 2001
13. Г.Буч. *Объектно-ориентированный анализ и проектирование.* – М., Бином, 1998, 558 с.
14. Миллер Т., Пауэл Д. *Использование DELPHI 3.* – К.: Диалектика, 1997.
15. Дж.Мартин. *Организация баз данных в вычислительных системах.* – М., Мир, 1980. 662 с.
16. К.Дейт. *Введение в системы баз данных.* – М., Наука, 1980.
17. Д.Мейер. *Теория реляционных баз данных.* - М., Мир, 1987. 608 с.

ПІДХОДИ ДО ОПИСУ І СИНТЕЗУ СИМЕТРИЧНИХ ЗОБРАЖЕНЬ

Олег Березький

ТАНГ, 46016, м.Тернопіль, Львівська 11, тел. 33-08-30, e-mail ob@tanet.edu.te.ua

Анотація. В доповіді обгрунтовано необхідність опису і синтезу зображень, виходячи із генеративної моделі розпізнавання при побудові образних комп'ютерів. Розглянуто клас симетричних зображень (зображень-орнаментів) на смузі і площині Запропоновано мову опису, матричну і рекурсивну моделі синтезу складних симетричних зображень.

ВСТУП

Широке розповсюдження сучасних інформаційних технологій вимагає розробки нових методів відбору, оброблення та передачі інформації. Це приводить до необхідності комплексного розв'язання проблем інформаційних технологій та обчислювальної техніки.

Одним із перспективних напрямків розвитку обчислювальної техніки є створення принципово нових комп'ютерів [1], яким властиве образне мислення (розуміння людської мови, просторових сцен, зображень, рукописних текстів тощо). Комп'ютера, який має такі властивості, називають образним (ОК). ОК містить декілька каналів сприйняття інформації, серед яких є зоровий. Завдяки цьому каналу, ОК сприймає і розпізнає тексти, аналізує зображення і сцени.

В основі побудови алгоритмів розпізнавання зображень автори проекту ОК пропонують генеративну модель, яка полягає в початковому відтворенні зображень даного класу і подальшому автоматичному порівнянні досліджуваного зображення із згенерованим. Але для цього необхідно структурувати і описати досліджуваний клас зображень. У доповіді розглянуто клас симетричних зображень і запропоновано мову опису та алгоритми синтезу зображень даного класу.

Класи зображень, наділених симетричною структурою, широко представлені в природі, мистецтві та інших галузях людського життя. Важливе дослідження класів складних зображень, наділених симетричною структурою, належить наукам: математиці, фізиці, кристалографії, хімії, біології та ін. Особливо широкі класи таких зображень зустрічаються в науково-технічних розробках при побудові систем штучного інтелекту, в представленні, описі, обробці та розпізнаванні образів.

Ці класи зображень відображають великі обсяги різних природних об'єктів та реалізованих фізичних і технічних процесів - одновимірних, двовимірних і трьохвимірних. Прагнення якомога ширше охопити вивчення зображень, наділених складною структурою, привело до створення різних підходів, теорій опису, представлення, моделювання, синтезу різних класів об'єктів та процесів в науково-технічних розробках, пов'язаних з розпізнаванням образів та побудовою систем штучного інтелекту.

Велика кількість плоских зображень-орнаментів на площині описується сімнадцятьма групами, структури яких описуються двомірними федорівськими групами та на смузі відповідно сімома півторамірними групами [2].

Зображення-орнамент складається з ритмічно впорядкованих однакових елементів, тобто володіє певною симетричною структурою (рис.1) [3]. В її основі лежать наступні структурні складові: орнамент, підорнамент, рапорт та мінімальний рисунок. Для орнаменту і його складових буде дійсною схема, де N - кількість підорнаментів в орнаменті, K, ..., L - кількість рапортів в підорнаментах, m, p, - кількість мінімальних рисунків в рапорті.

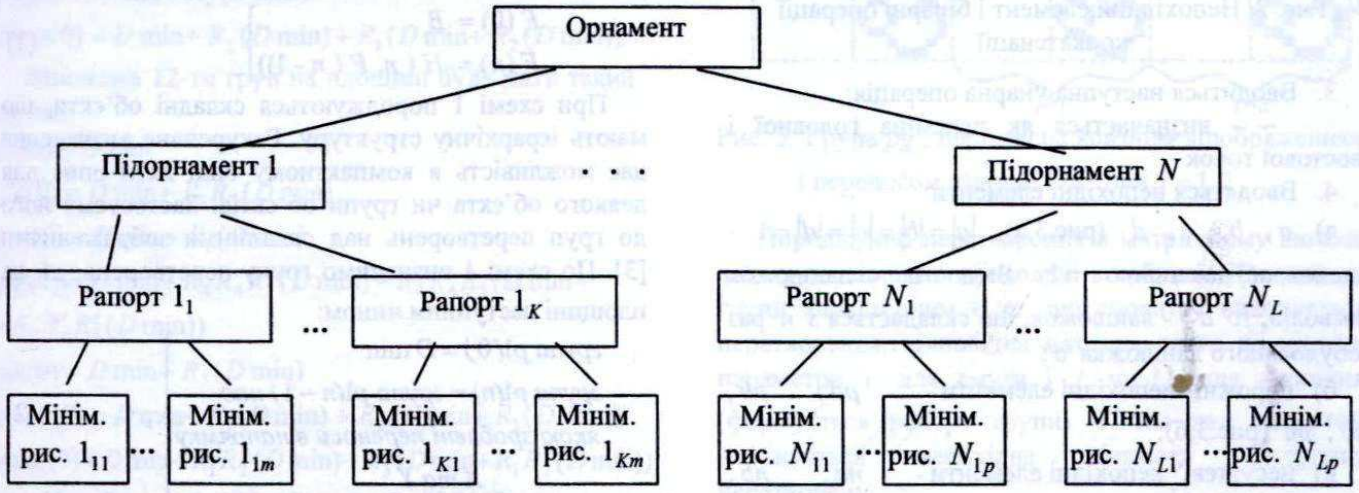


Рис. 1. Структура орнаменту

МОВА ОПИСУ ЗОБРАЖЕНЬ

Для представлення мінімального рисунку запропоновано мову опису зображень[3]

1. Кожен непохідний елемент помічається в двох різних точках - головній $г$ і хвостовій x (рис.2,а). Причому непохідні елементи дотикаються і накладаються тільки в головних чи хвостових точках.

2. Вводяться бінарні операції конкатенації (з'єднання):

- Операція $a+b$ (рис.2,б - головна точка "a" дотикається до хвостової точки "b").

- Операція $a \oplus b$ (рис.2,в - головна точка "a" співпадає з хвостовою точкою "b").

- Операція $a \times b$ (рис.2,г - хвостова точка "a" дотикається до хвостової точки "b").

- Операція $a \otimes b$ (рис.2,д - хвостова точка "a" співпадає з хвостовою точкою "b").

- Операція $a-b$ (рис.2,е - головна точка "a" дотикається до головної точки "b").

- Операція $a \Theta b$ (рис.2,є - головна точка "a" співпадає з головною точкою "b").

- Операція $a * b$ (рис.2,ж - головна точка "a" дотикається до головної точки "b" і хвостова точка "a" дотикається до хвостової точки "b".

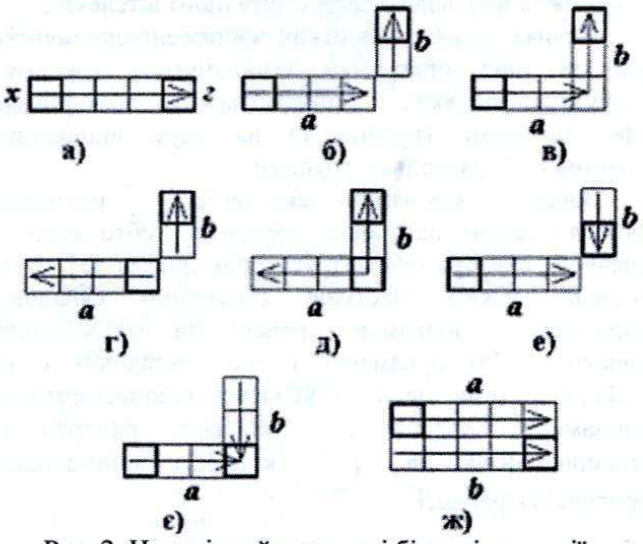


Рис. 2. Непохідний елемент і бінарні операції конкатенації

3. Вводиться наступна унарна операція:

~ - визначається як переміна головної і хвостової точок

4. Вводяться непохідні елементи:

а) a, b, c, d (рис.3,а). $|a|=|b|=|c|=|d|=1$ - ланцюжки довжиною $=1$. Якщо a - ланцюжок символів, то a^n - ланцюжок, що складається з n раз побудованого ланцюжка a;

б) "порожні" непохідні елементи - pa, pb, pc, pd (рис.3,б);

в) "несуттєві" непохідні елементи - na, nb, nc, nd (рис.3,в)

Непохідні елементи "порожній" і "несуттєвий" з'єднують непохідні елементи, що не пересікаються і досить корисні для опису простих геометричних зв'язків. "Порожній" використовується при обривах і є з'єднуючим "простором" між образами. Коли необхідно описати зв'язок між непохідними елементами, що не перетинаються, а вони є розділені іншими об'єктами, то ці останні об'єкти визначають як несуттєві елементи.

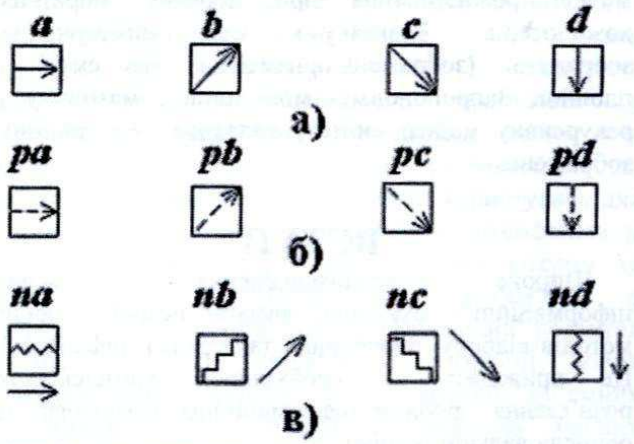


Рис. 3. Непохідні елементи: а)основні; б)«порожні»; в)«несуттєві»

РЕКУРСИВНА МОДЕЛЬ

Для отримання рекурсивних моделей використаємо рекурсивну схему [4]:

$$\left.\begin{aligned} f(0) &= b \\ f(n) &= h(n, f(n-1)) \end{aligned}\right\}$$

де b - натуральне число; f, h - функції, значення яких також є натуральними числами.

Побудуємо схему не тільки для числових функцій, а й для об'єктів іншої природи. Нехай B - довільний об'єкт класу Ψ, $B \in \Psi$, F - функція чи оператор, аргументом якої є натуральне число, а результатом - об'єкт з Ψ; H - функція 2-х аргументів натурального числа і об'єкту з Ψ, а результатом - об'єкт з Ψ, тоді рекурсивним визначенням F на класі Ψ буде наступне:

$$\left.\begin{aligned} F(0) &= B \\ F(n) &= H(n, F(n-1)) \end{aligned}\right\} \qquad (1)$$

При схемі 1 породжуються складні об'єкти, що мають ієрархічну структуру. Рекурсивне визначення дає можливість в компактному виді дати опис для деякого об'єкта чи групи об'єктів. Застосуємо його до груп перетворень над складними зображеннями [3]. По схемі 1 визначимо групу перетворень $p1$ на площині наступним чином:

$$\left.\begin{aligned} &\text{група } p1(0) = D\min \\ &\text{група } p1(n) = \text{група } p1(n-1) \text{ над} \\ &\text{якою зроблені переноси в напрямку} \\ &\qquad X \text{ та } Y \end{aligned}\right\}, \qquad (2)$$

а групи смуги таким чином:

$$\left.\begin{array}{l} \text{група } p1(0) = D\min \\ \text{група } p1(n) = \text{група } p1(n-1) \text{ над} \\ \text{якою зроблено перенос в напрямку } X \end{array}\right\} \quad (3)$$

Оператор H (комбінація відображень (рис.4) R_1R_1' та R_2R_2' - у схемі 2 і R_1R_1' - у схемі 3 дає можливість за один такт рекурсії на n-му кроці проходити декілька елементарних змін в об'єкті $F(n-1)$.

Приведемо кожну з груп площини і смуги до схем 2 і 3. Тобто покажемо, що кожна з них є підгрупою $p1$, домовившись, що знак "+" означає об'єднання зображень, R_1, R_1', R_2, R_2', R_3, R_3', R_4, і R_4', (рис. 4) - відображення у сторонах і діагоналях прямокутника і паралельних до діагоналей прямих, $D\min$ - мінімальний рисунок. Надалі приводимо лише першу тотожність рекурсивних схем (рапорт), друга буде для всіх груп аналогічною.

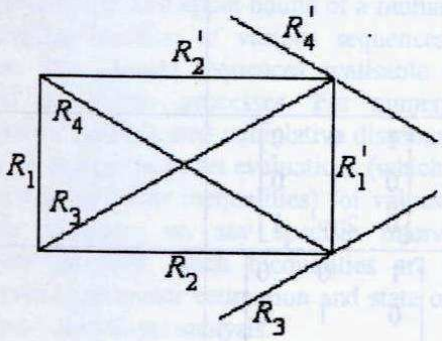


Рис. 4. Представлення осей відображень

Тоді множина 7-и груп на смузі буде представлятися таким чином:

$p1(0) = D\min$

$pg(0) = D\min + R_1R_1'R_2(D\min)$

$p1m(0) = D\min + R_1(D\min)$

$p2(0) = D\min + R_1R_2(D\min)$

$pmg(0) = D\min + R_1(D\min) + R_1R_2(D\min + R_1(D\min))$

$pm(0) = D\min + R_2(D\min)$

$pmm(0) = D\min + R_2(D\min) + R_1(D\min + R_2(D\min))$.

Множина 12-ти груп на площині буде мати такий вигляд:

$p1(0) = D\min$

$p2(0) = D\min + R_1R_2(D\min)$

$pm(0) = D\min + R_1(D\min)$

$pg(0) = D\min + R_4R_4'R_3'(D\min) + R_4R_4'R_3(D\min +$
$+R_4R_4'R_3'(D\min))$

$cm(0) = D\min + R_3(D\min)$

$pmm(0) = D\min + R_1(D\min) + R_2(D\min + R_1(D\min))$

$pmg(0) = D\min + R_1R_2(D\min) + R_3(D\min + R_1R_2(D\min))$

$pgg(0) = D\min + R_1R_1'R_2(D\min) + R_2R_2'R_1'(D\min +$
$+R_1R_1'R_2(D\min))$

$cmm(0) = D\min + R_3(D\min) + R_4(D\min + R_3(D\min))$

$p4(0) = D\min + R_3R_2'(D\min) + R_1R_2(D\min +$
$+R_3R_2'(D\min))$

$p4m(0) = D\min + R_3(D\min) + R_1(D\min + R_3(D\min)) +$
$+R_2(D\min + R_3(D\min) + R_1(D\min + R_3(D\min)))$

$p4g(0) = D\min + R_3(D\min) + R_3R_2'(D\min +$
$+R_3(D\min)) + R_1R_2(D\min + R_3(D\min) + R_3R_2'(D\min +$
$+R_3(D\min)))$.

МАТРИЧНА МОДЕЛЬ

Матриця перетворень порядку 3×3 в загальному випадку, має вигляд $\begin{bmatrix} a & b & p \\ c & d & q \\ m & n & s \end{bmatrix}$, де a, b, c і d здійснюють відповідно відображення, поворот; m і n виконують зміщення, а p і q - одержання проекцій. Елемент s проводить повну зміну масштабу. Розглянемо групу pg, яка формується здійсненням ковзного відображення (його проводимо у 2 етапи: 1) відображення відносно OX; 2) перенос на відстань x') [3]:

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ x' & 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ x' & 0 & 1 \end{bmatrix}$$

і необхідної кількості, яка задається параметром i, переносів цих зображень вздовж осі OX (рис.5):

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 2ix' & 0 & 1 \end{bmatrix}.$$

Параметр i приймає цілі значення.

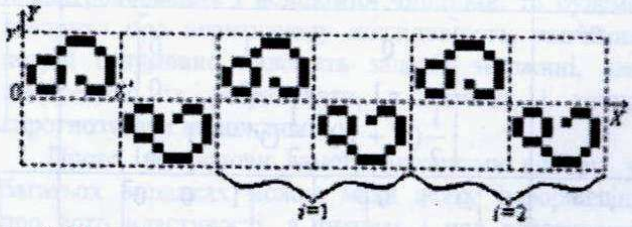


Рис. 5. Група pg, породжена ковзним відображенням і переносом вздовж OX при $i = 1, 2$.

Породжуючі перетворення в матричному вигляді для груп площини приведені в таблиці. Для кожної групи характерним є те, що спочатку виконується перетворення, відповідна матриця якого не містить параметра i для смуги і i чи j для площини (формується рапорт групи). Потім над рапортом виконується необхідна кількість наступних перетворень.

Назва групи	Породжуючі перетворення в матричному вигляді				
	I	II	III	IV	V
$p1$	$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ ix' & 0 & 1 \end{bmatrix}$	$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & jy' & 1 \end{bmatrix}$			
$p2$	$\begin{bmatrix} -1 & 0 & 0 \\ 0 & -1 & 0 \\ x' & y' & 1 \end{bmatrix}$	$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ ix' & 0 & 1 \end{bmatrix}$	$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & jy' & 1 \end{bmatrix}$		
pm	$\begin{bmatrix} -1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$	$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 2ix' & 0 & 1 \end{bmatrix}$	$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & jy' & 1 \end{bmatrix}$		
pg	$\begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ x' & \frac{1}{2}(y'-x') & 1 \end{bmatrix}$	$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ ix' & iy' & 1 \end{bmatrix}$	$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ jx' & -jy' & 1 \end{bmatrix}$		
cm	$\begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}$	$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ ix' & 0 & 1 \end{bmatrix}$	$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & jy' & 1 \end{bmatrix}$		
pmm	$\begin{bmatrix} -1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$	$\begin{bmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$	$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 2ix' & 0 & 1 \end{bmatrix}$	$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 2jy' & 1 \end{bmatrix}$	
pmg	$\begin{bmatrix} -1 & 0 & 0 \\ 0 & -1 & 0 \\ 2x' & y' & 1 \end{bmatrix}$	$\begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}$	$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ ix' & iy' & 1 \end{bmatrix}$	$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ jx' & -jy' & 1 \end{bmatrix}$	
pgg	$\begin{bmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ x' & y' & 1 \end{bmatrix}$	$\begin{bmatrix} -1 & 0 & 0 \\ 0 & -1 & 0 \\ 2x' & 2y' & 1 \end{bmatrix}$	$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 2ix' & 0 & 1 \end{bmatrix}$	$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 2jy' & 1 \end{bmatrix}$	
cmm	$\begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}$	$\begin{bmatrix} 0 & -1 & 0 \\ -1 & 0 & 0 \\ y' & y' & 1 \end{bmatrix}$	$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ ix' & 0 & 1 \end{bmatrix}$	$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & jy' & 1 \end{bmatrix}$	
$p4$	$\begin{bmatrix} 0 & 1 & 0 \\ -1 & 0 & 0 \\ \frac{1}{2}(x'+y') & \frac{1}{2}(y'-x') & 1 \end{bmatrix}$		$\begin{bmatrix} -1 & 0 & 0 \\ 0 & -1 & 0 \\ x' & y' & 1 \end{bmatrix}$	$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ ix' & 0 & 1 \end{bmatrix}$	$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & jy' & 1 \end{bmatrix}$
$p4m$	$\begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}$	$\begin{bmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$	$\begin{bmatrix} -1 & 0 & 0 \\ 0 & 1 & 0 \\ 2x' & 0 & 1 \end{bmatrix}$	$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 2ix' & 0 & 1 \end{bmatrix}$	$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 2jy' & 1 \end{bmatrix}$

ВИСНОВОК

У роботі проведено структуризацію, опис та синтез класу симетричних зображень-орнаментів.

ЛІТЕРАТУРА

1. Вінцюк Т.К. Образний комп'ютер: концепції, методологія, підходи // Оптико-електронні інформаційно-енергетичні технології. - 2001. - №1. - С.125-139.
2. Федоров Е.С. Симметрия и структура кристаллов: Основные работы. - М.: Изд-во Акад. наук СССР, 1949. - 631 с.
3. Грицик В.В., Березька К.М., Березький О.М. Моделювання та синтез складних симетричних зображень // Інформаційні технології і системи. - 1999. - Т.2., №1. - С. 84-100.
4. Мальцев А.И. Алгоритмы и рекурсивные функции. - М.: Наука, 1986. - 367 с.

ХАОТИЧНІ ПОДІЇ, ПРОЦЕСИ ТА ЇХ ХАРАКТЕРИСТИКИ

Михайло Личак
Інститут космічних досліджень НАН та НКА України (ІКД НАНУ-НКАУ)
Проспект Академіка Глушкова, 40, Київ 03680, Україна
Тел.: +380 44 2661291 Факс: +380 44 2664124
adapt@space.is.kiev.ua

ABSTRACT

M. Lychak. Chaotic events, processes and their characteristics. The phenomena and experiments with uncertain results at their recurrence are considered. The new approach is constructed on the basis of the set theory and concept chaotic algorithm of selection of elements among the set. The concepts of chaotic events and chaotic sequences are introduced, and there is a lower and upper edge of an own cumulative distribution function of sequence of chaotic events on elementary events set lower and upper bound of a mutual cumulative distribution function of various sequences of chaotic events. The chaotic sequences realisable in time are named as chaotic processes. For numerical chaotic sequences, the indicated cumulative distribution function is shown to determine set evaluations (which are solution the system of linear inequalities) for values of terms of chaotic sequence on any specific interval of their adjacent numbers. Such inequalities are used for a guaranteed parameter estimation and state of objects on the basis interval-set analysis.

The computer simulation of different types of chaotic sequences is conducted and their characteristics are experimentally determined.

ВСТУП

Розглядаються явища, досліди і експерименти, результати яких є неоднозначними при повторюваності, хоча їх параметри і зовнішні фактори при цьому є незмінними. Більшість дослідників застосовують імовірнісні моделі невизначеності. З іншого боку, почав інтенсивно розвиватися так званий інтервальний аналіз, де використовується припущення, що самі чисельні результати повністю невизначені, а відомі лише деякі інтервали, яким вони безумовно належать.

Пропонується підхід до побудови математичної моделі таких подій на основі теорії множин і понятті алгоритму хаотичного вибору елементів множини. Враховуючи якісну відмінність від імовірнісного підходу, говориться про хаотичні події і процеси. Виділено клас хаотичних подій, за своїми характеристиками близьких до випадкових подій, Однак при цьому ніде не використовується поняття ймовірності і не вводиться що-небудь аналогічне. Більше того, цей клас охоплює встановлені коливання в режимі "хаосу" в нестійких складних нелінійних динамічних системах.

Закладаються теоретичні основи використання множинного (нестохастичного) підходу до рішення задач оцінювання, розпізнавання та керування.

1. ХАОТИЧНІ ПОДІЇ ТА ЇХ ВЛАСНІ ХАРАКТЕРИСТИКИ

Будемо вважати, що можливі прояви досліджуваного явища, результати розглядуваного досліду або спостереження утворюють деяку замкнуту обмежену множину елементарних подій X_0. Реалізація того чи іншого прояву, результату або спостереження є елемент x цієї множини

$$x \in X_0 . \qquad (1)$$

Зокрема, такою множиною може бути деякий скінчений набір раціональних чисел, замкнутий інтервал або об'єднання деякого числа замкнутих інтервалів на числовій вісі.

Якщо явище, дослід чи спостереження реалізується багато разів при збереженні основних факторів або умов, то цьому буде відповідати деяка послідовність $x_1, x_2, \ldots, x_n$ (n - поточний номер повторення) елементів множини X_0, які вибираються з неї згідно деякому алгоритму вибору.

Коли цей алгоритм вибору відомий, то говорять про детерміновану послідовність, для якої всі значення можуть бути наперед розраховані, а значить спрогнозировані відповідно даному алгоритму.

Якщо ж такий алгоритм вибору невідомий, а вибір елементів послідовності x_n $(n = 1,2,\ldots)$ з множини X_0 здійснюється під впливом неконтрольованих і невідомих чинників, то будемо говорити про невизначену послідовність, значення котрої безумовно належать заданій множині, але попередньо їх розрахувати, а значить і точно спрогнозувати неможливо.

Проте, не знаючи самого алгоритму вибору, у багатьох випадках можна мати деяку інформацію про його властивості, а значить і про властивості послідовності x_n $(n = 1,2,\ldots)$, що формується ним.

Введемо для цієї послідовності поняття інтервалу зв'язаних елементів $[n, n+S-1]$, де ціле число S задає кількість сусідніх членів послідовності на цьому інтервалі, коли на елементи, що їм відповідають, накладаються деякі загальні обмеження.

Позначимо

$$X_n^T = (x_n; \ldots; x_{n+S-1}), \qquad (2)$$

деякий вектор, що характеризує «стан» послідовності, яка аналізується на інтервалі «зв'язаності». Відповідно можна ввести

послідовність векторів $X_n^T\ (n=1,2,\ldots)$, що відображають зміну цього стану в процесі вибору.

Тоді

$$X_n \in \mathcal{X}_0^{(S)} \quad \forall \quad n=1,2,\ldots, \tag{3}$$

де обмежена множина $\mathcal{X}_0^{(S)}$ в просторі E^S складена з елементів, що є всіми можливими комбінаціями S компонент векторів X_n, зформованих шляхом їхнього вибору з $\mathcal{X}_0$.

Визначення 1. Хаотичними подіями на множині $\mathcal{X}_0$ будемо називати члени всякої послідовності $x_n\ \ (n=1,2,\ldots)$, утвореної таким вибором елементів множини $\mathcal{X}_0$, що для заданого інтервалу пов'язаних елементів $[n, n+S-1]$, де $S \geq 1$ деяке ціле число, вектор «стану» X_n із (2) для усіх $n=1,2,\ldots$ належить деякій обмеженій множині $\mathcal{X}_S$, тобто

$$X_n \in \mathcal{X}_S \quad \forall \quad n=1,2,\ldots, \tag{4}$$

що є підмножиною множини $\mathcal{X}_0^{(S)}$:

$$\mathcal{X}_S \subset \mathcal{X}_0^{(S)}. \tag{5}$$

Послідовності $x_n\ \ (n=1,2,\ldots)$, складені з хаотичних подій будемо називати хаотичними, а множину $\mathcal{X}_S$ з (4) будемо називати репроективною. Виконання (5) означає, що в цьому випадку алгоритм вибору елементів множини $\mathcal{X}_0$ забороняє частину можливих комбінацій з S елементів у якості компонент векторів X_n.

Нехай множина $\mathcal{X}_0$ є упорядкованою, тобто для будь-яких двох різних її елементів установлене правило, по якому один із цих елементів вважається попереднім другому. Позначимо $a<b$, якщо $a \in \mathcal{X}_0$ і $b \in \mathcal{X}_0$, але a є елементом $\mathcal{X}_0$, що передує b. Крім того, справедливі умови:

1. якщо $a<b$, або $b>a$, то не має місця $b<a$ або $a>b$ (асиметричність);
2. якщо $a<b$ або $b>a$ і $b<c$ або $c>b$, то $a<c$ або $c>a$ (транзитивність).

Позначення $a \leq b$ або $b \geq a$ означає, що b може збігатися з a

Будь-яка множина дійсних чисел, наприклад, упорядкована по їх величині.

На упорядкованій множині $\mathcal{X}_0$ введемо таку функцію від $x \in \mathcal{X}_0$ і членів будь-якої невизначеної послідовності $x_n\ \ (n=1,2,\ldots)$, що

$$F(x,x_n) \equiv \begin{cases} 1 & при \quad x > x_n, \\ 0 & при \quad x \leq x_n. \end{cases} \tag{6}$$

Визначення 2. Хаотичними подіями L - типу на упорядкованій множині $\mathcal{X}_0$ будемо називати хаотичні події, для яких існують деякі функції $P_н(x,N) \geq 0$ і $P_в(x,N) \leq 1$, де $x \in \mathcal{X}_0$, $N=1,2,\ldots$, – послідовні значення довжини інтервалів пов'язаних елементів, що для таких членів будь-якої хаотичної послідовності $x_n\ \ (n=1,2,\ldots)$ справедлива система нерівностей

$$P_н(x,N) \leq \frac{1}{N}\sum_{i=0}^{N-1} F(x,x_{n+i}) \leq P_в(x,N)\ \forall\ n=1,2,\ldots,$$
$$N=1,2,\ldots. \tag{7}$$

Функції $P_н(x,N)$ і $P_в(x,N)$ будемо називати відповідно нижньою і верхньою гранями функції розподілу хаотичних подій L - типу на множині елементарних подій $\mathcal{X}_0$.

Наслідок 1 (із визначення 2). Якщо для всіх $x \in \mathcal{X}_0$ існує

$$\lim_{N\to\infty} P_н(x,N) = \lim_{N\to\infty} P_в(x,N) = P(x), \tag{8}$$

то функція розподілу хаотичних подій $P(x)$ буде відповідати функції розподілу випадкових подій на множині елементарних подій $\mathcal{X}_0$.

2. МНОЖИННІ ОЦІНКИ ЗНАЧЕНЬ ЧИСЛОВИХ ХАОТИЧНИХ ПОДІЙ

Нехай множина елементарних подій $\mathcal{X}_0$ має вигляд замкнутого інтервалу на числовій осі $[-\Delta;\Delta]$, де $\Delta>0$ - деяке число, тобто будь-яка послідовність $x_n\ \ (n=1,2,\ldots)$, що вибирається з цієї множини задовільняє умову

$$|x_n| \leq \Delta = const \quad \forall \quad n=1,2,3,\ldots\ . \tag{9}$$

Розглянемо тепер випадок хаотичної послідовності, для якої виконуються нерівності виду (7).

Вирахуємо

$$\int_{-\Delta}^{\Delta} F(x,x_n)dx = \int_{x_n}^{\Delta} dx = \Delta - x_n, \tag{10}$$

звідки

$$\int_{-\Delta}^{\Delta}[\frac{1}{N}\sum_{i=0}^{N-1} F(x,x_{n+i})]dx = \frac{1}{N}\sum_{i=0}^{N-1}\int_{-\Delta}^{\Delta} F(x,x_{n+i})dx =$$
$$= \frac{1}{N}(\Delta\cdot N - \sum_{i=0}^{N-1} x_{n+i}) = \Delta - \frac{1}{N}\sum_{i=0}^{N-1} x_{n+i}\ .$$

Проінтегровані нерівності виду (7) для цього випадку набудуть вигляду

$$\int_{-\Delta}^{\Delta} P_н(x,N)dx \leq \Delta - \frac{1}{N}\sum_{i=0}^{N-1} x_{n+i} \leq \int_{-\Delta}^{\Delta} P_в(x,N)dx, \tag{11}$$

які можна переписати в вигляді системи нерівностей

$$m_н(N) \leq \frac{1}{N}\sum_{i=0}^{N-1} x_{n+i} \leq m_в(N) \tag{12}$$

$\forall n=1,2,\ldots,\quad N=1,2,\ldots,$

де

$$\begin{cases} m_н(N) = \Delta - \int\limits_{-\Delta}^{\Delta} P_в(x,N)dx, \\ m_в(N) = \Delta - \int\limits_{-\Delta}^{\Delta} P_н(x,N)dx. \end{cases} \quad (13)$$

Очевидно, що система лінійних нерівностей (12) конструктивним чином в явному виді виділяє репроективну множину X_S типу поліедра, а (13) установлюють взаємозв'язок між гранями функції розподілу хаотичних подій L - типу (функціями $P_н(\cdot)$ і $P_в(\cdot)$) і спільними обмеженнями на значення членів будь-якої частини хаотичної послідовності певної довжини.

Відзначимо, що якщо виконується (8) і функція $P(x)$ диференційована, то

$$\lim_{N\to\infty} m_н(N) = \lim_{N\to\infty} m_в(N) = m_0 = \int\limits_{-\Delta}^{\Delta} x\frac{dP(x)}{dx}dx, \quad (14)$$

де m_0 - математичне сподівання послідовності випадкових подій на множині X_0, що виділяється умовою (9).

Це випливає з того факту, що по визначенню математичного сподівання випадкової величини

$$m_0 = \int\limits_{-\Delta}^{\Delta} x\cdot\frac{dP(x)}{dx}dx = xP(x)\Big|_{-\Delta}^{\Delta} - \int\limits_{-\Delta}^{\Delta} P(x)dx =$$

$$= \Delta - \int\limits_{-\Delta}^{\Delta} P(x)dx. \quad (15)$$

3. ВЗАЄМНІ ХАРАКТЕРИСТИКИ ХАОТИЧНИХ ПОДІЙ

Розглянемо тепер той випадок, коли паралельно реалізовані два хаотичних алгоритми вибору елементів упорядкованої множини X_0. Відповідно отримуємо дві хаотичні послідовності: $x_n^{(1)}$ і $x_n^{(2)}$ $(n = 1,2,\ldots)$. Нехай це будуть хаотичні події L-типу, для яких існують нижні грані функції розподілу хаотичних подій $P_н^{(1)}(x,N_1)$ і $P_н^{(2)}(x,N_2)$, а також верхні грані - $P_в^{(1)}(x^{(1)},N_1)$ і $P_в^{(2)}(x^{(2)},N_2)$. Відповідно, якщо це числові послідовності, що задовільняють умову (9), то існують $m_н^{(1)}(N_1)$ і $m_н^{(2)}(N_2)$, а також $m_в^{(1)}(N_1)$ і $m_в^{(2)}(N_2)$, що обмежують середньоарифметичне значення будь-яких N_1 чи N_2 сусідніх членів вибраної хаотичної послідовності.

Проте, можна ще на упорядкованій множині X_0 ввести таку функцію від $x^{(1)} \in X_0$ і $x^{(2)} \in X_0$ та членів хаотичних послідовностей $x_n^{(1)}$ і $x_n^{(2)}$ $(n = 1,2,\ldots)$, що

$$F(x^{(1)},x^{(2)},x_n^{(1)},x_n^{(2)}) =$$

$$= \begin{cases} 1 \quad при \quad x^{(1)} > x_n^{(1)} \quad і \quad x^{(2)} > x_n^{(2)}, \\ 0 \quad при \quad x^{(1)} \le x_n^{(1)} \quad або \quad x^{(2)} \le x_n^{(2)}. \end{cases} \quad (16)$$

Визначення 3. Зв'язаними хаотичними подіями L - типу на упорядкованій множині X_0 будемо називати членів хаотичних послідовностей $x_n^{(1)}$ і $x_n^{(2)}$ $(n = 1,2,\ldots)$, для яких існують такі функції $P_н(x^{(1)},x^{(2)},N) \ge 0$ і $P_в(x^{(1)},x^{(2)},N) \le 1$, де $x^{(1)} \in X_0$, $x^{(2)} \in X_0$, числа $N = 1,2,\ldots$, що для них справедлива система нерівностей

$$P_н(x^{(1)},x^{(2)},N) \le \frac{1}{N}\sum_{i=0}^{N-1} F(x^{(1)},x^{(2)},x_{n+i}^{(1)},x_{n+i}^{(2)}) \le$$

$$\le P_в(x^{(1)},x^{(2)},N) \quad \forall \quad n = 1,2,\ldots. \quad (17)$$

Функції $P_н(\cdot)$ і $P_в(\cdot)$ будемо називати відповідно нижньою і верхньою гранями взаємної функції розподілу таких хаотичних подій.

Наслідок 1 (із визначення 3). Якщо для всіх $x^{(1)} \in X_0$ і $x^{(2)} \in X_0$ існує

$$\lim_{N\to\infty} P_н(x^{(1)},x^{(2)},N) = \lim_{N\to\infty} P_в(x^{(1)},x^{(2)},N) =$$

$$= P(x^{(1)},x^{(2)}), \quad (18)$$

то взаємна функція розподілу хаотичних подій $P(x^{(1)},x^{(2)})$ буде відповідати взаємної функції розподілу деяких двох випадкових подій на множині елементарних подій X_0.

За функцією $F(\cdot)$ у вигляді (16) для членів двох хаотичних послідовностей $x_n^{(1)}$ і $x_n^{(2)}$ можна визначити аналогічні функції у вигляді (6) від членів будь-якої з цих послідовностей:

$$\begin{cases} F(x^{(1)},x_n^{(1)}) = \lim\limits_{x^{(2)}\to+\infty} F(x^{(1)},x^{(2)},x_n^{(1)},x_n^{(2)}), \\ F(x^{(2)},x_n^{(2)}) = \lim\limits_{x^{(1)}\to+\infty} F(x^{(1)},x^{(2)},x_n^{(1)},x_n^{(2)}). \end{cases} \quad (19)$$

Відповідно

$$\begin{cases} P_н(x^{(1)},N) = \lim\limits_{x^{(2)}\to\infty} P_н(x^{(1)},x^{(2)},N), \\ P_в(x^{(1)},N) = \lim\limits_{x^{(2)}\to\infty} P_в(x^{(1)},x^{(2)},N). \end{cases} \quad (20)$$

Розглянемо тепер випадок, коли множина елементарних подій X_0 має вигляд замкнутого інтервалу на числовій осі $[-\Delta;\Delta]$, тобто хаотичні послідовності $x_n^{(1)}$ і $x_n^{(2)}$ задовільняють умові виду (9).

Вирахуємо подвійний інтеграл

$$\int\limits_{-\Delta}^{\Delta}\int\limits_{-\Delta}^{\Delta} F(x^{(1)},x^{(2)},x_n^{(1)},x_n^{(2)})dx^{(1)}dx^{(2)} =$$

$$= \int\limits_{x_n^{(1)}}^{\Delta} dx^{(1)}\cdot\int\limits_{x_n^{(2)}}^{\Delta} dx^{(2)} = (\Delta - x_n^{(1)})(\Delta - x_n^{(2)}). \quad (21)$$

Звідси

$$\int_{-\Delta}^{\Delta}\int_{-\Delta}^{\Delta}\left[\frac{1}{N}\sum_{i=0}^{N-1}F(x^{(1)},x^{(2)},x_{n+i}^{(1)},x_{n+i}^{(2)})\right]dx^{(1)}dx^{(2)} =$$

$$= \frac{1}{N}\sum_{i=0}^{N-1}(\Delta - x_{n+i}^{(1)})(\Delta - x_{n+i}^{(2)}). \qquad (22)$$

Подвійний інтеграл від нерівностей (17) дає наступну систему нерівностей виду

$$\int_{-\Delta}^{\Delta}\int_{-\Delta}^{\Delta}P_н(x^{(1)},x^{(2)},N)dx^{(1)}dx^{(2)} \le \frac{1}{N}\sum_{i=0}^{N-1}(\Delta - x_{n+i}^{(1)})(\Delta - x_{n+i}^{(2)}) \le$$

$$\le \int_{-\Delta}^{\Delta}\int_{-\Delta}^{\Delta}P_e(x^{(1)},x^{(2)},N)dx^{(1)}dx^{(2)} \ \forall\, n = 1,2,\ldots, N = 1,2,\ldots \qquad (23)$$

Ці нерівності встановлюють зв'язок між гранями взаємної функції розподілу і нелінійними обмеженнями на значення членів взаємозв'язаних хаотичних числових послідовностей.

4. ХАОТИЧНІ ДИСКРЕТНІ ПРОЦЕСИ

При розгляді членів хаотичної послідовності x_n $(n = 1,2,\ldots)$, передбачалось, що номер n визначає лише номер реалізації явища, досліду чи спостереження. Проте можна вважати, що реалізація відбувається в часі і кожному номеру відповідає момент часу t_n, тобто

$$x_n = x(t_n), \quad t_1 < t_2 < t_3 < \ldots, \qquad (24)$$

Визначення 4. Дискретну часову послідовність виду (24) із заданими моментами часу t_n $(n = 1,2,\ldots)$, де значення x_n складають хаотичну послідовність з деякими нижньою і верхньою гранями функції розподілу хаотичних подій L - типу на заданій множині елементарних подій X_0, будемо називати хаотичним дискретним процесом.

Найчастіше реалізація в часі хаотичних подій відбувається послідовно через рівний проміжок часу Δt, який називають періодом квантування за часом. Тоді

$$t_n = t_0 + n\Delta t, \quad n = 1,2,\ldots, \qquad (25)$$

де t_0 - деякий початковий відлік часу, з якого починається спостередження за хаотичною подією. Для такого хаотичного дискретного процесу знання характеристик виду (7) і величини Δt дозволяє в подальшому оцінювати його вплив на об'єкти та вимірювальні пристрої. Зокрема, коли x_n є числами, що задовольняють умову (9), то справедливі нерівності (12), які використовуються при рішенні задач оцінювання, розпізнавання та керування.

В той же час, існує відмінність хаотичних процесів від хаотичних подій, а саме, важливість залежності чи незалежності їх характеристик від початкового часу t_1 (з якого починається спостереження – вимірювання результатів певного явища чи досліду).

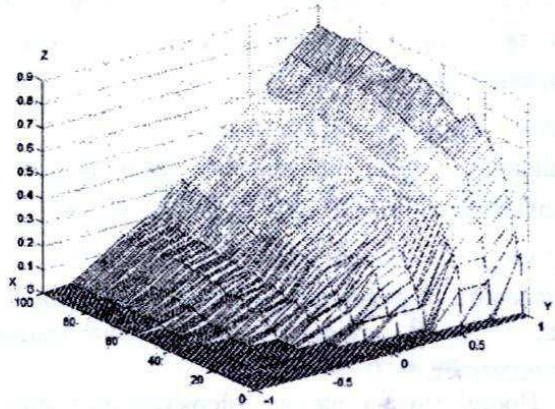
Визначення 5. Дискретний хаотичний процес виду (24) будемо називати стаціонарним, якщо значення нижньої і верхньої граней функції розподілу хаотичних подій x_n при будь-яких $N = 1,2,\ldots$ не залежать від початкового моменту часу t_1.

Тобто, якщо будь-яку кількість членів хаотичної послідовності x_n на початку буде вилучено і визначення системи нерівностей (7) почнеться для $n \ge n_1 > 0$, де n_1 деяке ціле число, то значення функцій $P_1(\cdot)$ і $P_2(\cdot)$ для стаціонарного хаотичного процесу залишаться незмінними. Природно, що ті процеси, для яких проявляється залежність цих характеристик від початкового моменту часу, називатимуться нестаціонарними.

6. ВИСНОВКИ

Введені поняття хаотичних подій і хаотичних послідовностей, які мають нижню і верхню грані власної функції розподілу послідовності хаотичних подій на множині елементарних подій, а також нижню і верхню грані взаємної функції розподілу різних послідовностей хаотичних подій. Хаотичні події, що реалізуються за часом, називаються хаотичними процесами. Для числових хаотичних послідовностей вказані функції розподілу визначають множинні оцінки (що задаються системою лінійних нерівностей) для значень членів хаотичної послідовності на будь-якому заданому інтервалі сусідніх їх номерів. Такі нерівності можуть використовуватись для гарантованого оцінювання параметрів і стану об'єктів на основі інтервально-множинного аналізу. Проведено комп'ютерне моделювання різних типів хаотичних послідовностей і експериментально визначені їх характеристики.

На рисунку приведено графік нижньої грані функції розподілу числової хаотичної послідовності, згенерованої за допомогою стандартного генератора випадкових чисел, рівномірно розподілених в інтервалі [-1,1].



При цьому рівні x в функції (6) вибирались дискретними з кроком 0,2, а $N = \overline{1,100}$. Кількість членів згенерованої послідовності чисел, на якій вираховувались значення функції $P_н(\cdot)$ в (7) була рівною 3600 .

САМООРГАНІЗАЦІЯ КОЛЕКТИВІВ ВИРІШУВАЛЬНИХ ПРАВИЛ В ЗАДАЧАХ РОЗПІЗНАВАННЯ

В.І.Васильєв

Міжнародний науково-навчальний центр ЮНЕСКО інформаційних технологій та систем, просп. Акад. Глушкова, 40, Київ, МСП 680, 03680, Україна.

Тел. (044) 266-4187, E-mail:vas175@irtc.uran.net.ua

АНОТАЦІЯ

Розглянуто можливість використання колективів вирішувальних правил в задачах розпізнавання образів. Показано, що в алгоритмах, які базуються на методі граничних спрощень та методі групового врахування аргументів, можуть самоорганізовуватись колективи вирішувальних правил, до яких можуть бути застосовані всі відомі правила формування колективних рішень.

ВСТУП

До методів розпізнавання звертаються частіше всього в тих випадках, коли не вдається побудувати формальну теорію, яка б описувала досліджувані об"єкти, та використати класичні математичні методи. Досить довго перважна більшість застосувань теорії розпізнавання була пов"язана з недостатньо формалізованими областями – медициною, геологією, соціологією, хімією та інш. Для вирішення окремої задачі чи класу задач на основі правдоподібних міркувань пропонується нестрогий, але змістовно розумний метод вирішення та побудований на його основі алгоритм.

Колективи вирішувальних правил та їх самоорганізація

Загальне захоплення проблемами розпізнавання привело до появи великої кількості евристичних алгоритмів, які з успіхом вирішували різні конкретні задачі без їх серйозного обгрунтування. При цьому стало очевидним, що поява кожного нового евристичного алгоритму можна розглядати як деякий експеримент, а з усією множиною цих експериментів працювати як з новою для математики множиною об"єктів, тобто вивчати з допомогою строгих математичних процедур вирішення недостатньо формалізованих задач. В результаті сформувався принципово новий підхід до розв"язання задачі розпізнавання, який полягає в переході від дослідження окремих некоректних алгоритмів до дослідження принципів формування колективних рішень. В результаті виник новий метод колективного розпізнавання, в основу якого була покладена алгебраїчна теорія алгоритмів [1]. Найбільш характерним і найбільш практично важливим результатом нового підходу став той факт, що практично завжди ефективність колективного рішення виявлялась вищою від ефективності розв"язку будь-якого конкретного алгоритму із сформованого колективу.

Введення операцій над розпізнавальними операторами забезпечує можливість розширення висхідної сукупності алгоритмів, так як кожний новий оператор породжує новий алгоритм.

Розширені таким чином розпізнаючі алгоритми володіють досить сильними коректуючими властивостями [1]. Навіть у тих випадках, коли у висхідному сімействі розпізнаючих алгоритмів відсутній алгоритм, що правильно вирішує дану задачу, то при виконанні деяких, які можуть бути просто провірені, допущень відносно висхідної навчаючої інформації, алгоритм, що правильно вирішує задачу, все ж існує в

розширеному наборі, причому такий алгоритм може бути виписаний в явному вигляді.

Зустрічаються випадки, коли деякі конкретні алгоритми можуть виробляти відразу цілу множину вирішувальних правил, на базі яких можуть бути сформовані колективи. Тут будуть розглянуті два таких алгоритми: метод групового врахування аргументів (МГВА) [3] і метод граничних спрощень (МГС) [4], які можуть генерувати множини майже рівноцінних по якості, але суттєво різних по структурі, вирішувальних правил, на базі яких можуть бути організовані колективи, що приймають рішення.

В основу методу групового врахування аргументів (МГВА) покладено декілька основних принципів: неостаточність проміжних рішень та єдиність остаточного рішення. Одночасно з оптимізацією коефіцієнтів апроксимуючого поліному відбувається оптимізація його складності, що дозволяє не турбуватися про вибір його структури. Спеціальний критерій (зовнішнє доповнення) виступає перепоною на шляху переускладнення моделі, так як відкидає все, що не здатна відтворити коротка навчаюча вибірка (відкидання «сміття»). В результаті залишається тільки те, що може бути надійно підтверджено конкретною вибіркою.

В якості апроксимуючого поліному у МГВА частіше всього використовується поліном виду:

$$y = \alpha_0 + \sum_{i=1}^{m} \alpha_i x_i + \sum_{i=1}^{m} \sum_{j=1}^{m} \alpha_{ij} x_i x_j + \sum_{i=1}^{m} \sum_{j=1}^{m} \sum_{k=1}^{m} \alpha_{ijk} x_i x_j x_k + \ldots, \quad (1)$$

де x_i – складова вектора, що описує об"єкт. Це зручно, так як з допомогою такого поліному можна досягти досить точної апроксимації любої дифференційованої функції.

Цей складний поліном замінюється множиною простих функцій:

$$y_1 = f(x_1, x_x); y_2 = f(x_1, x_3); \ldots; \quad y_n = f(x_{m-1}, x_m), \quad (2)$$

де $n = C_m^2$, причому функція f скрізь однакова.

Досить часто в якості функції f вибирається така залежність:

$$y = f(x_i x_j) = \alpha_0 + \alpha_1 x_i + \alpha_2 x_j + \alpha_3 x_i x_j \quad (3)$$

Коефіцієнти в цих залежностях знаходять по МНК на точках навчаючої послідовності. Побудовані моделі з певними коефіцієнтами α розглядаються як самостійні розпізнавачі та перевіряються на перевірочній послідовності, яка є зовнішнім доповненням і не використовується при визначенні коефіцієнтів. В результаті цієї перевірки вибираються N кращих моделей і по них будуються більш складні моделі наступного ряду. В другому ряді розглядаються більш складні моделі такого виду:

$$z_1 = f(y_1, y_2); z_2 = f(y_1, y_3); \ldots; \quad z_{n_1} = f(y_{n_1-1}, y_n), \quad (4)$$

де $n_1 = C_n^2$;

Тут функції f залишаються попередніми, тобто відповідають співвідношенню (3), але аргументами цих функцій виступають змінні y_i. Коефіцієнти цих моделей визначаються по МНК на точках тієї самої навчаючої послідовності, а потім ці моделі перевіряються на точках перевірочної послідовності і з них вибираються N кращих, які використовуються в якості аргументів слідуючого третього ряду. Складність поліномів зростає від ряду до ряду, незважаючи на те, що коефіцієнти цих поліномів визначаються по одних і тих же точках навчаючої послідовності. При цьому число визначених коефіцієнтів значно переважає число точок навчаючої послідовності. Якби не було зовнішнього доповнення, тобто самовідбору по перевірочній вибірці, то алгоритми МГВА, дякуючи складності синтезуючих поліномів, могли б абсолютно точно апроксимувати розділяючу

функцію в усіх точках навчаючої вибірки, але при цьому не лишилось би ніяких гарантій задовільного розпізнавання нових точок. Зовнішнє доповнення зупиняє роботу алгоритмів відразу ж по досягненню єдиного мінімуму похибок на перевірочній послідовності. Тим самим вибирається модель оптимальної складності, яка встановлює компроміс між складністю вирішувальної функції та об"ємом інформації, що використовується при синтезі моделі. Але в тому ряді, в якому з"являється модель, що забезпечує мінімум похибок на навчаючій послідовності, існує ще по крайній мірі C_n^2 моделей, які також дають непогані результати. Якщо вибрати з них L найкращих моделей, то можна сформулювати колектив вирішувальних правил, ранг якого дорівнюватиме L. До цього колективу можна застосувати всі ті правила, які використовуються при формуванні колективних рішень, розглянутих в рамках алгебраїчного підходу.

В методі граничних спрощень (МГС) використовується принцип редукції, детально розглянутий в [4]. В основу цього принципу покладено максимальне спрощення (редуціювання) висхідної задачі, в результаті чого синтезується такий простір малої розмірності, в якому лінійне вирішувальне правило найкращим чином разділяє навчаючу вибірку досить великого об"єму. Цей простір синтезується послідовно шляхом відбору і перевірки різних комбінацій властивостей об"єктів [3]. Спочатку розглядаються всі пари із m властивостей (таких пар буде C_m^2). Вибирається в деякому розумінні найкраща пара і до неї підлаштовуються всі інші властивості. Серед отриманих на базі кращої пари трійок, що аналізуються, знову вибирається найкраща і до неї підлаштовуються інші властивості. Такий процес продовжується аж до побудови простору, в якому навчаюча вибірка буде або безпомилково, або найкращим чином розділятися лінійним вирішувальним правилом.

Різновидністю МГС є альфа-процедура [3], яка функціонує за слідуючими правилами. Розглянемо всі пари із m властивостей. Для кожної пари (x_i, x_j) та для кожної точки навчаючої послідовності

$$\breve{x}_{ij} = \rho_{ij}\cos\left(\beta_{ij} + \alpha_{ij}^*\right), \qquad (5)$$

де $\rho_{ij} = \sqrt{x_i^2 + x_j^2}$; $\quad \beta_{ij} = arctg\dfrac{x_j}{x_i}$.

Кут α_{ij} змінюється в інтервалі від 0^0 до 180^0 і являється змінним, що настроюється в процедурі, параметром лінійного вирішувального правила, яке діє в площині (x_i, x_j). Вибирається оптимальне значення

$$\alpha_{ij}^* = \arg\max_s \omega_{ijs} \qquad (6)$$

де ω_{ijs} число правильно розпізнаних об"єктів при допомозі вирішувального правила, що діє в площині (x_i, x_j) при конкретному значенні s. Серед всіх пар вибирається така, в площині якої відбулось найкраще розділення. Цей формалізм практично здійснює проектування точок навчаючої послідовності на напрям $\breve{x}_{ij}$, який повертається навколо початку координат на кут α_{ij} і вибирається такой кут α_{ij}^*, який забезпечує найкраще лінійне розділення вибірки при її проекції на напрямок $\breve{x}_{ij}$.

Після цього до вибраної пари прилаштовуються послідовно всі інші властивості. Припустимо, що була вибрана пара x_i, x_j. Тоді розглядаються пари $(\breve{x}_{ij}, x_k)$, де x_k – одна із властивостей, які заново випробовуються. Серед всіх пар знову вибирається така, що максимізує число правильних відповідей і вже вона визначає трійку x_i, x_j, x_k. Далі розглядаються пари $(\breve{x}_{ijk}, x_t)$, де x_t – одна із властивостей, які ще лишились. Така процедура продовжується доти, поки в площині

однієї з пар не відбудеться або безпомилкове, або оптимальне розділення образів. Тут же ми відмітимо, що кінцевий результат, до якого приводить альфа-процедура, у великій мірі залежить від вибору висхідної пари властивостей. Стандартна альфа-процедура передбачає вибір початкової пари, виходячи з її оптимальності в розумінні найкращого розділення навчаючої послідовності в площині вибраної пари. Але це далеко не завжди гарантує найкращий кінцевий результат, тобто не завжди простір, в якому вперше наступить лінійне розділення, буде мінімальної розмірності при умові, що початкова пара вибрана оптимальною. А це вказує на те, що не завжди доцільно вибирати тільки одну оптимальну пару. Якщо вибрати декілька пар і на їх основі продовжити синтез простору, то буде сформовано декілька просторів різної структури, тобто з різними координатами, в яких можуть бути отримані досить хороші результати. А це означає, що вибір декількох висхідних пар приведе до того, що буде отримано цілу множину рішень однієї й тієї ж задачі, тобто буде отримано колектив вирішувальних правил, на базі якого можна буде сформувати колективні рішення. Якщо вибрати L висхідних пар, то можна побудувати колектив рангу L. Висхідні пари можна вибрати по мірі погіршення якості лінійного розділення навчаючої послідовності і тоді можна сподіватися, що серед отриманої множини рішень будуть і оптимальні. Разом з тим, колективне рішення може дати виграш навіть тоді, коли серед отриманих рішень, які увійшли до колективу, не було жодного, яке вирішувало б задачу безпомилково.

ВИСНОВКИ

Отже, як алгоритми МГВА, так і алгоритми МГС по своїй природі такі, що можуть генерувати не одне, а цілу множину рішень. Якщо вказати тільки ранг колективу, який треба сформувати, то відбудеться самоорганізація колективу наперед вказаного рангу. Причому наперед ніхто не вказує, якої структури вирішувальне правило повинно ввійти в цей колектив. Алгоритми самі вибирають ці правила і самі формують їх структуру. Конструктор повинен вказати тільки ранг коллективу.

Породжувальні при допомозі МГВА та МГС вирішувальні правила такі, що в кожному з них вирішувальна функція може бути представлена у вигляді

$$y = F(X). \qquad (7)$$

Тут X – вектор, який представляє об"єкт розпізнавання. Складові цього вектора в різних алгоритмах можуть суттєво відрізнятися по своїй структурі та можуть містити складні комбінації груп висхідних складових їх множини $(x_1,\ldots,x_m)$. До такої родини вирішувальних правил может бути в повній мірі застосована алгебраїчна теорія, а також будь-які методи прийняття колективних рішень, які дозволять істотно розширити можливості алгоритмів МГВА и МГС.

ЛІТЕРАТУРА

1. *Журавлев Ю.И. Об алгебраическом подходе к решению задач распознавания классификации // Проблемы кибернетики: Сб.ст.М.: Наука, 1978. Вып.33.с.5-68.*

2. *Ивахненко А.Г. Долгосрочное прогнозирование и управление сложными системами. – Киев.: Техника: 1975. – с.311.*

3. *Васильев В.И. Теория редукции в проблемах экстраполяции// Проблемы управления и информатики. – 1996. №1,2. – с.239-251.*

ОПТИМІЗАЦІЯ ПАРАМЕТРІВ НАВЧАННЯ ЗА ВИБОРКАМИ МАЛОГО ОБСЯГУ

А.С.Краснопоясовський, І.В.Шелехов
Сумський Державний Університет
Римського-Корсакова 2, м Суми Україна 40007
Тел. +38 0542 214084, факс +38 0542 334058
E-mail: kras@kpm.sumdu.edu.ua

Розглядається підхід до оптимізації параметрів навчання в рамках екстремально-інформаційного методу функціонально-статистичних випробувань (МФСВ), який ґрунтується на прямій оцінці інформаційної здатності системи розпізнавання.

ВСТУП

Основним недоліком відомих методів розпізнавання образів є відсутність у процесі навчання корекції вхідного математичного опису з метою побудови високоточного класифікатора [1]. Цей недолік усунено в МФСВ [2,3], в рамках якого розглянемо оптимізацію параметрів навчання за навчальними виборками малого обсягу, що має важливе практичне значення.

1. ПОСТАНОВЛЕННЯ ЗАДАЧІ

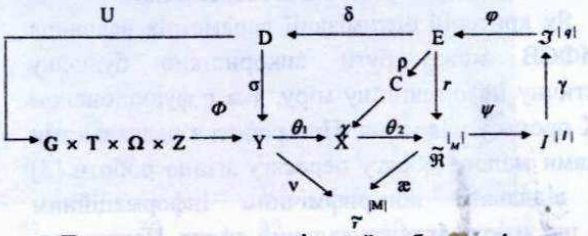
Нехай відомо алфавіт класів розпізнавання $\{X_m^o \mid m=\overline{1,M}\}$ і структурований вектор параметрів навчання $g=<g_1,\ldots,g_\xi,\ldots,g_\Xi>$, які впливають на точнісні характеристики (ТХ) класифікатора. Такими параметрами є безпосередньо геометричні параметри контейнерів і параметри нормалізації образів. Під контейнером K_m^o класу X_m^o згідно роботи [4] тут розуміється його замкнена роздільна гіперповерхня, для якої деяким способом визначено геометричний центр. Для нечіткого розбитя $\widetilde{\Re}^{|M|}$ бінарного простору ознак $\Omega_Б$, яке допускає перетиняємість класів розпізнавання, необхідно виконання таких умов:

1. $(\forall X_m^o \in \widetilde{\Re}^{|M|})[X_m^o \neq \varnothing]$;
2. $(\exists X_k^o \in \widetilde{\Re}^{|M|})(\exists X_l^o \in \widetilde{\Re}^{|M|})[X_k^o \neq X_l^o \to$
$\to X_k^o \cap X_l^o \neq \varnothing]$, $k,l=\overline{1,M}$;
3. $(\forall X_k^o \in \widetilde{\Re}^{|M|})(\forall X_l^o \in \widetilde{\Re}^{|M|})[X_k^o \neq X_l^o \to$
$\to KerX_k^o \cap KerX_l^o = \varnothing]$, тобто ядра класів не перетинаються;
4. $\bigcup_{X_m^o \in \widetilde{\Re}^{|M|}} X_m^o \subseteq \Omega_Б \subset \mathbf{X}$, де **X**–простір Хеммінга.

Так само відома дійсна або ціла навчальна матриця $\| y_{m,i}^{(J)} \|$. Треба шляхом нормалізації матриці $\| y_{m,i}^{(J)} \|$ для класу X_m^o знайти максимальне значення критерію функціональної ефективності (КФЕ) E_m^*, яке визначає оптимальні в інформаційному (точнісному) розумінні параметри навчання.

2. МАТЕМАТИЧНА МОДЕЛЬ ОПТИМІЗАЦІЇ ПАРАМЕТРІВ НАВЧАННЯ

Подамо універсум випробувань у вигляді декартова добутка $W = G \times T \times \Omega \times Z$, де G - вхідна змінна; T - множина моментів зняття інформації; Ω –простір ознак; Z –множина станів. Введемо оператор $\Phi : \overline{W} \to Y$ формування вхідного математичного опису Y класифікатора – навчальної матриці $\| y_{m,i}^{(J)} \|$ і оператор $\theta : Y \to \widetilde{\Re}^{|M|}$ побудови розбиття $\widetilde{\Re}^{|M|}$, який є композицією двох операторів: $\theta_1 : Y \to X$ і $\theta_2 : X \to \widetilde{\Re}^{|M|}$, де X – бінарна вибіркова множина. Тоді діаграма відображень множин, які застосовуються в процесі навчання за МФСВ, має вигляд



(1)

Тут оператор ν апріорної побудови і оператор æ апостеріорної побудови нечіткого покриття $\widetilde{L}^{|M|}$,

яке задається алфавітом $\{X_m^o\}$, замикають комутативне кільце діаграми. Нехай оператор класифікації $\Psi : \tilde{\Re}^{|M|} \to I^{|l|}$ перевіряє основну статистичну гіпотезу про належність реалізацій класу X_m^o, де $I^{|l|}$ – множина виходів (гіпотез); l– кількість гіпотез. Для обчислення КФЕ навчання оператор $\gamma : I^{|l|} \to \Im^{|q|}$ шляхом оцінки прийнятих гіпотез формує множину $\Im^{|q|}$, де $q = l^2$ – кількість ТХ, а оператор $\varphi : \Im^{|q|} \to E$ обчислює множину E значень КФЕ, який є функціоналом від ТХ. Ітераційний процес оптимізації параметрів розбиття реалізується оператором $r : E \to \tilde{\Re}^{|M|}$ шляхом пошуку максимуму критерію E_m^*. Завдяки принципу композиції кожний параметр навчання має відповідний контур оптимізації в діаграмі (1). При цьому обов'язковим є виконання умови повної композиції: множина E є загальною для всіх контурів оптимізації. Серед інших параметрів навчання, які впливають на достовірність класифікатора, розглянемо систему полів контрольних допусків $\{\delta_{K,i} \mid i = \overline{1,N}\}$ на значення ознак розпізнавання і рівні селекції $\{\rho_m\}$ координат еталонних векторів (ЕВ), які визначають центри контейнерів. Нехай ітераційний процес оптимізації системи контрольних допусків, яка утворює множину D, реалізується оператором $\delta : E \to D$ і оператором $\sigma : D \to Y$, який змінює значення реалізацій образу в процесі навчання. Нехай значення рівнів селекції утворюють множину C, а ітераційний процес їх оптимізації реалізується послідовно оператором $\rho : E \to C$ і оператором $\chi : C \to X$, який змінює бінарну навчальну матрицю. Оператор $U : E \to G \times T \times \Omega \times Z$ регламентує процес навчання.

3. КРИТЕРІЇ ОПТИМІЗАЦІЇ

Як критерій оптимізації параметрів навчання за МФСВ може бути використано будь-яку статистичну інформаційну міру, яка є функціоналом від ТХ процесу навчання. При роботі з навчальними вибірками малого обсягу перевагу згідно роботи [5] варто віддавати логарифмічним інформаційним мірам, що мають згладжувальний ефект. Наприклад, для двоальтернативної системи оцінок і рівноймовірних гіпотез згідно принципу неповноти апріорної інформації Бернуллі-Лапласа можна застосовувати, наприклад, модифікацію нормованого критерію за Шенноном у вигляді [6]:

$$E = 1 + \frac{1}{2}\left(\frac{\alpha}{\alpha + D_2}\log_2\frac{\alpha}{\alpha + D_2} + \frac{\beta}{D_1 + \beta}\log_2\frac{\beta}{D_1 + \beta} + \right. \quad (2)$$
$$\left. + \frac{D_1}{D_1 + \beta}\log_2\frac{D_1}{D_1 + \beta} + \frac{D_2}{\alpha + D_2}\log_2\frac{D_2}{\alpha + D_2}\right),$$

де α, β, D_1, D_2 –ТХ: помилки першого і другого роду, перша і друга достовірності відповідно. Критерій (2) є нелінійним і неоднозначним функціоналом ТХ, що вимагає оцінки в процесі навчання робочої області визначення його значень.

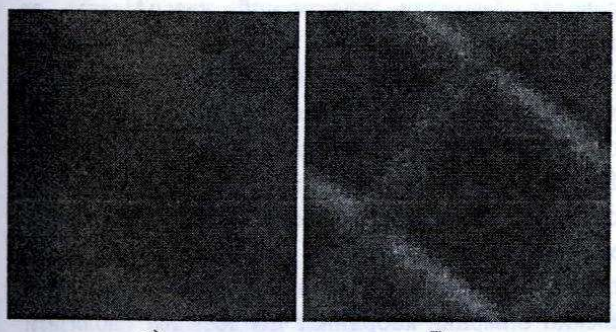
4. ОСОБЛИВОСТІ РЕАЛІЗАЦІЇ АЛГОРИТМУ ОПТИМІЗАЦІЇ ПАРАМЕТРІВ НАВЧАННЯ

Алгоритм навчання у рамках МФСВ полягає в послідовній реалізації ітераційних процедур оптимізації параметрів навчання шляхом пошуку максимуму КФЕ навчання для кожного контуру оптимізації в діаграмі (1). З метою спрощення класифікатора як наближення "точного" контейнера, наприклад, класу X_m^o може розглядатися гіперсфера, центром якої є вершина ЕВ $x_m \in X_m^o$, що відповідає прийнятній для задач контролю і управління гіпотезі компактності (як чіткої, так і нечікої) реалізацій образу. Тоді як параметри оптимізації за діаграмою (1) розглядаються радіуси $\{d_m\}$ контейнеріів, контрольні допуски на ознаки розпізнавання і рівні селекції $\{\rho_m\}$ координат ЕВ.

5. ПРИКЛАД РЕАЛІЗАЦІЇ АЛГОРИТМУ НАВЧАННЯ ЗА МФСВ

Розглянемо задачу оптимізації геометричних параметрів контейнера у вигляді гіперсфери на прикладі автофокусування растрового електронного мікроскопа РЕМ-103 виробництва ВАТ SELMI (м.Суми, Україна) за зображенням зразка, що досліджується. Ідея автофокусування полягає в побудові на кожному кроці настроювання мікроскопа оптимального контейнера для початкового розфокусованого зображення (класс X_1^o) як за своїми реалізаціями, так і реалізаціями поточного зображення (класс X_S^o), отриманими на S-му кроці настроювання, шляхом пошуку максимуму інформаційної міри ррізноманітності цих класів. Як параметр настроювання розглядався струм управляючої обмотки об'єктивної лінзи. На рис.1а

наведено початкове розфокусоване зображення об'єкту «Гратка», а на рис.1б – зображення цього об'єкту, отриманого після першого кроку настроювання..



а) б)

Рис.1. Зображення об'єкту «Гратка»: а)початкове розфукусоване зображення; б) поточне зображння

У процесі реалізації алгоритму навчання за діаграмою (1) було побудовано отимальний контейнер $K_1^o \subset X_1^o$, який забезпечує максимальну асимптотичну достовірність розпізнавання в режимі екзамену реалізацій класу X_1^o. Оптимальний радіус d_1^* контейнера K_1^o, як видно з рис.2, дорівнює 35 кодовим одиницям відстані при значенні КФЕ $E_1^* = 0{,}809$. На рис.2 темна область графіка позначає робочу область його визначення, в якій $D_1, D_2 > 0{,}5$. В області I присутні тільки реалізації класу X_1^o, в області II присутні реалізації класів X_1^o і X_S^o, а в області III – тільки реалізації класу X_S^o. В табл.1 наведено значення критерію E_1, першої достовірності D_1 і помилки другого роду β, параметрів навчання ρ_1, d_1, δ і міжцентрова кодова відстань $d_c = d(x_1 \oplus x_S)$. Як видно з табл.1, оптимальне значення рівня селекції дорівнює $\rho_1^* = 0{,}54$, а оптимальне поле контрольних допусків для *i*-ої ознаки дорівнює $x_{1,i} \pm \delta^*$, де δ^*=12 градаціям яскравості.

Таблиця 1. Результати оптимізації контейнера K_1^o

ρ_1	E_1	D_1	β	d_1	δ	d_c
0,40	0,674989	0,80	0,00	39	12	55
0,42	0,710603	0,84	0,00	33	14	4
0,44	0,710603	0,84	0,00	33	14	38
0,46	0,710603	0,84	0,00	31	14	38
0,48	0,701025	0,83	0,00	30	14	43
0,50	0,754381	1,00	0,12	38	12	45
0,52	0,794417	1,00	0,09	37	12	48
0,54	**0,809527**	**1,00**	**0,08**	**35**	**12**	**48**
0,56	0,809527	1,00	0,08	35	12	51
0,58	0,765397	0,99	0,08	33	12	53
0,60	0,794417	1,00	0,09	33	12	51

Початкові координати ЕВ дорівнювали вибірковим середнім відповідних навчальних виборок. Оскільки навчальна матриця для кожного класу складалася тільки із 100 реалізацій, то центри розсіювання реалізацій не збігалися з вершинами початкових ЕВ.

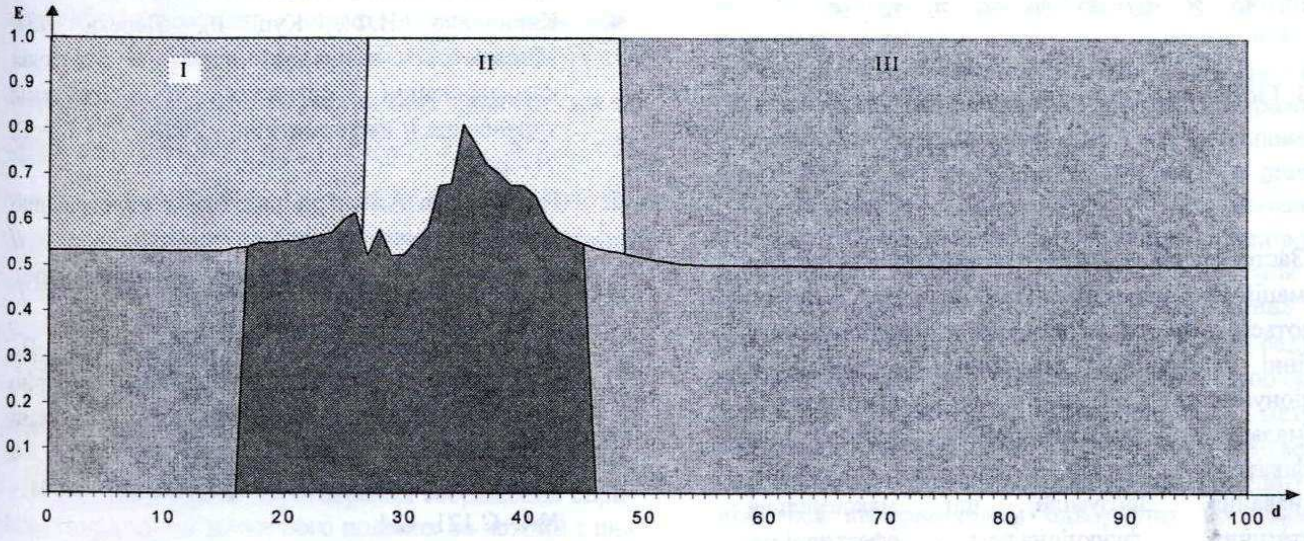
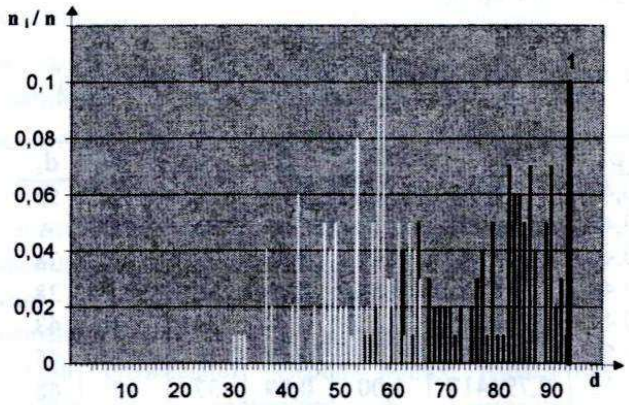
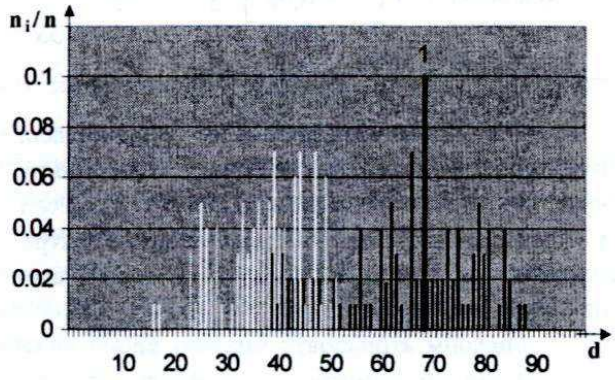


Рис. 2, Залежність КФЕ від радіуса контейнера K_1^o

На рис.3 показано гістограми розподілів реалізацій класів X_1^o (темні дискрети) і X_S^o (світлі дискрети). Тут мітка 1 позначає положення вершини початкового вектора x_1 при ρ_1=0,50 (рис.3а) і положення вершини цього вектора при оптимальному рівні селекції $\rho_1^* = 0{,}54$, яке, як видно з рис.3б, є більш наближеним до центру розсіювання реалізацій класу X_1^o.



а)



б)

Рис. 3, Гістограма розподілу реалізацій класів X_1^o і X_S^o : а) ρ=0,50; б) ρ=0,54

ПІДСУМОК

Застосування МФСВ для розв'язання задачи інформаційного синтезу систем розпізнавання, що навчаються, дозволяє оптмізувати в інформаційному розумінні просторово-часові параметри функціонування системи, які забезпечують максимальну асимтотичну достовірність класифікатора за умови перетиняємості класів розпізнавання. Зрозуміло, що максимальна асимптотична функціональна ефективність класифікатора збігається з фактичною тільки за умови однакової структурованості навчальної та екзаменаційної матриць і однакових показників їх статистичної сталості та статистичної однородності. На відміну від відомих параметричних і непараметричних методів розпізнавання МФСВ дозволяє будувати ефективний класифікатор за репрезентативними в розумінні Паретто навчальними виборками малого обсягу, що важливо для розв'язання практичних задач контролю та управління. Основне обмеження на використання МФСВ полягає в наявності у контейнерів класів розпізнавання деяким способом визначеного геометричного центра Показано, що вибір оптимального в інформаційному розумінні рівня селекції бінарних координат ЕВ дозволяє наблизити центр контейнера з центром розсіювання реалізацій відповідного класу, що призводить до підвищення достовірності класифікатора..

ЛІТЕРАТУРА

1. Васильев В. И. Распознающие системы. Справочник. 2е изд., перераб. и доп. – Киев: Наукова думка, 1983.– 422 с.
2. Краснопоясовський А. С., Заговора С. О. Розпізнавання електронограм за методом функціонально – статистичних випробувань // Оброблення сигналів і зображень, та розпізнавання образів: Праці п'ятої Всеукраїнської міжнародної конференції. – Київ, 2000. – С. 37 – 41.
3. Краснопоясовский А.С. Классификационная настройка сложной системы // Управляющие системы и машины.–2002.–№1.–С. 9–15.
4. Кириченко Н.Ф., Куц Р., Лепеха Н.П. Множества принадлежности в задачах классификации сигналов // Проблемы управления и информатики.– 2001. – № 5. – С. 71– 84.
5. Колмогоров А.Н. Три подхода к определению понятия «количество информации» // Проблемы передачи информации, 1961.–№3.–С.27–35.
6. Краснопоясовський А. С., Черниш А. В., Сластушевський О.Ю. Про вибір критерію функціональної ефективності системи розпізнавання, що навчається // Радиоэлектроника и информатика, 2001.– №4.–С.121–124.

ІНФОРМАТИВНІСТЬ АФІННО-ІНВАРІАНТНОЇ ГЕОМЕТРИЧНОЇ МОДЕЛІ ПРОЕКЦІЙНИХ ЗОБРАЖЕНЬ В ЇХ МОРФОЛОГІЧНОМУ АНАЛІЗІ

В.М.Корчинський
Дніпропетровський національний університет
49106, м. Дніпропетровськ, вул. Наукова, 13
Тел.: (0562)43-36-30 Факс: (0562)46-55-23
e-mail: buser @ ap1.net-rff.dsu.dp.ua

АНОТАЦІЯ

Визначені інформаційні характеристики (нормовані семіінваріанти функцій яскравості) та їх лінійні комбінації, найбільш інформативні при позиційно-інваріантному розпізнаванні проекційних зображень, одержаних в умовах дальньої фотограмметрії. Запропоновано спосіб визначення оптимальної вимірності простору моделі для забезпечення ідентифікації та аналізу таких зображень із заданою достовірністю.

1. ВСТУП

В роботах [1,2] було розвинено інформаційно-геометричну модель подання форм зображень проекційної природи, інваріантну стосовно їх афінних перетворень, спричинених зміною позиційних умов проекціювання. У відповідності з нею множина розподілів функцій яскравості таких зображень, одержаних при різних позиційних умовах проекціювання (ракурс, відстань), подається точками квадрик у просторі інформаційних характеристик – безрозмірних семіінваріантів $S_{n,m}$ парного порядку N від функцій яскравості, нормування яких здійснюється за правилом $Z_i = S_{i-1,N+1-i} \big/ S_{0,0}^{n}$, $i = \overline{1, N+1}$, де $n = N/2$. З геометричних позицій такі квадрики являють собою сфери при псевдоевклідовій метризації простору інформаційних характеристик $\mathbf{R}_n^{N+1}$, радіуси яких є афінно-інваріантними інформаційними ознаками. Розпізнавання зображення зводиться до визначення належності точки його подання псевдосферам, що визначаються еталонним зображенням у просторах моделі різних вимірностей (при різних N), причому афінні перетворення зображення спричинять змішення точки його подання по кожній з цих псевдосфер. Необхідною умовою реалізації моделі є лінійна незалежність семіінваріантів функцій яскравості, на основі яких визначаються компоненти вектора його подання у просторі моделі, а збільшення його вимірності виправдане лише у випадку реального зростання інформативності. Між тим, існуючі оцінки свідчать, що лінійна незалежність семіінваріантів заданого порядку N може існувати лише в деякій області простору моделі, за межами якої використання усіх компонент вектора $\mathbf{Z}$ як інформаційних характеристик зображення є надлишковим. Ця обставина ставить проблему визначення інформативності подання проекційних зображень в рамках афінно-інваріантної моделі їх просторових форм з метою оптимізації алгоритмів морфологічного (позиційно-інваріантного) аналізу. Відповідні результати пов'язані з дослідженням структури перетворень вектору подання його геометричної форми, індукованих афінними перетвореннями у картинній площині.

2. ФІЗИЧНА МОДЕЛЬ ФОРМОУТВОРЕННЯ ПРОЕКЦІЙНИХ ЗОБРАЖЕНЬ

Розглядаються проекційні зображення, одержані в умовах дальньої фотограмметрії – відстані між точками об'єкту, відеодатчиком та картинною площиною вважаються суттєво більшими, ніж довжина хвилі випромінення – носія відеоінформації. Об'єкт приймається плоским, а просторовий розподіл інтенсивності випромінення елементів його поверхні описується на феноменологічному рівні – за допомогою індикатриси розсіювання (випромінювання). Вважається, що яскравість зображення пропорційна інтенсивності випромінення. Відповідність між точками об'єкту та його проекційного зображення приймається афінною, $\mathbf{r} = \mathbf{A} \cdot \mathbf{x} + \mathbf{b}$, де $\mathbf{r}, \mathbf{x}$ – радіуси-вектори відповідних точок зображення та об'єкту в картинній та предметній площинах; $\mathbf{A}$ – матриця однорідних геометричних перетворень; $\mathbf{b}$ – вектор перенесень. (Випадок проективної відповідності реалізується використанням однорідних координат в картинній площині.)

Інтегральний зв'язок між функцією яскравості зображення $I(\mathbf{r})$ та скалярною яскравістю $I_0(\mathbf{r})$ в рамкках наведених припущень має вигляд [3]

$$I(\mathbf{r}) = C \int_{-\infty}^{\infty} I_0(\mathbf{B}\cdot\mathbf{x}) P(\mathbf{r}-\mathbf{x}) d\mathbf{x}, \quad (1)$$

де $P(\mathbf{x})$ – функція розсіювання точки (ФРТ); $\mathbf{B} = \mathbf{A}^{-1}$; $C = (\lambda)^{-4}(Hd)^{-2}$, λ – довжина хвилі випромінення - носія відеоінформації, H – висота, з якої здійснюється фіксація зображення, d – відстань між відеодатчиком та картинною площиною.

Співвідношення (1) описує неточкову (нелокальну) відповідність між точками предметної та картинної площини і є основою для обчислення компонент вектору подання зображення $\mathbf{Z}$.

3. МЕТРИЧНІ ВЛАСТИВОСТІ ПРОСТОРУ МОДЕЛІ

Зв'язок семіінваріантів функції яскравості зображення S_{nm}, скалярної яскравості об'єкту $S_{nm}^{(0)}$ та ФРТ s_{nm}, обчислений на основі співвідношень (1) за методикою [1,2], має вигляд

$$S_{nm} = \sum_{k=0}^{n} \sum_{l=0}^{m} C_n^k C_m^l b_{11}^k b_{21}^{n-k} \times$$
$$\times\, b_{22}^l b_{21}^{m-l} S_{k+m-l,n-k+l}^{(0)} + s_{nm}, \quad (2)$$

де C_n^m – біноміальні коефіцієнти; b_{ij} – компоненти матриці $\mathbf{B}$. Відповідні перетворення вектору $\mathbf{Z}$, індуковані зміною позиційних умов проекціювання, у матричній формі:

$$\mathbf{Z}_2 = k_{12}^n \mathbf{M}_2 \mathbf{M}_1^{-1} \mathbf{Z}_1 + K_2 \mathbf{z}, \quad (3)$$

де $\mathbf{M}$ – матриця перетворення (2); індекси 1, 2 відповідають еталонному зображенню та зображенню, що аналізується; $k_{12} = C_1/C_2$; $\mathbf{z}$ – вектор подання ФРТ; $K_2 = \left[C_2 Y_2 M_{00}^{(2)}\right]^{-n}$; $Y_2 = \det(\mathbf{B})$. Через M_{00} позначено нульовий декартів момент функції яскравості зображення $\mathbf{Z}_2$. У випадку точкової (локальної) відповідності між об'єктом та його проекційним зображенням $\mathbf{z} = \mathbf{0}$.

Компоненти вектору $\mathbf{Z}$ в силу їх означення інваріантні стосовно паралельних перенесень зображення.

Інваріантні багатовиди другого степеня у просторі $\mathbf{R}_n^{N+1}$ для $k = 1$, $\mathbf{z} = \mathbf{0}$ встановлені в роботах [1,2]. Узагальнюючи відповідні результати на що випадок, що розглядається, одержуємо афінно-інваріантну квадрику у вигляді

$$(\mathbf{Z} - K_2\mathbf{z})^T \cdot \mathbf{G} \cdot (\mathbf{Z} - K_2\mathbf{z}) = \frac{1}{k_{12}^2} F_N^2, \quad (4)$$

де F_N^2 – афінно-інваріантна інформаційна ознака геометричної форми зображення. Через $\mathbf{G}$ в (4) позначено матрицю метричного тензору простору моделі $\mathbf{R}_n^{N+1}$ з елементами:

$$g_{ij} = (-1)^{n+1-i} \frac{n!}{(i-1)!} \frac{n!}{(N+1-i)!} \delta_{N+2-i,j}, \quad (5)$$

де $\delta_{i,j}$ - символ Кронекера.

Співвідношення (4) дає можливість трактувати афінно-інваріантну квадрику як псевдосферу радіуса $k_{12}^{-1}|F_N|$ з центром в точці $K_2\mathbf{z}$ у просторі $\mathbf{R}_n^{N+1}$ з метричним тензором (5).

4. СТРУКТУРА ПЕРЕТВОРЕНЬ ВЕКТОРІВ ПОДАННЯ ПРОЕКЦІЙНИХ ЗОБРАЖЕНЬ

Розглянемо лінійні багатовиди у просторі $\mathbf{R}_n^{N+1}$, інваріантні стосовно афінних перетворень функції яскравості проекційного зображення.

Твердження 1. Перетворення вектора подання довільного проекційного зображення у псевдоевклідовому просторі моделі $\mathbf{R}_n^{N+1}$, індукованого афінним перетворенням зображення в картинній площині, має принаймні одну інваріантну пряму.

Оскільки підпростір, ортогональний до інваріантного, є також інваріантним підпростором [4], визначена пряма є інваріантною стосовно перетворення з матрицею $\mathbf{M} = \mathbf{M}_2\mathbf{M}_1^{-1}$. В загальному випадку кількість інваріантних прямих та інваріантних гіперплощин перетворення $\mathbf{Z}_1 \to \mathbf{Z}_2$ дорівнює відповідно кількості дійсних та комплексних коренів характеристичного рівняння матриці $\mathbf{M}$.

Побудова інваріантних підпросторів дозволяє оцінювати інформаційні властивості окремих компонент векторів подання геометричних форм зображень. Інваріантність гіперплощини у просторі $\mathbf{R}_n^{N+1}$ означає, що точки подання зображень, розташовані у ній, залишаються належними цій гіперплощині при їх афінних перетвореннях; розташування нормалі $\mathbf{n}$ до такої гіперплощини залишається незмінним. Внаслідок цього комбінація компонент вектора подання зображення, яка визначається вектором нормалі, $\mathbf{n}\cdot\mathbf{G}\cdot\mathbf{Z}$, є сталою до такого перетворення і має найбільшу інформативність. Оскільки цей вектор визначає інваріантну пряму, перетворення вектора подання зображення вздовж неї є тотожним – відповідне власне число матриці перетворення $\mathbf{M}$ дорівнює одиниці. Аналогічно, у випадку, якщо матриця перетворення точок подання зображень у просторі моделі має власне число, яке дорівнює одиниці, найбільш інформативною є комбінація компонент вектора подання, що визначається відповідним власним вектором матриці $\mathbf{M}$ (останній має зміст напрямного вектора відповідної інваріантної прямої).

Отже, інформативність комбінації компонент вектора подання зображення, яка визначається напрямним вектором довільної інваріантної прямої, тим вища, чим ближче відповідне власне число матриці перетворення $\mathbf{M}$ у просторі моделі до одиниці.

В загальному випадку за показник інформативності окремих компонент вектору подання зображення у просторі $\mathbf{R}_n^{N+1}$, яка визначається k-ю інваріантною прямою, пропонується прийняти величину $\lambda_k \Big/ \sum_{i=1}^{P} |\lambda_i|$, де P - кількість власних чисел матриці $\mathbf{M}$. Показник інформативності лінійних комбінацій компонент векторів подання зображень при їх однопараметричних афінних перетвореннях в картинній площині мають дорівнює:

– при масштабуванні вздовж осей картинної системи координат 0.987 (найбільш інформативна компонента Z_{n+1});

– при косому зсуві вздовж осей картинної системи координат 0.91 (найбільш інформативні компоненти Z_1 та Z_{N+1});

– при повороті з довільним центром 0.93 (найбільш інформативна комбінація $Z_1 + Z_{n+2}$).

Як випливає з співвідношень (3), (4), точки подання зображень, зафіксованих у різних спектральних діапазонах, розташовані на афінно-інваріантних псевдосферах з різними радіусами та центрами. Відповідність псевдосфер встановлюється співвідношенням (3): точка $\mathbf{Z}_2$ подання зображення, що аналізується, та еталонного зображення $\mathbf{Z}_1$ пов'язані проективною відповідністю з центром у точці $\mathbf{z}$. За наявності додаткових афінних перетворень зображення, зумовлених зміною позиційних умов фіксації, точка $\mathbf{Z}_2$ зміщується вздовж лінії перетину відповідної квадрики з інваріантною площиною.

На рис. 1-4 подані напівтонові зображення фрагментів земної поверхні, одержані при однакових позиційних параметрах проекціювання, але на різних довжинах хвиль в оптичному діапазоні (зображення на рис.1 прийняте за еталонне) відповідно при $\lambda_1 = 0.8$ мкм, $\lambda_2 = 0.93$ мкм, $\lambda_3 = 0.97$ мкм, $\lambda_4 = 1.1$ мкм. Компоненти векторів подання зображень у просторі $\mathbf{R}_1^3$ дорівнюють:

$\mathbf{Z}_1 = (0.35, -1.23, 3.25)^T$; $\mathbf{Z}_2 = (1.37, 2.31, 1.75)^T$;

$\mathbf{Z}_3 = (2.59, 0.93, 2.25)^T$; $\mathbf{Z}_4 = (3.86, 0.62, 4.72)^T$,

що добре узгоджується з наведеним ідентифікаційним критерієм. Зазначимо, що на основі векторів подання трьох зображень фіксованого об'єкту та відомих параметрах їх фіксації можна відновити вектор подання функції розсіювання точки $\mathbf{z}$ (див. співвідношення (3)).

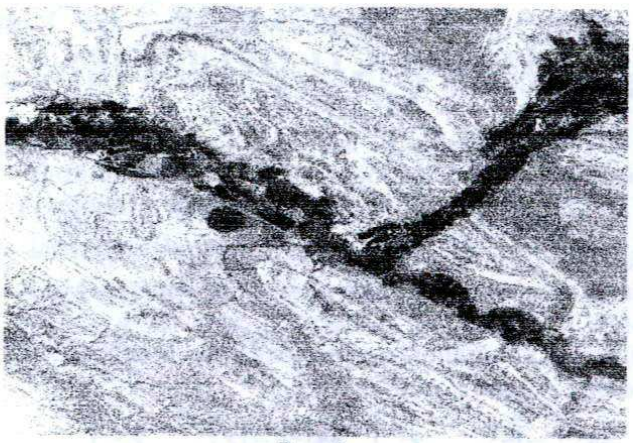


Рис. 1

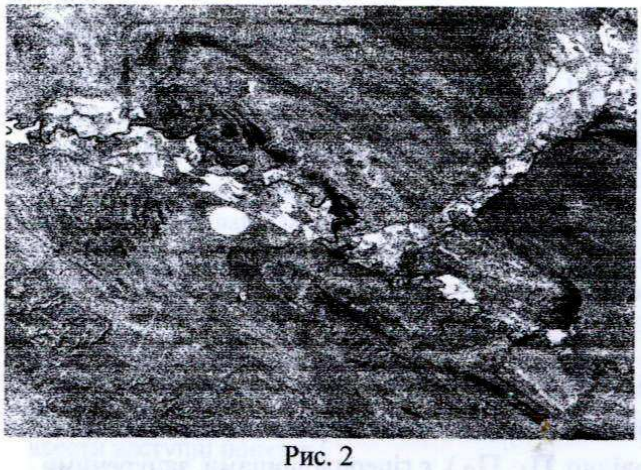


Рис. 2

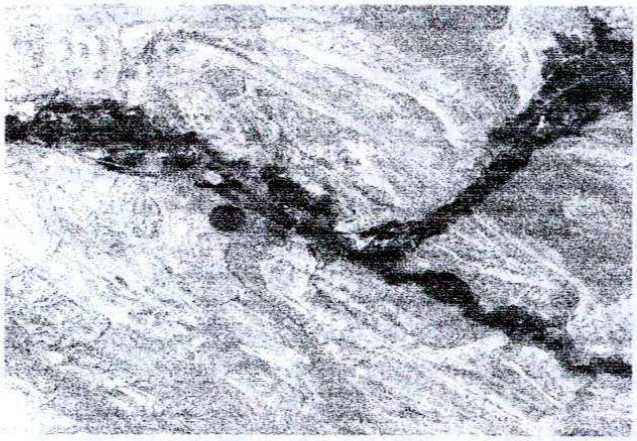


Рис. 3

Рис. 4

5. ОПТИМІЗАЦІЯ ВИМІРНОСТІ ПРОСТОРУ МОДЕЛІ

Розглянемо m векторів подання фіксованого проекційного зображення у просторах моделі різних вимірностей(навчальну вибірку) $\mathbf{z}_k \in \mathbf{R}_n^{N+1}$, $\mathbf{Z}_k \in \mathbf{R}_l^{L+1}$, де $L > N$, $k = \overline{1,m}$; індекс 1 відповідає еталонному зображенню. Множини цих векторів подаємо у вигляді матриць, рядки яких відповідають окремим векторам, а стовпці – їхнім компонентам

$$\mathbf{M}^{(1)} = \left(\mathbf{z}_1^T, \ldots, \mathbf{z}_m^T\right)^T, \mathbf{M}^{(2)} = \left(\mathbf{Z}_1^T, \ldots, \mathbf{Z}_m^T\right)^T,$$

позначаючи елементи матриць відповідно через $m_{ij}^{(1)}$ та $m_{ik}^{(2)}$ ($j = \overline{1, N+1}$, $k = \overline{1, L+1}$). Оскільки вектори подання зображень фіксованого об'єкту, одержаних за різних умов проекціювання, пов'язані між собою перетворенням (1), мають місце співвідношення

$$\bar{\mathbf{z}}_k = \mathbf{B}_k \cdot \mathbf{z}_1, \bar{\mathbf{Z}}_k = \mathbf{C}_k \cdot \mathbf{Z}_1 \quad (k = \overline{1,m}), \quad (7)$$

де через $\mathbf{B}_k, \mathbf{C}_k$ позначені матриці перетворень k - го зображення у просторах $\mathbf{R}_n^{N+1}, \mathbf{R}_l^{L+1}$.

Лінійні простори вимірності m, базиси яких визначаються стовпцями матриць $\mathbf{M}^{(1)}, \mathbf{M}^{(2)}$ (відповідно Π_1, Π_2), є гіперплощинами, зануреними у псевдоевклідів простір $\mathbf{R}_{l+1}^{L+3}$. Базисні вектори цих просторів позначаємо як $\left\{\mathbf{e}_i : i = \overline{1, N+1}\right\}$ та $\left\{\mathbf{E}_i : i = \overline{1, L+1}\right\}$ відповідно. Параметричні рівняння гіперплощин Π_1, Π_2 мають вигляд:

$$\mathbf{r} = \mathbf{r}_1 + \mathbf{A}_1 \cdot \mathbf{t}; \mathbf{r} = \mathbf{r}_2 + \mathbf{A}_2 \cdot \mathbf{t}, \quad (8)$$

де $\mathbf{t} = (t_1, \ldots, t_{L+3})$ – вектор параметрів; $\mathbf{r}_1$, $\mathbf{r}_2$ – довільні точки Π_1 та Π_2. У співвідношенні (8) позначено: $\mathbf{A}_1 = \left(\mathbf{P}^{(1)}\mathbf{e}_1, \ldots, \mathbf{P}^{(1)}\mathbf{e}_{N+1}\right)$;

$\mathbf{A} = \left(\mathbf{P}^{(2)}\mathbf{E}_1, \ldots, \mathbf{P}^{(2)}\mathbf{E}_{L+1}\right)$, де $\mathbf{P}^{(1)}, \mathbf{P}^{(2)}$ – матриці проекціювання просторів $\mathbf{R}_n^{N+1}$, $\mathbf{R}_l^{L+1}$ на простір $\mathbf{R}_{l+1}^{L+3}$.

Квадрат косинуса кута між площинами Π_1 та Π_2 обчислений у псевдо евклідовій метриці простору $\mathbf{R}_{l+1}^{L+3}$ дорівнює максимальному власному числу матриці

$$\left(\tilde{\mathbf{A}}_1^T \mathbf{G} \mathbf{A}_1\right)^T \left(\tilde{\mathbf{A}}_1^T \mathbf{G} \mathbf{A}_2\right)^T \left(\mathbf{A}_2^T \mathbf{G} \mathbf{A}_2\right)^T \left(\mathbf{A}_2^T \mathbf{G} \tilde{\mathbf{A}}_1\right)^T,$$

де $\mathbf{G}$ – матриця метричного тензора простору $\mathbf{R}_{l+1}^{L+3}$; $\tilde{\mathbf{A}}_1$ – матриця проекцій напрямних векторів гіперплощини Π_1 на гіперплощину Π_2.

Твердження 2. Мірою додаткової інформації про форму проекційного зображення, яка вноситься при збільшенні порядку моделі, є кут між гіперплощинами, що визначаються стовпцями ін-формаційних матриць подання множини зображень у моделюючих просторах відповідних вимірностей.

Алгоритм визначення оптимальної вимірності простору моделі базується на обчисленні кута між гіперплощинами Π_1 та Π_2 при збільшенні порядку моделі за умови лінійної не залежності стовпців інформаційних матриць. Оптимальній за інформативністю вимірності простору моделі відповідає нульове значення цього кута.

6. ВИСНОВКИ

1. Афінні перетворення проекційних зображень, зумовлені зміною позиційних умов проекціювання, спричиняють однорідні лінійні перетворення інформаційних характеристик зображень (безрозмірних семіінваріантів функцій яскравості), структура яких визначає інваріантні лінійні багатовиди у просторі моделі.

2. Перетворення вектору подання довільного проекційного зображення у псевдоевклідовому просторі моделі $\mathbf{R}_n^{N+1}$, індукованого афінним перетворенням зображення у картинній площині, має принаймні одну інваріантну пряму. Усі афінно-інваріантні гіперплощини простору моделі ортогональні інваріантнім прямим в рамках псевдоевклідової метризації простору моделі.

3. Мірою додаткової інформації про форму та розподіл яскравості проекційного зображення, яка забезпечується при використанні інваріантної геометричної моделі при збільшенні її порядку, є кут між гіперплощинами, що визначаються стовпцями інформаційних матриць подання навчальної вибірки зображень у моделюючих просторах відповідних вимірностей.

ЛІТЕРАТУРА

1. Михайленко В.Є., Корчинський В.М. *Концепція геометричного об'єкту в морфологічному аналізі проекційних зображень* // Прикладна геометрія та інженерна графіка. - К.: КДТУБА, 1997. - Вип. 61. - С.59-63.
2. Корчинский В.М. *Обратная задача идентификации изопланатических изображений* // Математические модели и современные информационные технологии. - К.: Ин-т математики НАН Украины, 1998. - С. 69-71.
3. Рытов С.М., Кравцов Ю.А., Татарский В.И. *Введение в статистическую радиофизику: Случайные поля*. - М.: Наука, 1978. - 464 с.
4. Гельфанд И.М. *Лекции по линейной алгебре*. - М.: Наука, 1966. - 364 с.

Кластерізація багатовимірних та багатопараметрових зображень "без вчителя" на основі генетичного алгоритму

О. М. Ахметшин. Е. В. Пирогов

Дніпропетровський Національний Університет, Україна, 49050, м. Дніпропетровськ, пер. Науковий 13

Електронна пошта : akhm@mail.dsu.dp.ua

Abstract

In article offered new unsupervised clastering method multimeasure and multiparameter data based on genetic algorithm. The model is capable to an estimation of amount clasters in data. The article consist of the theoretical description of the offered model and her parameters, examples of clastering multiparameter NMR medical image and multimeasure image of earth surface.

Анотація

В роботі запропоновано новий метод кластерізації "без вчителя" багатовимірних та багатопараметрових даних оснований на генетичному алгоритмі. Отримана модель здатна до оцінки кількості кластерів в даних. Робота складається з теоретичного опису запропонованої моделі, її параметрів та прикладів кластерізвції багатопараметрового ЯМР медичного зображення і багатовимірного зображення земної поверхні.

1. Вступ

Серед методів клатерізації особливе місце займають алгоритми, що дають змогу оцінити кількість кластерів в даних. Оцінка кількості кластерів в даних, що аналізуються для багатьох прикладних задач розпізнання є вирішальним для висновків про належність об'єктів до однієї групи.

Актуальними задачами, які потребують алгоритмів самоорганізуючоїся кластерізації є задачі медичної діагностики, прикладні задачі економіки, дистанційні задачі геофізики та ин.

Але при великій потребі в алгоритмах даного типу їх кількість досі мала. Це пояснюється складністю побудови моделі кластерізації в якій будуть одночасно покращуватись параметри кластерів відносно міжкластерної відстані та внутрішньокластерні щільності і при цьому буде варіюватись кількість кластерів з метою знаходження глобального оптимуму. Прикладом моделі для кластерізації без вчителя є самоорганізуючі карти Кохонена [2].

В роботі представлено новий метод кластерізації "без вчителя" на основі генетичного алгоритму.

2. Метод

Вибором моделі генетичного алгоритму для вирішення задачі самоорганізуючоїся кластерізації стало те, що генетичні алгоритми – евристичні алгоритми що здатні знаходити глобальний максимум функції. Евристична модель генетичного алгоритму ґрунтується на твердженні про те, що не вигідно отримувати якісно покращенні результати за допомогою поганих початкових даних.

Нехай необхідно кластеризовати багатовимірну вибірку $D = \{d_1 .. d_n\}$ де n – кількість даних в r-мірній виборці.

Для розробки алгоритму кластерізації "без вчителя" на основі моделі генетичного алгоритму потрібно ввести наступні поняття :

1. Популяція – множина реалізацій розбиття даних на кроці ітераційного процесу.

$$Q = \{g_1 .. g_m\} \quad (1)$$

2. Ген – реалізація розбиття даних на кластери.

$$g_i = L_1 .. L_{k\_i} \quad (2)$$

k_i – кількість кластерів в даній реалізації. L_i – індекс елемента даних, що вважається центром кластера.

За допомогою генетичного алгоритму буде виконуватись ітераційний пошук оптимального розбиття на кластери. На кожному кроці ітерації оцінюються отримані розбиття і для подальшого розвитку популяції покращуються ті розбиття, що є кращими в відношенні середньої міжкластерної відстані та середньої внутрішньокластерної щільності. Отримані таким чином нові реалізації розбиття будуть заміняти ті розбиття, що оцінені як погані на даному кроці ітераційного процесу.

Генетичний алгоритм складається з операторів селекції, схрещування та мутації [1].

Селекцією називається вибір двох генів з популяції для проведення операції схрещування. При цьому імовірність того, що ген буде вибрано обраховується за формулою :

$$P(g_i) = \frac{\sqrt{P_i^{dist\,2} + P_i^{dens\,2}}}{\sum_{i=1}^{m} \sqrt{P_i^{dist\,2} + P_i^{dens\,2}}} \quad (3)$$

P_i^{dist} - імовірність того, що ген буде вибрано завдяки кращій середній міжклатерній відстані.

$$P_i^{dist} = \frac{\sum_{l=1}^{k_i}\sum_{j=1}^{k_i} dist(d_{L_j^{g_i}}, d_{L_l^{g_i}})}{k_i - 1} \quad (4)$$

$dist(d_{L_j^{g_i}}, d_{L_l^{g_i}})$ - відстань між двома кластерами.

P_i^{dens} - імовірність того, що ген буде вибрано завдяки кращій середній внутріщньокластерній щільності. При цьому для оцінки щільності в кластері використовується середньоквадратичне відхилення даних, що належать кластеру відносно центра кластера :

$$P_i^{dens} = \frac{1 - P_i^{std}}{\sum_{j=1}^{M} 1 - P_j^{std}} \quad (5)$$

Обраховується як обернена імовірності вибору гена відносно середньоквадратичного відхилення в кластері.

$$P_i^{dens} = \frac{\sum_{j=1}^{k_i} std(d_{L_j^{g_i}}, U_{L_j^{g_i}})}{k_i - 1} \quad (6)$$

$std(d_{L_j^{g_i}}, U_{L_j^{g_i}})$ -середнькавдраничне відхилення обраховане для $U_{L_j^{g_i}}$ – множини даних, що входить до кластера, відносно центра кластера $d_{L_j^{g_i}}$

Схрещування – операція, за допомогою якої знаходять нові реалізації розбиття на кластери. Операція проводиться над генами, що були вибрані за допомогою оператора селекції. Нехай вибрано два гена g^1 та g^2 виберемо один з цих генів як головний. Тобто такий, кластери якого будуть переноситись до нового гена незмінними. Нехай це g^1

Обробка головного гена. Спочатку проводять операцію видалення з розбиття таких кластерів, відсутність яких покращить середню міжкластерну відстань в реалізації та середню внутрішньокластеру щільність. Для цього обраховують обернені імовірності видалення кластерів з гена g^1 :

$$P_i^{g^1} = \frac{\sqrt{P(g^1)_i^{dist^2} + P(g^1)_i^{dens^2}}}{\sum_{i=1}^{m}\sqrt{P(g^1)_i^{dist^2} + P(g^1)_i^{dens^2}}} \quad (7)$$

де $$P(g^1)_i^{dist} = \frac{\sum_{j=1}^{k_{g^1}} dist(d_{L_i^{g^1}}, d_{L_j^{g^1}})}{\sum_{k=1}^{k_{g^1}}\sum_{j=1}^{k_{g^1}} dist(d_{L_j^{g^1}}, d_{L_k^{g^1}})} \quad (8)$$

$$P(g^1)_i^{dens} = \frac{std(d_{L_i^{g^1}}, U_{L_i^{g^1}})}{\sum_{k=1}^{k_{g^1}} std(d_{L_k^{g^1}}, U_{L_k^{g^1}})} \quad (9)$$

Кластери, що потрібно видалити із розбиття задовольняють умові :

$$P_i^{g^1} < w / k_{g^1} \quad (10)$$

де w – **коефіцієнт жорсткості видалення. w→1.**

Після видалення з головного гена поганих кластерів формується **новий ген g^*** , до якого входять всі кластери, що залишились в головному.

Синтез нових кластерів. Синтез нових кластерів проводять на основі кластерів, що вже є в новому гені та кластерів, що є в **допоміжному гені g^2.**

Для того, щоб встановити, з яких саме кластерів необхідно формувати нові кластери знаходять матрицю схрещування :

$$B = \begin{bmatrix} b_{11}..k_{1k_{g^*}} \\ b_{k_{g^2}1}..k_{k_{g^2}k_{g^*}} \end{bmatrix} \quad (11)$$

де :

$$b_{ij} = \begin{cases} 0 - \textit{немає схрещування} \\ 1 - \textit{схрещуються } i-\textit{й кластер з нового гена та } j-\textit{й кластер з допоміжного} \end{cases} \quad (12)$$

Ця матриця формується за допомогою ймовірностей схрещування :

$$P_{ij}^{cross} = P_i^{g^*} * P_j^{g^2} \quad (13)$$

де $P_i^{g^*}$ та $P_i^{g^2}$ обраховуються за формулою (7).

Після створення матриці схрещування для кожної пари кластерів активних для цієї матриці проводять схрещування за наступною схемою :

1. Не хай схрещуються два кластери задані індексами L_1 та L_2 тоді їх бінарне перетворення :

$$L_1 \to LB_1\{lb_1^1 .. lb_{b\_max}^1\} \quad (14)$$
$$L_2 \to LB_2\{lb_1^1 .. lb_{b\_max}^1\}$$

Де $b\_max = \log_2(n)$

2. Формують бінарну матрицю допустимих нащадків :

$$LB = \begin{bmatrix} lb_1^1, lb_2^2, lb_3^2, .., lb_{b\_max}^2 \\ lb_1^1, lb_2^1, lb_3^2, .., lb_{b\_max}^2 \\ \dots\dots\dots\dots \\ lb_1^1, lb_2^1, lb_3^1, .., lb_{b\_max}^2 \end{bmatrix} \xrightarrow{int} LI = \begin{bmatrix} l_1 \\ \dots\dots \\ l_{b\_max-1} \end{bmatrix} \quad (15)$$

З матриці нащадків вибирають такого нащадка, який знаходиться як надалі від батьків. Тобто до нової реалізації розбиття вноситься такий кластер, що знаходиться як надалі від кластерів, що його утворили.

3. Після отримання нового кластеру L^* його вносять до g^*

Сформований ген g^* потрібно внести до популяції – для цього з популяції виноситься ген з найгіршими характеристиками міжкластерної відстані та внутрішньокластерної щільності. Імовірність того, що ген буде видалено з популяції обернена імовірності того, що ген буде брати участь в схрещуванні. Тобто :

$$\overline{P}(g_i) = \frac{\sqrt{\overline{P}_i^{dist\,2} + \overline{P}_i^{dens\,2}}}{\sum_{i=1}^{m} \sqrt{\overline{P}_i^{dist\,2} + \overline{P}_i^{dens\,2}}} \qquad (16)$$

де

$$\overline{P}_i^{dist} = \frac{1 - P_i^{dist}}{\sum_{j=1}^{M} 1 - P_j^{dist}} \qquad (17)$$

$$\overline{P}_i^{dens\,2} = \frac{1 - P_i^{dens}}{\sum_{j=1}^{M} 1 - P_j^{sens}} \qquad (18)$$

Оператор мутації використовується для того, щоб знайдені в процесі еволюціонування популяції розбиття прагнули до глобального оптимуму. Тобто оператор мутації – цілеспрямоване внесення в популяцію випадкової інформації з метою виведення системи із локальних оптимумів. В залежності від наявних даних імовірність мутації можливо варіювати з метою отримання більш швидкої збіжності ітераційного процесу.

4. Параметри моделі.

Вхідними параметрами моделі є :

1. **Розмір популяції**. Цей параметр відповідає за швидкість сходження еволюційного процесу до стаціонарного стану. Встановлення його дуже малим буде гальмувати процес завдяки невеликій кількості реалізацій, що будуть порівнюватись між собою. Але вибір великої популяції може гальмувати час виконання ітерацій.
2. **Коефіцієнт жорсткості видалення** – відповідає за видалення "поганих" кластерів на етапі синтезу нового гена. **W∈(0,1)** Є регулювальником сходження алгоритму. При **W=1** як і при **W → 0** алгоритм буде розходитись, бо в першому випадку кількість кластерів в новому гені буде постійно зменшуватись, в другому – кількість кластерів буде постійно зростати.
3. **Імовірність мутації** – відповідає за виведення популяції з локального мінімуму. Дуже велике значення цього коефіцієнта буде лише погіршувати вже знайдені розбиття і може привести до розходження алгоритму, дуже мале значення призведе до того, що процес пошуку буде затримано в локальному оптимумі, і навіть до того, що локальний оптимум буде вибрано, як глобальний. Для більшості задач імовірність мутації 10% така ж як для природної мутації.

Важливим аспектом є встановлення критерію зупинки алгоритму. Такий критерій може бути отримано за допомогою аналізу стабільності прогресу середньої міжкластерної відстані та середнього внутрішньомластерної щільності.

5. Використання метода для кластерізації багатовимірних та багатопараметрових зображень.

Багатовимірні та багатопараметрові зображення використовуються в багатьох галузях прикладної науки. Це можуть бути медичні мультизображення чи мультизображення земної поверхні. Аналіз таких зображень з метою виділення інформації про наявність об'єктів різних типів – важлива задача діагностики.

Було оброблено багатопараметрове ЯМР медичне зображення, що складалось з трьох гомографічних зображень (рис 1). Для кластерізації було використано метрику Евкліда. Отримане кластерізоване зображення виявило наявність 49 кластерів (рис 2.а). Ітераційний алгоритм ініціювався за допомогою популяції з 10 реалізацій. Центри кластерів в початкових реалізаціях було задано випадково, максимальна кількість кластерів в початковій реалізації 20. Імовірність мутації 10%. Коефіцієнт жорсткості видалення в операторі схрещування 0.9. Зупинка ітераційного процесу проводилась за допомогою аналізу усталеності прогресу міжкластерної відстані та прогресу внутрішньокластерної щільності (рис 2.б рис 2.в). Отримано 588 ітерацій.

Також було кластерізовано багатовимірне зображення земної поверхні яке мало 3 параметри (рис 3.а). Клатерізоване зображення містить 32 кластери (рис 3.б). Розмір популяції містив 10 реалізацій. Початкова популяція формувалась випадково з максимальною кількістю кластерів в реалізації 10. Імовірність мутації 10%. Коефіцієнт жорсткості видалення в операторі схрещування 0.9. Ітераційний процес стабілізувався на 303 ітерації.

Висновки.

Використання генетичного алгоритму як базового для моделі самоорганізуючоїся кластерізації – принціпово новий підхід до вирішення проблеми побудови алгоритмів кластерізації "без вчителя".

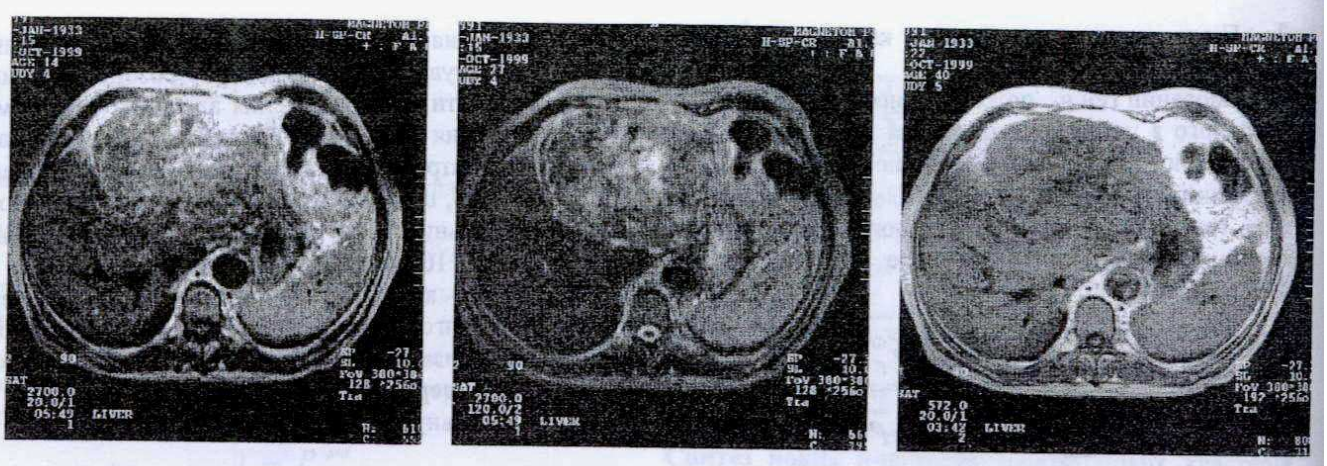


Рис. 1 Багатопараметрове ЯМР медичне зображення.

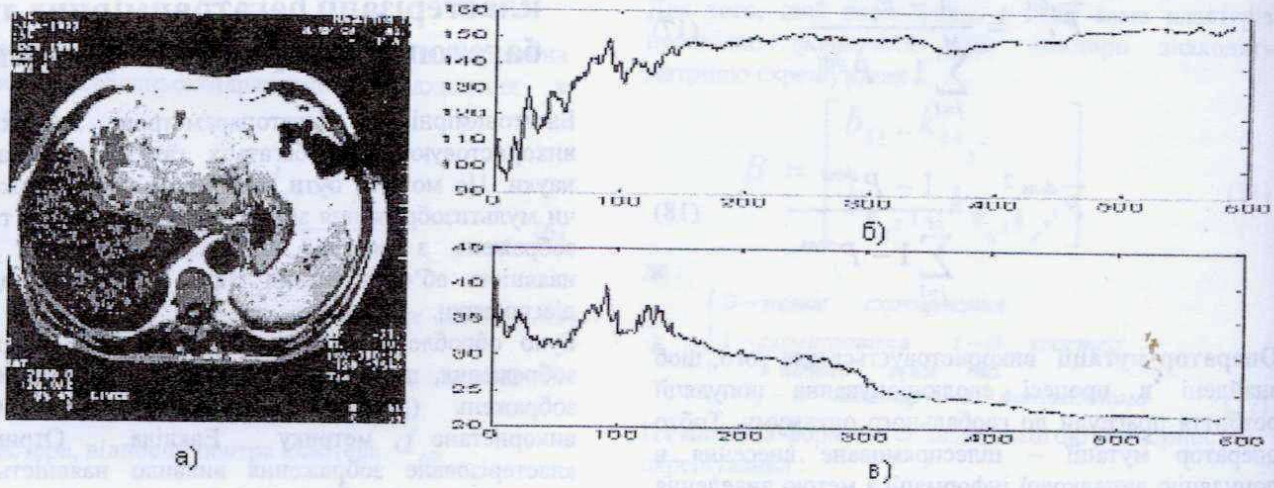


Рис. 2 Результат кластерізації багатопараметрового медичного зображення : а) нове кластері зоване зображення (49 кластерів); б) прогрес середньої міжкластерної відстані; в) прогрес середнього внутрішньокластерного середньоквадратичного відхилення.

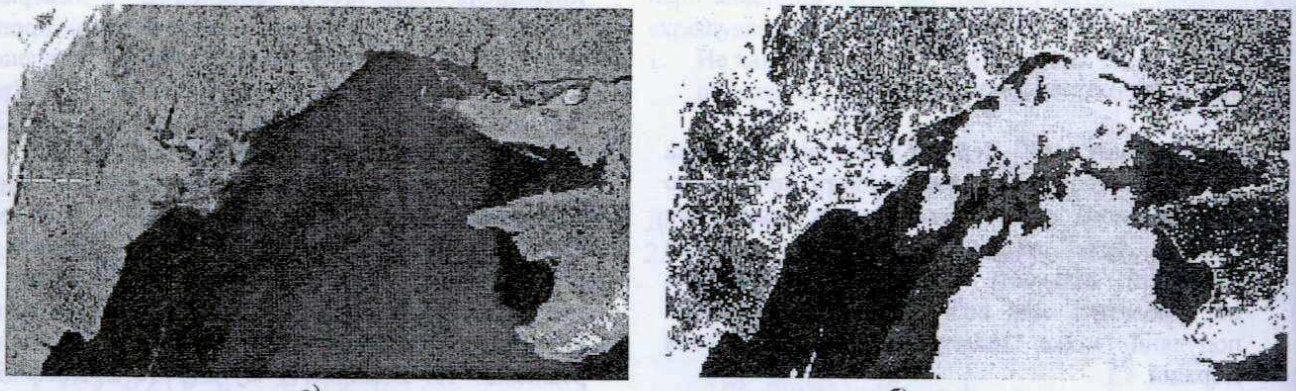


Рис. 3 Багатовимірне зображення земної поверхні (а) та нове кластері зоване зображення (б) 32 кластера.

В роботі розроблено алгоритм кластерізації багатовимірних даних "без вчителя", що базується на генетичному алгоритмі та виконана його комп'ютерна реалізація для медичних зображень та зображень земної поверхні.

Література

1. Goldberg D. E. Genetic algorithms in search, optimization and machine learning. Addison Wesleg, MA. 1989.
2. Kohonen D. Self-organizing maps. NY. Springer-Verlag, 1995

СИНТЕЗ ЗОБРАЖЕНЬ, ЩО ХАРАКТЕРИЗУЮТЬСЯ ТОЧКОВИМИ ТИПАМИ СИМЕТРІЇ

В.М.Хавалко

Державний науково-дослідний інститут інформаційної інфраструктури, м.Львів-31, а/с 5446, Тел./факс (0322) 63-03-11, victorkh@dndiii.lviv.ua

В статті розглянуто задачу синтезу зображень і складних геометричних об'єктів (СГО), які характеризуються точковими типами симетрії. Для розширення множини синтезованих зображень як непохідні елементи запропоновано використовувати фрактальні криві різних розмірностей та рівнів деталізації. Побудовано та програмно реалізовано інформаційно-аналітичну систему (ІАС) синтезу зображень, в тому числі і симетричних.

ВСТУП

В теорії і практиці побудови систем розпізнавання, оцінки, класифікації і синтезу зображень зустрічаються зображення різного походження, фізичної природи і структури. У зв'язку з цим розглядають прості і складні зображення. До простих зображень відносять зображення, які не підлягають подальшому розбиттю на простіші і використовуються як непохідні елементи для синтезу і опису зображень складнішої структури і геометричної форми. Досить часто такі зображення характеризуються тим чи іншим типом симетрії. Для побудови складних зображень та опису взаємозв'язків між структурними елементами таких зображень прийнято використовувати R-функції [1], а також структурний [2] і функціональний [3] підходи. Однак застосування непохідних елементів, які описуються засобами традиційної Евклідової геометрії, не може забезпечити достатньої різноманітності та повноти множини синтезованих зображень, а тому доцільно доповнити цю множину непохідними елементами, які не є традиційними в сенсі Евклідової геометрії. Такими непохідними елементами є криві та поверхні фрактальної геометрії [4, 5], при побудові яких прийнято використовувати рекурсивний підхід [6, 7] та системи ітераційних функцій [8, 9].

1. ТОЧКОВІ ТИПИ СИМЕТРІЇ

Відповідно до основних означень [1], якщо O – об'єкт в R^n, а $Q: R^n \to R$ перетворення, яке визначається формулою

$$x' = Qx = Ax + b,$$

де $x = (x_1, \ldots, x_n)$, $x' = (x_1', \ldots, x_n')$, $b = (b_1, \ldots, b_n)$,

$$A = \begin{pmatrix} a_{11} & \ldots & a_{1n} \\ \ldots & \ldots & \ldots \\ a_{n1} & \ldots & a_{nn} \end{pmatrix},$$

то об'єкт $O \subset R^n$ характеризується деяким типом симетрії, якщо існує хоча б одне перетворення $Q \in \{Q_i\}$, відмінне від одиничного перетворення, таке що $Q * O = O$.

Крім цього, якщо $f = 0$ – рівняння СГО O, який симетричний відносно перетворення Q, то вважаємо, що рівняння $f = 0$ має такий самий тип симетрії (тобто $f = 0$ симетричне відносно Q), якщо $f(Qx) = f(x)$.

Рівняння $f = 0$ будемо називати кососиметричним відносно Q, якщо

$$f(Qx) = -f(x).$$

Загальноприйнята така класифікація за типами симетрій:

- Точкові типи симетрії:
  - симетрія відносно точки;
  - симетрія інверсії;
  - симетрія відносно площини;
  - симетрія відносно осі.
- Симетрії трансляційного типу.

Перетворення симетрії, які зберігають нерухомими хоча би одну точку простору прийнято називати точковими. Очевидно, що будь-яке точкове перетворення має вигляд

$$x' = Ax. \qquad (1)$$

В інших випадках говорять про симетрію трансляційного типу.

1.1. Симетрія відносно осі

Для синтезу СГО, які характеризуються симетрією відносно осі, запропоновано новий алгоритм, який базується на матриці перетворення симетрії порядку N відносно осі Ox_1.

$$C_{x,N} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos 2\pi/N & \sin 2\pi/N \\ 0 & -\sin 2\pi/N & \cos 2\pi/N \end{pmatrix}.$$

Означення 1. СГО O є симетричним відносно прямої C, яка називається віссю симетрії, якщо після повороту на кут $\alpha_N = 2\pi/N$, де N – ціле число більше одиниці, він займає початкову область простору. N називається порядком симетрії осі.

1.2. Симетрія інверсії

Симетрії цього типу відповідає перетворення (1) з матрицями:

$$A_1 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{pmatrix};\ A_2 = \begin{pmatrix} -1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{pmatrix};$$

$$A_3 = \begin{pmatrix} -1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 1 \end{pmatrix}. \qquad (2)$$

Якщо $f(x) = f(x_1, x_2, x_3) = 0$ – рівняння СГО або зображення O, який володіє симетрією в сенсі її інваріантності відносно перетворень (2). При цьому саме рівняння $f(x) = 0$ відносно перетворень (2) може бути і не симетричним. Очевидно, що рівняння $f(x_1, -x_2, -x_3) = 0$ також буде рівнянням СГО O при $i = 1$, $f(-x_1, x_2, -x_3) = 0$ при $i = 2$ і $f(-x_1, -x_2, x_3) = 0$ при $i = 2$. Тоді симетричні рівняння для A_1, A_2 та A_3 відповідно матимуть вигляд:

$$f_1(x) = 0.5 * [f(x_1, x_2, x_3) + f(x_1, -x_2, -x_3)] = 0,$$
$$f_2(x) = 0.5 * [f(x_1, x_2, x_3) + f(-x_1, x_2, -x_3)] = 0, \qquad (3)$$
$$f_3(x) = 0.5 * [f(x_1, x_2, x_3) + f(-x_1, -x_2, x_3)] = 0.$$

Замість операції додавання, згідно з [1], в рівняннях (3) можна використовувати довільний тип операцій об'єднання або перетину, які грунтуються на R-функціях.

1.3. Симетрія відносно площини

Симетрії цього типу відповідає перетворення (1) з матрицями:

$A_1 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{pmatrix}$ – для СГО, які симетричні відносно площини x_1Ox_2;

$A_2 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$ – для СГО, які симетричні відносно площини x_1Ox_3;

$A_3 = \begin{pmatrix} -1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$ – для СГО, які симетричні відносно площини x_2Ox_3.

Методики описані в попередніх пунктах, справедливі для цього типу точкової симетрії і є загальним прикладом, який використовується в більшості випадків. Однак в плані обчислювальних затрат цей тип симетрії приводить до громіздких виразів, що ускладнює реалізацію в реальному часі.

1.4. Симетрія відносно точки

Синтез зображень симетричних відносно точки [1] полягає в побудові такого рівняння $f(x_1, x_2, x_3) = 0$, для якого справедливі такі припущення.

Якщо $f_1(x_1, x_2, x_3) = 0$ – довільне рівняння СГО O_1, яке характеризується симетрією відносно перетворення (1), де

$$A = \begin{pmatrix} -1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{pmatrix},$$

але саме рівняння не є симетричними відносно цього перетворення, то рівняння СГО O_1 можна записати у вигляді

$$f(x_1, x_2, x_3) = 0.5 * [f(x_1, x_2, x_3) + f(-x_1, -x_2, -x_3)] = 0$$

і це рівняння буде симетричним відносно типу симетрії, що розглядається. Інший підхід до побудови зображень, які характеризуються симетрією відносно точки і який також базується на властивостях R-функцій, розглянуто в [1].

2. НЕПОХІДНІ ЕЛЕМЕНТИ ФРАКТАЛЬНОЇ ГЕОМЕТРІЇ

Сучасні методи синтезу складних зображень використовують дійсні функції для опису елементарних об'єктів до яких застосовується ряд операцій, для синтезу більш складних зображень. Задання або вибір множини елементарних об'єктів, а також опис основних операцій, які дозволяють проводити перетворення над об'єктами, створювати плавні об'єкти та ін. належать до основних задач при синтезі складних зображень [10, 11].

В цій роботі пропонується доповнити традиційну, в сенсі Евклідової геометрії, множину непохідних елементів основними типами кривих фрактальної геометрії. Насамперед, це продиктовано тим, що фрактальна геометрія серед рекурсивних підходів є глобальним та потужним засобом, а протягом багатьох років розвитку розроблено велику кількість класів і типів методів та алгоритмів для роботи з СГО, які характеризуються високою ступінню складності і для опису яких потрібно застосовувати інший апарат ніж Евклідова геометрія.

До основних типів фрактальних кривих та об'єктів належать: просторово-заповнюючі і традиційні лінійні криві, а також різноманітна множина тривимірних об'єктів (губка Менгера, трикутник Серпінського, множини Жуліа та ін.).

До традиційних одновимірних фрактальних кривих, які широко використовуються при синтезі складних природних структур, належать крива Коха, $3/2$ - крива та ін. Ці, а також багато інших фракталів відносяться до класу геометричних фракталів, які є найбільш наочними і зрозумілими. Поняття елементарного об'єкта в фрактальній геометрії дещо відрізняється від такого ж поняття в інших підходах синтезу та представлення зображень. Суть процесу побудови фрактальних кривих полягає в тому, що на початку задається відрізок певної довжини (ініціатор), кожна лінійна частина якого на подальших етапах побудови замінюється генератором, який з допомогою коефіцієнтів масштабування набуває потрібних розмірів [12]. На наступних етапах побудови кожен відрізок ламаної попереднього етапу замінюється генератором зменшеним в 4 рази для $3/2$ -кривої і в 3 рази для

кривої Коха. Процес побудови триває до тих пір, поки не досягається задовільна точність, або, коли зміни при переході від одного етапу до іншого стають непомітними для ока спостерігача.

3. СИСТЕМИ ІТЕРАЦІЙНИХ ФУНКЦІЙ

Ще одним способом побудови фракталів є задання систем ітераційних функцій (СІФ). Систем, які відображають одну багатовимірну множину в іншу. Найбільш проста система ітераційних функцій складається з афінних перетворень площини:

$$u = a * x + b * y + e$$
$$v = c * x + d * y + f.$$

Отже, зображення кодується декількома простими перетвореннями (в нашому випадку афінними), тобто коефіцієнтами цих перетворень.

Якщо систему ітераційних функцій замінити оператором Хатчінсона W, то ітераційний процес можна записати наступним чином

$$A_{k+1} = W(A_k),\ k = 0,1,2,\ldots,$$

де A_0 – початкова множина (зображення), а $\{A_k\}$ – послідовність множин (зображень), отримана багаторазовим застосуванням W. Отримуємо послідовність, яка прямує до кінцевого зображення A_∞, яке називається атрактором СІФ і залишається інваріантним щодо системи ітераційних функцій. В термінах W це означає

$$A_\infty = W(A_\infty).$$

Кінцеве зображення (атрактор) A_∞ і буде фракталом.

4. ПРОГРАМНА РЕАЛІЗАЦІЯ СИСТЕМИ СИНТЕЗУ СИМЕТРИЧНИХ ЗОБРАЖЕНЬ

На основі концепцій побудови ІАС синтезу зображень[13], розбиття таких систем на рівні абстракції [14], а також запропонованих алгоритмів синтезу СГО симетричної структури створено ІАС синтезу складних зображень і алгоритмічне забезпечення до цієї системи. В розробленій ІАС задача синтезу розв'язується шляхом виконання двох наступних етапів: 1) побудова аналітичного опису СГО, використовуючи множини елементарних геометричних об'єктів, операцій та відношень [10, 11]; 2) візуалізація синтезованого СГО або інше застосування отриманого аналітичного виразу.

Користувач ІАС взаємодіє з системою за допомогою найвищого рівня абстракції – графічного інтерфейсу користувача (ГІК). Ця взаємодія відбувається через спроектовану мову, яка може бути як текстовою або візуальною, так і поєднувати ці два види. На цьому рівні доступні засоби для побудови, модифікації, архівації і знищення як простих, так і складних об'єктів, а також присутні засоби для аналізу проектування.

Побудована ІАС функціонує таким чином. Нехай потрібно синтезувати чи відтворити СГО. Користувач, за допомогою ГІК, вибирає непохідні елементи та операції чи відношення з відповідних бібліотек. Кожен непохідний елемент визначений за допомогою параметризованої функції дійсних змінних. Відповідні параметри, під час вибору елемента з бібліотеки, в інтерактивному режимі задає користувач. Для підвищення ефективності роботи системи користувач повинен чітко знати логіку побудови СГО, який він хоче синтезувати. Якщо потрібного непохідного елемента в бібліотеці не має, то користувач, використовуючи СІК, може доповнити бібліотеку необхідним елементом. Для цього передбачено використання стандартних математичних функцій ('exp', 'sin', 'sqr', 'sqrt', 'cos' та ін.). Таким чином, користувач, за допомогою ІАС, задає аналітичну конструкцію, яка описує синтезований СГО. При цьому основні теоретико-множинні операції задаються використовуючи зарезервовані символи ('|' – об'єднання, '&' - перетин, '\' - віднімання, '~' - заперечення, '@' – Декартовий добуток та ін.). Згідно з заданою аналітичною формулою та вибраною гамою кольорів ІАС синтезує зображення. При необхідності, користувач може побудувати поверхню, яка відповідає синтезованому СГО.

У випадку синтезу СГО симетричної структури користувач ІАС, використовуючи вікно вибору типів симетрії (рис.1), має можливість вибрати відповідний тип симетрії, який буде притаманним синтезованому СГО.

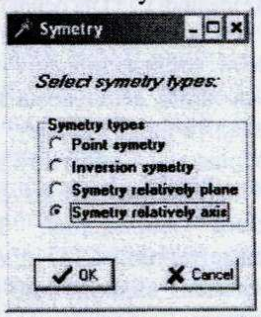


Рис.1. Вікно вибору типів симетрії

Після цього користувачу необхідно задати початкові координати синтезованого зображення (рис.2), а також порядок симетрії, тип непохідних елементів, які при цьому будуть використовуватись і рівень рекурсії у випадку вибору фрактальних кривих як непохідних елементів (рис.3).

Рис.2. Вікно задання початкових координат

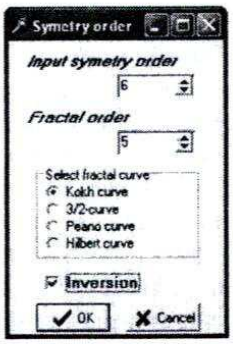


Рис.3. Вікно задання параметрів синтезованого СГО

Множина побудованих, за допомогою ІАС, зображень і СГО симетричної структури (рис.4, рис.5), які характеризуються широкою різноманітністю і крім цього СГО є інваріантними щодо різних форм представлення непохідних елементів.

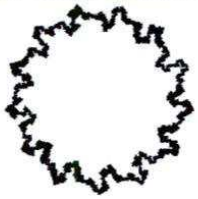


Рис.4. Синтезований СГО на основі $3/2$ -кривої

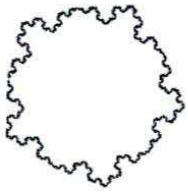


Рис.5. Синтезований СГО на основі кривої Коха

Синтезовані зображення зручні в зберіганні, адже описуються лише декількома коефіцієнтами – параметрами аналітичного виразу, що відповідає даному типу зображень. Саме типу зображень тому, що змінюючи параметри отримуємо можливість синтезувати різні зображення. Так на рис.6. зображено осесиметричний СГО п'ятого порядку, а як непохідний елемент використано криву Гільберта четвертого порядку. Крім цього візуалізація синтезованих зображень спрощується завдяки внутрішньому представленню побудованих аналітичних виразів.

Рис.6. Осесиметричний СГО на основі кривої Гільберта

ВИСНОВКИ

Запропоновані алгоритми поєднують елементи функціонального, структурного і рекурсивного підходів. Саме таке поєднання призводить до різкого зменшення обчислювальних затрат, та затрат, які необхідні на зберігання синтезованих СГО і зображень. Застосування функціонального підходу до синтезу зображень складної геометричної форми і структури , а систем ітераційних функцій – до опису непохідних елементів таких синтезованих СГО і зображень дає можливість досягнути високих коефіцієнтів стиску і дозволяє ефективним чином кодувати отримувані зображення.

ЛІТЕРАТУРА

1. Рвачов В.Л. *Теория R–функций некоторые ее приложения*.- Киев: Наукова думка,1982.- 456 с.
2. Александров В.В., Горский Н.Д. *Алгоритмы и программы структурного метода обработки данных*.- Ленинград: Наука, 1983.- 208 с.
3. Pasko A., Adzhiev V., Sourin A., Savchenko V. *Function representation in geometric modelling: concepts, implementation and applications* // The Visual Computer.- 1995.- vol.11.- №8.- P.429-446.
4. Mandelbrot Benoit B. *The Fractal Geometry of Nature*.- WH Freeman and company, 1977.- 468P.
5. Massopust P .R. *Fractal functions, fractal s urfaces, and wavelets*.- San Diego, Calif.: Acad. Press, 1994.- 383 P.
6. Александров В.В., Горский Н.Д. *Представление и обработка изображений. Рекурсивный подход*.- Ленинград: Наука,1985.- 192с.
7. Мальцев А.И. *Алгоритмы и рекурсивные функции*.- М.:Наука, 1986.- 367с.
8. Michael Barnsley *Fractals everywhere*.- Academic press: Jones and Bartlett Publishers, 1988.- 394P.
9. Peitgen H.-O., Jürgens H., Saupe D. et. al. *Fractals for the Classroom: Strategic Activities*.- New York: Springer-Verlag, 1999.- Vol. Two.- 187 P.
10. Хавалко В.М. *Використання теоретико-множинних операцій для синтезу та перетворень геометричних об'єктів* // Вісник ДУ "Львівська політехніка": Комп'ютерна інженерія та інформаційні технології.- 1999.- №386.- С.164-172.
11. Хавалко В.М. *Проблема вибору множини елементарних об'єктів для ефективного синтезу складних зображень* // Вісник ДУ "Львівська політехніка": Комп'ютерна інженерія та інформаційні технології.- 2000.- №392.- С.131-136.
12. Peitgen H.-O., Richter P.H. *The beauty of fractals: images of complex dynamical systems*.- Berlin: Springer, 1986.- 199 P.
13. Хавалко В.М. *Інформаційно-аналітична система синтезу та обробки зображень: концептуальний підхід* // Труды Одесского политехнического университета.- 2002.- Вып. 4(16).- С.86-90.
14. Хавалко В.М., Манишин І.Р. *Інформаційно-аналітична система як ефективний засіб синтезу складних зображень* // Комп'ютерні технології друкарства.- 2000.- №4.- С.301-308.

СЕГМЕНТАЦІЯ ДАНИХ З ВИКОРИСТАННЯМ ВЕЙВЛЕТ-АНАЛІЗУ

Г.М. Востров, М.В. Полякова, В.В. Любченко
ОНПУ, 65044, Україна, Одеса, пр. Шевченка, 1, (0482) 288-478, (0482) 685-152,
vvl@lmpo.ospu.odessa.ua

У представленій доповіді розглядається сегментація економічних часових рядів за допомогою вейвлет-аналізу. У якості меж сегментів використовуються моменти зміни властивостей випадкового процесу, що описує часовий ряд. Отримана сегментація використовується для підвищення ефективності виділення трендів економічних часових рядів.

1. ВСТУП

При вирішенні задач аналізу і обробки економічних часових рядів, зокрема при побудові ефективних алгоритмів стиснення або прогнозування, виникає проблема вивчення внутрішньої структури часових рядів. Під внутрішньою структурою будемо розуміти безліч схожих сегментів часового ряду із заданою на йому сукупністю n-арних відносин.

Ефективність структурного аналізу часових рядів залежить від точності оцінки коефіцієнтів тренда, тому вирішення задачі аналізу часових рядів передусім пов'язане з його виділенням. Автори досліджували часові ряди з властивостями, що стрибкоподібно змінюються. Момент зміни властивостей часового ряду відповідає точці розриву тренда. Тому використання безперервної функції для апроксимації тренда в цій ситуації некоректне. Необхідно використати кусочну апроксимацію у вигляді сплайнового тренду. Для цього передусім необхідно виконати виявлення змін властивостей випадкового процесу.

При аналізі сигналів, що характеризують випадкові процеси, можна використати вейвлет-перетворення, яке дає можливість отримати спектральну інформацію на вибраній частоті, враховуючи при цьому, що частота або амплітуда сигналу можуть змінюватися.

2. МАТЕМАТИЧНА МОДЕЛЬ ВИПАДКОВОГО ПРОЦЕСУ

Часові ряди, що описують динаміку зміни цін на фондовому ринку, можна розглядати як одновимірні випадкові процеси з властивостями, що змінюються. На нашій думку ці сигнали можна описати з використанням наступної моделі. Передбачимо, що випадковий процес можна представити у вигляді сукупності сегментів даних з властивостями, що змінюються. На кожному з сегментів часовий ряд являє собою нестаціонарний процес, що повільно міняється, зі стаціонарними або близькими до стаціонарних приростами, причому характеристики процесу на межах сегментів стрибкоподібно міняються. Будемо вважати, що на кожному з сегментів, де властивості ряду міняються повільно, часовий ряд описується моделлю [1]:

$$y(t)=c(t)+\sum_i a_i \sin(\omega_i t+\theta_i)+n(t) \qquad (1)$$

где $c(t)$ — тренд,

a_i, ω_i, θ_i — амплітуда, частота, зсув по фазі i-ої циклічної компоненти,

$n(t)$ — іррегулярна компонента,

t — деякий параметр, наприклад, час або просторова координата, який може приймати дискретні значення t=1, …, N. Вважається, що подібне уявлення (1) справедливе в межах кожного з сегментів, де властивості ряду міняються повільно. Тоді точками зміни властивостей часового ряду будемо вважати точки, в яких спостерігаються стрибки трендової компоненти, а також параметрів циклічних компонент — амплітуди, частоти. Передбачимо, що в якості іррегулярної компоненти виступає білий шум, а також, що різкі зміни розділені у часі досить, щоб були можливі їх індивідуальні виявлення.

При вирішенні задачі виявлення зміни властивостей сигналів передусім перевіримо, чи адекватна модель (1) реальним даним. Модель (1) передбачає, що ряд містить тренд і циклічні компоненти. Для перевірки даних на наявність тренду звичайно перевіряють, чи містить спектральна щільність пік на нульовій частоті. Щоб перевірити, чи присутні в даних періодичні компоненти, можна, виділивши тренд, перевірити, чи містить значущі піки спектральна щільність.

Перевірка адекватності моделі (1) реальним даним проводилася для часових рядів, що описують динаміку зміни цін на фондовому ринку. Спектральна щільність сигналів до виділення тренда, як правило, має пік на нульовій частоті, говорячий про наявність тренду, а спектральна щільність сигналів після виділення тренда має піки, що свідчать про наявність циклічних компонент.

Перевірка фрагментів сигналу між стрибками показала, що вони мають стаціонарні прирости. Цей факт, а також приведені вище міркування дозволяють зробити висновок про адекватність часових рядів, що описують динаміку

зміни цін на фондовому ринку, моделі (1). У цій моделі повільна зміна властивостей сигналу виражена трендом, а як стрибкоподібні зміни будемо розглядати зміну частоти і амплітуди циклічних компонент.

3. ВИЯВЛЕННЯ ЗМІН ВЛАСТИВОСТЕЙ ВИПАДКОВОГО ПРОЦЕСУ

3.1. Виявлення стрибків середнього

Задача визначення моменту стрибкоподібної зміни трендової компоненти (стрибка середнього) розглядалася нами в [2]. Якщо для перетворення Фурье ми маємо тільки необхідну умову глобальної регулярності, то, використовуючи вейвлет-перетворення, можна зробити висновок про локальну і глобальну регулярність сигналу. При реалізації вейвлет-методів виявлення зміни властивостей намагаються оцінити локальну асимптотичну поведінку вейвлет-перетворення $T(b,a)$ початкового сигналу для різних x при $a \to 0$, де a — рівень (масштаб) вейвлет-перетворення [3]. Для того, щоб уникнути цієї проблеми, пропонуємо замість оцінки асимптотики використати порогові значення для кожного рівня вейвлет-перетворення. Крім того, в [4] показано, що при визначенні локальної регулярності функції краще використати розкладання по надмірних вейвлет-базисах, в яких неінваріантність до зсуву відсутна (безперервний випадок) або спостерігається в меншій мірі (дискретний випадок), ніж при використанні звичайного дискретного вейвлет-перетворення (ДВП). Тому ми будемо використати стаціонарне вейлет-перетворення (СВП). Це перетворення використовувалося для знаходження точок сингулярності в [2] і дало непогані результати.

Визначимо поріг для ідентифікації значущих максимумів СВП, і, отже, значущих стрибків сигналу. Розглянемо функцію що описує ідеальний стрибок:

$$f(x) = \begin{cases} 1, x > 0, \\ 0, x \le 0. \end{cases}$$

Згідно [3], вейвлет-перетворення функції $f(x)$ має вигляд

$$T(b,a) = \int_{-b/a}^{\infty} \psi(t)dt, \qquad (2)$$

де $\psi(t)$— вейвлет-функція. У випадку $f(x) = \begin{cases} A_1, x > x_0, \\ A_0, x \le x_0. \end{cases}$ внаслідок заміни змінних в (2) маємо $T(b,a) = A \int_{(x_0-b)/a}^{\infty} \psi(t)dt$, де $A=A_1-A_0$ — величина стрибка в точці x_0. При вирішенні задачі виявлення зміни властивостей, коли величина стрибка невідома, одним з підходів є апріорне завдання мінімальної величини стрибка. Визначимо мінімальний розмір стрибка α_{thr}, наприклад, в процентах від $\max_{t=1,\dots,N} y(t) - \min_{t=1,\dots,N} y(t)$, где $y(t)$, $t=1, \dots, N$ — відліки часового ряду.

Тоді як порогову величину приймемо

$$T(a) = \alpha_{thr} \max_b \left| \int_{(x_0-b)/a}^{\infty} \psi(t)dt \right|, \qquad (3)$$

(3) дозволяє обчислити поріг для визначення значущих максимумів, а оцінка моменту зміни прийме вигляд:

$$\hat{x}_0 = 2^j K + b,$$

де $b = \arg\max_b \int_{(x_0-b)/a}^{\infty} \psi(t)dt$.

3.2. Виявлення моментів змін параметрів циклічних компонент

Розглянемо задачу визначення моменту стрибкоподібної зміни параметрів циклічної компоненти, що має вигляд $h(x)=A\sin(\omega_0 x)$, где A — амплітуда, ω_0 — частота синусоїди, x — просторова координата або час. Якщо вважати, що вейвлет $\psi(x)$ антисимметричен, то вейвлет-перетворення синусоїди $h(x)$ [3]:

$$|Th(b,a)| = A\left|\cos\omega_0 b\right| \left|\hat{\psi}(a\omega_0)\right| \qquad (4)$$

где $\hat{\psi}(\omega)$ — перетворення Фур'є вейвлет-функції $\psi(x)$, a — масштаб вейвлет-перетворення, b — параметр зсуву.

Знайдемо поріг для визначення моменту стрибкоподібної зміни амплітуди циклічної компоненти. Нехай амплітуда синусоїди стрибкоподібно змінюється з A_0 на A_1 та задана відносна зміна амплітуди синусоїди: $\delta = \frac{A_0}{A_1}$. Без обмеження спільності можна вважати, що $0<\delta<1$, тобто $A_0<A_1$. Тоді відносну величину стрибка амплітуди можна оцінити в термінах стандартного відхилення вейвлет-коефіцієнтів. Нехай $std_0(T)(a)$ — стандартне відхилення вейвлет-коефіцієнтів до стрибка, $std_1(T)(a)$ — стандартне відхилення вейвлет-коефіцієнтів після стрибка. Будемо вважати, що сталася зміна амплітуди циклічної компоненти з частотою, відповідною даному масштабу вейвлет- перетворення, якщо

$$\Delta std(T)(a) = \frac{std_0(T)(a)}{std_1(T)(a)} =$$

$$\frac{A_0\, std(|\cos(\omega_0 b)|)\,|\hat{\psi}(a\omega_0)|}{A_1\, std(|\cos(\omega_0 b)|)\,|\hat{\psi}(a\omega_0)|} = \frac{A_0}{A_1} \le \delta$$

Знайдемо поріг для визначення моменту стрибкоподібної зміни частоти циклічної компоненти. Виберемо такий ортонормований вейвлет-базис, що функція $|\hat{\psi}(\omega)|$ має тільки один

глобальний максимум при ω>0, тоді вона монотонно спадає по обидві сторони цього максимума. Цією властивістю володіють, наприклад, вейвлети Добеші з *мінімальною фазою. З урахуванням цієї* властивості, а також рівності (4) можна затверджувати, що стрибкоподібна зміна частоти синусоїди з ω_0 на ω_1 спричиняє за собою зміну стандартного відхилення вейвлет-перетворення на всіх масштабах, для яких $a \in supp|\hat{\psi}(a\omega_0)|$ або $a \in supp|\hat{\psi}(a\omega_1)|$, де *supp* — носій функції, причому максимальний стрибок стандартного відхилення буде спостерігатися для масштабів a близьких до ω_m/ω_0 і ω_m/ω_1, де $\omega_m = arg\max_{\omega}|\hat{\psi}(\omega)|$.

Нехай задана відносна зміна частоти синусоїди: $\gamma = \frac{\omega_0}{\omega_1}$. Без обмеження спільності можна вважати, що $0<\gamma<1$, тобто $\omega_0<\omega_1$. Оцінимо відносну величину зміни частоти синусоїди у термінах стандартного відхилення вейвлет-коефіцієнтів. Нехай $std^0(T)(a)$ — стандартне відхилення вейвлет-коефіцієнтів до зміни, $std^1(T)(a)$ — стандартне відхилення вейвлет-коефіцієнтів після зміни.Тоді

$$\Delta std(T)(a) = \frac{std^0(T)(a)}{std^1(T)(a)} = \frac{A\, std(|\cos(\omega_0 b)|)\,|\hat{\psi}(a\omega_0)|}{A\, std(|\cos(\omega_1 b)|)\,|\hat{\psi}(a\omega_1)|}$$

Врахуємо, що $std(|\cos(\omega_0 b)|) = std(|\cos(\omega_1 b)|)$. Тоді

$$\Delta std(T)(a) = \frac{|\hat{\psi}(a\omega_0)|}{|\hat{\psi}(a\omega_1)|}. \qquad (6)$$

Для подальшого аналізу формули (6) виразимо значення $|\hat{\psi}(a\omega)|$ через a і ω. Зазначимо, що вибраний нами для аналізу вейвлет Добеши [4] другого порядку з мінімальною фазою не має аналітичного уявлення, значення цієї функції можна обчислити за допомогою рекурентної формули з будь-якою мірою точності. Аналогічні проблеми виникають і при обчисленні її модуля перетворення Фур'є $|\hat{\psi}(\omega)|$. Тому як аналітичний вираз для $|\hat{\psi}(\omega)|$ будемо використати апроксимацію цієї функції, враховуючи при цьому графічне уявлення $|\hat{\psi}(\omega)|$. На відрізку [0, 0.5] $|\hat{\psi}(\omega)|$ будемо апроксимувати квадратичною функцією $|\hat{\psi}(\omega)|_{[0,\ 0.5]} \sim \omega^2$, на відрізку (0.5, 4.5] — лінійною функцією: $|\hat{\psi}(\omega)|_{(0.5,\ 4.5]} \sim \omega$. У випадку $\omega>4.5$ функцію $|\hat{\psi}(\omega)|$ будемо аппроксимувати гіперболою: $|\hat{\psi}(\omega)|_{\omega>4.5} \sim 1/\omega$. Числа 0.5 и 4.5 виражають частоту в радіанах і отримані евристично, виходячи з графічного уявлення $|\hat{\psi}(\omega)|$.

Виходячи з приведених вище міркувань, будемо вважати, що зміна частоти циклічної компоненти сталося, якщо

$$\Delta std(T)(a) \le \gamma^n,$$

де γ визначається як ω_0/ω_1, n залежить від рівня вейвлет-перетворення i таким чином: $n = \begin{cases} 2, & i = 1, \\ 1, & i = 2, \\ -1, & i > 2. \end{cases}$

У разі виявлення зміни обох параметрів циклічної компоненти — амплітуди і частоти, як порогову величину можна використати

$$\delta\gamma^n. \qquad (7)$$

3.3. Особливості алгоритму сегментації даних

Розроблений алгоритм сегментації даних буде мати наступні дві особливості. Перша особливість буде полягати в тому, що цей алгоритм буде багато масштабним. Це пов'язано передусім з тим, що необхідно визначити моменти зміни параметрів циклічних компонент, які характеризуються різними частотами і, отже, вимагають розгляду поведінки сигналу на різних частотних смугах. До того ж, для більшості задач виявлення змін, що використовують єдиний масштаб, типова суперечність, виникаюча при виборі вікна. Велике вікно дає можливість дістати більш точні статистичні оцінки. Мале вікно дозволяє збільшити швидкодію, більш точно оцінити момент зміни, а також визначати близько розташовані моменти змін. Щоб обійти цю проблему, виконується сегментація на різних масштабах, а потім результати сегментацій об'єднують по деякому критерію з використанням оптимізаційних алгоритмів. При цьому використовують ДВП або який-небудь інший пірамідальний алгоритм і об'єднання результатів сегментацій проводять, починаючи з найнижчого рівня, використовуючи співвідношення між батьківськими і дочірніми блоками коефіцієнтів піраміди або вейвлет-розкладання. Відомо, що сегментації на низьких рівнях мають мале число помилкових меж сегментів і добре характеризують великі однорідні регіони, в той час як сегментації на високих рівнях дозволяють більш точно локалізувати межі сегментів навіть у випадку близько розташованих меж. Тому при об'єднанні результатів сегментацій на різних масштабах, за основу приймають результат сегментації на самому великому масштабі, поступово уточнюючи межи сегментів за допомогою результатів сегментацій на малих масштабах.

У розробленому алгоритмі виявлення змін, ми використали СВП, яке дозволяє більш точно локалізувати межі сегментів вже на низьких масштабах в порівнянні з ДВП. До того ж нам потрібно виявити зміни властивостей сигналу на різних частотних смугах, тому буде не зовсім коректно приймати за основу результат сегментації

на самому низькому рівні, уточнюючи межі сегментів за допомогою сегментацій на малих масштабах. У зв'язку з цим при об'єднанні результатів сегментацій на різних масштабах використовувався наступний евристичний підхід.

1. Виконувалася сегментація на самому високому рівні СВП.
2. При виявленні змін властивостей на наступному рівні СВП перевірялося, чи не попадає момент зміни в місцевість $\delta = \sum_{j=1}^{j_0} k_j$, де j_0 — поточний рівень СВП, k_j —довжина фільтра на рівні j, якої-небудь точки зміни середнього, записаної на попередньому рівні, або в околицю, рівної розміру вікна L, якої-небудь точки зміни стандартного відхилення, записаної на попередньому рівні.
3. Якщо умова виконується, то відповідний момент зміни властивостей записується у вектор сегментації.

Таким чином, в нашому випадку ми використали об'єднання результатів сегментації на різних масштабах в зворотному порядку: від більш високих рівнів СВП до більш низьких. Щоб уникнути виникнення помилкових тривог (пересегментації), ми будемо використовувати два вигляду порогів, абсолютний і відносний, та робастну оцінку стандартного відхилення. Це ж дозволило уникнути використання оптимізаційної процедури, що підвищує обчислювальну ефективність алгоритму.

Другою особливістю алгоритму виявлення змін властивостей випадкового процесу буде його універсальність в тому значенні, що він буде виконувати виявлення як по диференційних, так і по інтегральних ознаках. Це дозволить врахувати представлення єдиної реалізації випадкового процесу у вигляді детермінований функції часу і виявляти її сингулярності (диференційний підхід), а також використати ймовірностно-статистичні характеристики випадкового процесу (інтегральний підхід).

Тоді алгоритм сегментації буде складатися з трьох основних частин:

1. Виявлення змін з використанням диференційної ознаки (моментів змін середнього з порогом (3)).
2. Виявлення змін з використання інтегральної ознаки (моментів змін стандартного відхилення з порогом (7)).
3. Основний алгоритм, який об'єднує сегментації по обох ознаках, а також виконує об'єднання результатів сегментацій на різних масштабах.

4. РЕЗУЛЬТАТИ

Визначення моментів змін властивостей дозволяє отримати розбиття часового ряду на сегменти, в межах кожного з яких властивості ряду мало міняються. Для кожного сегмента як оцінки коефіцієнтів тренду використовуються оцінки у вигляді сплайнів першого порядку та циклічних компонент [2]. У цьому випадку вдається дозволити суперечність між великою довжиною інтервалу часу, на якій виділяється тренд, і необхідністю використати для опису тренда поліному низького порядку. Часовий інтервал спостережень виявляється розбитим на відрізки невеликої довжини, на кожному з яких тренд описується поліномом та циклічними компонентами, причому поліноми зшиваються на межах відрізків. Коефіцієнти сплайну будуються рекурентно за допомогою методу найменших квадратів, коефіцієнти циклічних компонент знаходимо за допомогою циклічного аналізу [2].

Порівняємо ефективність виділення кусочного циклічного тренду з виділенням безперервного тренда сплайном першого порядку та циклічними компонентами на часових рядах, що описують динаміку зміни цін на фондовому ринку. Результати дослідження якості виділення тренда показують, що для часових рядів зі складною структурою і стрибкоподібними змінами властивостей кусочний тренд є більш точною оцінкою, тому імовірність того, що в залишку містяться частини тренда, низька.

5. ВИСНОВКИ

Застосування запропонованого підходу до виділення тренда економічних часових рядів підтвердило його ефективність. Результати роботи методу порівнювалися нами з іншими методами виявлення змін, зокрема методом Брандта [5]. Сегментація, отримана за допомогою запропонованого алгоритму дозволяє більш ефективно виділити тренд.

ЛІТЕРАТУРА

1. Harvey A.C. *Time series models*. — Cambridge: University Press, 1993.
2. Востров Г.Н., Полякова М.В., Любченко В.В. *Использование статистических и вейвлет-технологий в анализе временных рядов* // Автоматика-2000, Міжнародна конференція з автоматичного управління. — Львів, 11 — 15 вересня 2000р.: Праці в 7-ми томах. — Т.6. — С.116 — 122.
3. Tchamitchian P. *Wavelets: functions and operators* // Wavelets: Theory and applications / Erlebacher. G. (ed.)— Oxford: University Press, 1996.
4. Daubechies I. *Ten Lectures on Wavelets*. — New York: Academic Press, 1992.
5. *Обнаружение свойств сигналов и динамических систем*. /Под ред. М.Бассевиль, А.Банвениста.: — М.: Наука, 1983.

Сегментація текстурних зображень на основі методу векторної підтримки

Довженко Оксана Володимирівна

Дніпропетровський національний університет, Радіофізичний факультет, вул. Наукова 13, корпус 12, м. Дніпропетровськ 49050. тел. 056-7769092. e-mail: dovzhenko@mail.dsu.dp.ua

Мета даної роботи – дослідження інформаційних характеристик методу векторної підтримки стосовно до задачі сегментації текстур. Метод сегментації включає два етапи: на першому реалізується розпізавання і локалізація текстурних об'єктів зображення на основі модуля векторної підтримки з учителем, на другому - додаткова обробка результатів з використанням медіанної фільтрації і морфологічної обробки для виявлення невірно класифікованих блоків і більш зручного візуального аналізу результатів. Екпериментальні дослідження проводилися для різних типів текстурних зображень.

1. Вступ

Сегментація текстурних зображень відноситься до високоактуальних задачу в області комп'ютерного бачення, що приваблює увагу багатьох дослідників довгий час, оскільки і на даний момент є проблемою, далекою від кінцевого розв'язку. Вирішення задачі сегментації текстур провадиться за двома основними напрямками, до першого з котрих відноситься відшукання текстурної моделі, в рамках якої кожен піксел вихідного зображення представляється вектором характеристик (точка багатомірного простору ознак), причому віднайдена модель повинна задовільняти вимогам чутливого аналізу і локалізації подібних регіонів зображення, другий напрямок включає безпосередньо розробку методів сегментації. На даний момент для вирішення першого типу задач найбільш активно використовуються підходи, основані на моделюванні випадкових полів (Марківські, Гаусівські випадкові поля) [1], а також багаторівневій декомпозиції зображення за допомогою банків фільтрів Лапласа, Габора [2], вейвлет-перетворення і популярного останнім часом апарату орієнтованих пірамід [3]. Сегментація текстур, в залежності від наявності апріорної інформації про характер і тип вихідного зображення, може бути реалізована на основі методів навчання з учителем, куди можна віднести нейромережеві підходи [4], метод векторної підтримки або на основі адаптивних методів без учителя [5], однак, єдиного підходу, котрий би задовільняв вимогам високої чутливості аналізу, точності і низьких витрат ресурсів і машинного часу на даний момент не існує.

В даній роботі демонструється можливість застосування нового методу розпізнавання образів, названого як метод векторної підтримки (МВП) для вирішення задачі сегментації текстур, причому обгрунтуванням до використання саме цього методу є задовільні в розумінні точності і витрат машинного часу результати застосування МВП для вирішення задач класифікації [6].

Головна ідея методу – це побудова на основі модуля векторної підтримки оператора, що на основі апріорної інформації про характер ділянок, які треба локалізувати, здійснює аналіз зображення. Питання вибору розміру оператора досліджується в екпериментальній частині.

2. Метод векторної підтримки

Алгоритм векторної підтримки являє собою нелінійний варіант узагальненого портретного алгоритму, що був запропонований Вапніком на початку 60-х років. У зв'язку зі складною програмною реалізацією практично метод векторної підтримки не використовувався, і тільки в 1998 році Платтом [7] був розроблений алгоритм послідовної мінімальної оптимізації, що дозволив усунути обмеження прямого варіанта МВП, пов'язані з високими витратами машинного часу при вирішенні задачі квадратичного програмування.

В рамках спрощеного опису, МВП являє собою задачу побудови гіперплощини, що розділяє вхідні дані $\{x_i, y_i\}$, $i=1,...,m$; $y_i \in \{-1,+1\}$, $x_i \in R_n$ таким чином, що забезпечується максимальна відстань між найближчими точками "позитивного" і "негативного" піднаборів.

Нехай існує набір вхідних даних $x_1, x_2, \ldots, x_m$, де $x_i \in R_n$. Кожному вхідному значенню відповідає мітка з набору $y_1, y_2, \ldots, y_m$, де $y_i \in \{-1,+1\}$, що показує, до якого з двох класів належить те чи інше вхідне значення. Будемо говорити, що гіперплощина $(w \cdot x)+b$, що розділяє дані, існує тоді і тільки тоді, коли виконуються наступні умови:

$$(w \cdot x_i) + b \geq 0, \ для \ \ y_i = 1, \tag{1}$$

$$(w \cdot x_i) + b \leq 0, \ для \ \ y_i = -1, \tag{2}$$

або

$$y_i((w \cdot x_i) + b) \geq 1, \ для \ \ \forall i. \tag{3}$$

Оптимальною гіперплощиною буде така, для котрої відстань між найближчими точками з піднаборів, що помічені позитивними і негативними мітками та самою гіперплощиною була б максимальною.

Відстань між двома найближчими точками з різних піднаборів може бути визначена як:

$$d(w,b) = \min_{\{x_i|y_i=1\}} \frac{w \cdot x_i + b}{|w|} - \max_{\{x_i|y_i=-1\}} \frac{w \cdot x_i + b}{|w|},$$

З виразів (1) та (2) можна бачити, що значення максимума та мінімума відповідно дорівнюють 1 та –1. Тоді необхідно максимізувати відстань:

$$d(w,b) = \frac{1}{|w|} - \frac{-1}{|w|} = \frac{2}{|w|},$$

Задача відшукання максимальної відстані еквівалентна задачі мінімізації $|w|^2/2$ за умови, що виконуються обмеження (3).

Для вирішення цієї задачі використаємо перехід до форми Лагранжа, в результаті чого задача в двоїстій формі може бути записана як:

$$\min(\sum_i \alpha_i - \frac{1}{2}\sum_{i,j} \alpha_i \alpha_j y_i y_j x_i \cdot x_j)$$

За обмежень

$$\alpha_i \geq 0$$
$$\sum_i \alpha_i y_i = 0$$

Де α_i – множники Лагранжа. Кожному вхідному значенню може відповідати лише один такий множник. Ті значення вхідного набору, для котрих множники Лагранжа приймають ненульові значення, називаються векторами підтримки і використовуються для відшукання розділяючої гіперповерхні.

Розглянута вище задача оптимізації не має розв'язків, якщо вхідні дані не є такими, що можуть бути розділені лінійно. В даному випадку обмеження (1) та (2) модифікуються наступним чином:

$$(w \cdot x_i) + b \geq 1 - \xi_i, \ для \ \ y_i = 1,$$
$$(w \cdot x_i) + b \geq \xi_i - 1 \ для \ \ y_i = -1,$$
$$\xi_i \geq 0, \ для \ \ y_i = 1,$$

де ξ_i - це "послаблююча" змінна, що задає верхню межу похибки навчання. Схематичне зображення цього випадку задачі класифікації представлене на Рис. 1. Тоді задача мінімізації $|w|^2/2$ може бути сформульована як

$$\min(\frac{|w|^2}{2} + C(\sum_i \xi_i)),$$

Параметр C – параметр, що задається, і він визначає припустиму похибку класифікації. Якщо переходити до запису задачі в формі Лагранжа, то в двоїстому формулюванні вона буде виглядати наступним чином:

$$\min(\sum_i \alpha_i - \frac{1}{2}\sum_{i,j} \alpha_i \alpha_j y_i y_j x_i \cdot x_j),$$

за обмежень

$$0 \leq \alpha_i \leq C,$$
$$\sum_i \alpha_i y_i = 0.$$

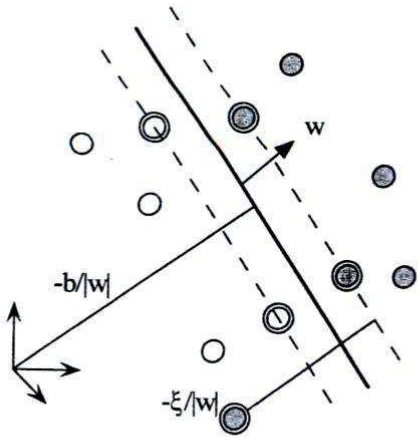


Рис. 1 Схематичне зображення розділяючої гіперплощини для лінійно-нероздільного випадку.

Наступний крок полягає в приведенні розглянутого алгоритма до більш узагальненого, тобто нелінійного варіанту, що досягається відображенням вхідних ознак x_i в деякий багатовимірний простір, причому його розмірність вище, ніж розмірність простору вхідних значень R^d. Це відображення може бути представленим як Φ: $X \rightarrow F$; після здійснення попередньої обробки вхідних даних використовується стандартний алгоритм МВП. Як було розглянуто раніше, для лінійного варіанту методу вхідні дані представляються в формі точкового добутку між вхідними ознаками x_i, x_j, причому $i \neq j$, що для нелінійного випадку: $k(x,x') = <\Phi(x), \Phi(x')>$, причому k називають функцією ядра.

Тепер виникає наступне питання: яким умовам мають задовільняти функції, що можуть бути вибраними в якості функцій ядер? Розглянемо теорему, що характеризує такі функції.

Теорема: нехай існує $k \in L_\infty(X^2)$, таке, що інтегральний оператор $T_k: L_2(X) \rightarrow L_2(X)$,

$$T_k f(\cdot) = \int_X k(\cdot, x) f(x) d\mu(x)$$

є додатнім.

Тоді, якщо

$$\int_{X \times X} k(x,x') f(x) f(x') dx dx' \geq 0$$

для всіх $f \in L_2(X)$, то ми можемо представити $k(x,x')$ як векторний добуток в деякому просторі відображення. Такі функції k будемо називати припустимими ядрами векторної підтримки.

Аналіз публікацій показує, що наступні функції, котрі задовільняють умовам припустимості ядер векторної підтримки, найчастіше використовуються для вирішення практичних задач.

1) Поліноміальні ядра k з $p \in N$:

$$k(x,x') = \langle x,x' \rangle^p .$$

2) Поліноміальні ядра k з $p \in N$ і $c>0$:

$$k(x,x') = \left(\langle x,x' \rangle + c\right)^p .$$

3) Ядра на основі радіальних базисних функцій

$$k(x,x') = e^{-\frac{\|x-x'\|^2}{2\sigma^2}} .$$

В описі теоретичних основ методу векторної підтримки був розглянутий випадок розділення даних на два класи, однак, в реальних задачах зазвичай постає питання класифікації вхідних ознак на значно більшу кількість класів, у зв'язку з чим виникає необхідність розширення апарату класифікації до т. з. мультикласифікатора або N-класифікатора, що дозволяє вирішувати задачу для випадку $N>2$, де N – кількість класів. Стандартний метод для конструювання N-класифікатора полягає в послідовному створенні N класифікаторів, кожен з котрих визначає розділяючу гіперповерхню для i-го класу, позначеного міткою "1" та всіх тестових входів, позначених "-1". Цей підхід скорочено називається "1-vs-r" (1-versus-rest) або "один проти багатьох".

Інший підхід, котрий вигідно відрізняється від попереднього більшою точністю класифікації, вимагає побудови $K = N(N-1)/2$ класифікаторів і скорочено позначається "1-vs-1" або "один проти одного". Метод 1-vs-1 базується на комбінації бінарних класифікаторів на основі архітектури орієнтованого ациклічного графу.

Кінцевий N-класифікатор представляє собою деревоподібну структуру з $N(N-1)/2$ вузлами, кожен з котрих - елемент множини булевих функцій $F = \{f:X \rightarrow \{0,1\}\}$, де X – простір вхідних даних, при чому будь-який i-й вузол у шарі $j<N$ зв'язаний з i-тим $(i+1)$-м вузлом $(j+1)$-го шару. Таким чином, вузли дерева повністю описують всі можливі двокласові комбінації на множині з N класів і кожна вхідна точка для віднесення її до якогось з класів повинна пройти шлях від кореневого вузла до найнижчого шару дерева з N виходами, котрий визначається як шлях оцінки. Недолік описаного підходу – це надлінійне зростання кількості операторів із зростанням значення N. З іншого боку, кожен операційний вузол преставляє собою окремий класифікатор, що дозволяє регулювання параметрів і похибки класифікації, завдяки чому забезпечується вища точність класифікації.

3. Вибір функції ядра

Важливим питанням при реалізації методу для врішення практичних задач є правильний вибір функції ядра, від котрого, в більшості випадків, залежить точність отримуваних результатів. Однак, єдиного підходу для визначення необхідної функції ядра, а також її параметрів, на даний момент немає, тому ця задача вирішується емпірично в кожному окремому випадку.

Для вирішення задачі сегментації текстур з учителем найліпші результати були отримані з використанням радіальних базисних функцій з вар'юванням коефіцієнту σ від 0.001 до 2.5 в залежності від розміру вікна оператора сегментації.

4. Експериментальна частина

Експериментальні дослідження було проведено з використанням різних типів текстурних зображень з відомої бази текстур Brodatz Collection [8]. Реалізація методу сегментації здійснювалася на основі алгоритму послідовної мінімальної оптимізації з використанням методу ациклічних графів для побудови мультикласифікатора.

Для кожного аналізованого зображення було досліджено можливість використання вікон операторів (window) різних розмірів: 4 × 4, 8 × 8, 16 × 16, 32 × 32. Необхідно зазначити, що розмір блока навчання визначає кількість операцій і, відповідно, обсяг витрат машинного часу. Ця обставина пов'язана з тим, що обчислення векторів підтримки відбувається для кожного кроку вікна оператора. Розмір вікна вибирався в залежності від характеру вихідного зображення, що сегментується.

5. Висновки

Як показують отримані результати (Рис. 2), метод векторної підтримки представляє собою чутливий аппарат для вирішення задачі сегментації текстурних зображень, котрий вигідно відрізняється від стандартних підходів високою точністю і низькими витратами машинного часу при реалізації його на основі алгоритму послідовної мінімальної оптимізації.

Однак, виникає цілий ряд додаткових питань, пов'язаних з розвитком методу сегментації, котрі в цілому визначають успішність результатів - а саме розробка автоматичного (адаптивного) методу вибору параметрів ядер, оскільки, як видно з Рис. 2е-л, точність сегментації (котра визначається кількістю невірно класифікованих блоків) суттєво змінюється при невеликому вар'юванні параметру σ ядра радіальної базисної функції.

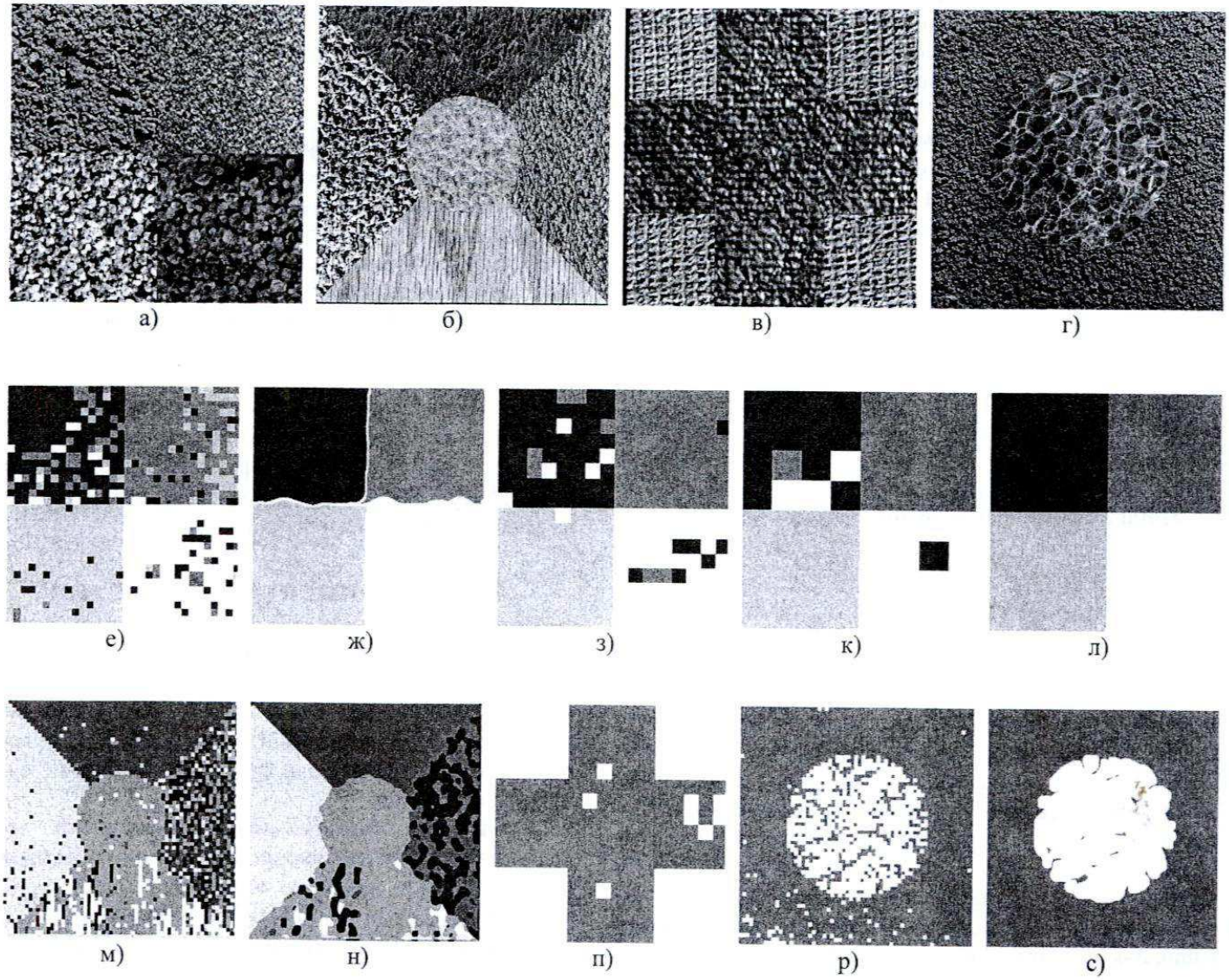


Рис. 2 Результати сегментації текстурних зображень на основі МВП:
а-г – вихідні зображення; е – результат сегментації зображення *а*: *window* = 8, σ = 1.5; ж – результат додаткової фільтрації і морфологічної обробки сегментованого зображення; з – *window* = 16, σ = 0.1; к – *window* = 32, σ = 0.05; л – *window* = 32, σ = 0.07; м - результат сегментації зображення *б*: *window* = 8, σ = 1.5; н – додаткова обробка результатів; п - результат сегментації зображення *в*: *window* = 16, σ = 0.05; р - результат сегментації зображення *г*: *window* = 8, σ = 1.0; с – додаткова обробка результатів.

Література

1. Geman S, Graffigne C., *Marcov random fields and their application to computer vision*. In Gleason, A. (Ed.), Proceedings of the International Congress of Mathematicians, pp. 1496-1517 Berkeley, California, 1989.
2. Jan Puzicha and Joachim M. Buhmann, *Multiscale Annealing for Real-Time Unsupervised Texture Segmentation*. Technical Report IAI-TR-97-4, Institut fur Informatik III, University of Bonn. April 1997.
3. Simoncelli E. P. and Portilla J. Texture *Characterization via Joint Statistics of Wavelet Coefficient Magnitudes.* // 5th IEEE Int'l Conf on Image Processing. - Chicago, IL. Oct 4-7, 1998.
4. Mario Koppen, Javier Ruiz-del-Solar, Pierre Soille, *Texture Segmantation by biologically-inspired use of Neural Networks and Mathematical Morphology*, Pattern recognition, v. 30, pp. 183-196, 1997.
5. Tanachit Tangsukson, Joseph P. Havlicek, *AM-FM image segmentation*, Image Processing, 2000. Proceedings. 2000 International Conference on, Vol. 2, pp. 104 – 107, Jan. 2001.
6. Ахметшин О. М. , Довженко О. В., *Підвищення ефективності класифікації в багатовимірному просторі ознак великої розмірності на основі методу векторної підтримки,* Збірка праць МКІМ-2002, т. 4, сс. 36-41, 2002.
7. J. Platt, *Fast training of support vector machines using sequential minimal optimization. In advances in kernel methods–support vector learning*, pp.185-208, Cambridge, 1999.
8. http://www.ux.his.no/~tranden/brodatz.html.

ДЕКОМПОЗИЦІЯ СКЛАДНИХ СХЕМ НА ОСНОВІ ІЄРАРХІЧНОЇ КЛАСТЕРИЗАЦІЇ

Р.П Базилевич, І.В.Подольський, П.Р.Базилевич
НУ "Львівська політехніка", вул. С.Бандери 12, Львів, 79013,
тел. (0322) 398578, ел. пошта: rbaz@polynet.lviv.ua

Анотація. Пропонується ряд алгоритмів декомпозиції складних схем, придатних для задач високих та надвисоких розмірностей. Методологічною основою є метод оптимального згортання схеми, який виявляє ієрархічно вкладені згустки схеми – її кластери. Деталізовано декілька алгоритмів початковго розбиття схеми на частини та його оптимізації. Замість оперування базовими елементами схеми, число яких може бути надзвичайно великим, використовуються кластери довільної розмірності. Це покращує якість розв'язку та зменшує обчислювальні затрати, підвищуючи швидкодію алгоритмів. Приводяться результати тестувань, які підтверджують ефективність запропонованої методики.

1. ВСТУП

Однією з важливих задач при розробці алгоритмів розпізнавання образів є розбиття складної схеми на частини з заданими обмеженнями. З математичної точки зору – це важковирішувана комбінаторна задача неполіноміальної складності (*NP* важка). У випадку її великих та надвеликих розмірностей (сотні тисяч та мільони складових елементів) для її ефективного розв'язання можливим є використання тільки наближених методів. Доцільним для цієї мети виявився метод оптимального згортання схеми [1,2], що дозволяє виявити ієрархічно вкладені кластери схеми. Кожна реальна електрична схема є неоднорідною, має певні згустки (кластери), в яких складові елементи зв'язані більш щільно, ніж у решти її частинах. При розв'язуванні різноманітних задач їх можна розглядати як макроелементи та описувати макромоделями. Це суттєво зменшує обчислювальну складність та сприяє покращенню якості розв'язку.

2. ФОРМУЛЮВАННЯ ЗАДАЧІ

Під розбиттям схеми, утвореної множиною $P=\{p_1,\ldots, p_n\}$ складових елементів, на частини вважаємо систему $\widetilde{P} = \{P_1,\ldots, P_k\}$, для якої

$$(\forall P_i \in \widetilde{P})[(P_i \subset P) \,\&\, (P_i \neq \varnothing)];$$

$$(\forall (P_i,P_j) \in \widetilde{P})[P_i \cap P_j = \varnothing]; \bigcup_{i=1}^{k} \bigcup_{j=1}^{n_i} p_{ij} = P.$$

Тут $P_i = \{p_{i1},\ldots, p_{in_i}\}$ – множина всіх елементів частини P_i. Задача полягає у знаходженні найкращого з можливих розбиттів, що задовольняє бажані обмеження. Вона може бути сформульована як знаходження такого розбиття $\widetilde{P}^*$, при якому досягається екстремум значення критерію Q у допустимій області D можливих розв'язків:

$$Q(\widetilde{P}^*) = extr\, Q(\widetilde{P}_i);\ \widetilde{P}_i \in D.$$

Як критерії та обмеження при розбитті можуть виступати: функційна завершеність утворених підсхем; еквівалентність виділених частин підсхемам з набору типових рішень, заданих бібліотекою (задача покриття); взаємна еквівалентність виділених підсхем з мінімізацією числа їх типів (задача типізації); число утворених підсхем; зовнішня зв'язність підсхем – сумарне число зв'язків між утвореними частинами; внутрішня зв'язність підсхем – сумарне число зв'язків всередині підсхем та їх середнє значення на одну підсхему; окремі характеристики кожної підсхеми: число її складових елементів, число зовнішніх зв'язків та деякі інші величини (сумарна площа всіх елементів, сумарний об'єм, сумарне тепло, що виділяється; множини елементів, які повинні бути призначені в різні підсхеми, або навпаки – в одну підсхему; прив'язка за характеристиками зв'язності до певних зовнішніх об'єктів, наприклад роз'ємів).

Основна ідея розвинутого підходу полягає в тому, що спочатку будується дерево оптимального згортання T^R за висхідною стратегією, яке виділяє кластери схеми, після чого за низхідною та комбінованою стратегіями на основі попередньо отриманої інформації здійснюється розбиття схеми на частини. Замість аналізу схеми на рівні початкових елементів здійснюється комбінаторний аналіз ієрархічно вкладених кластерів. Базуючись на розвинутому підході, задача розв'язується за три етапи: побудова дерева оптимального згортання; початкове розбиття; оптимізація розбиття.

3. АЛГОРИТМИ ПОЧАТКОВОГО РОЗБИТТЯ

Використовуючи дерево оптимального згортання можна реалізувати декілька алгоритмів отримання початкового розв'язку для задачі розбиття схеми на частини: послідовного та паралельно-послідовного виділення частин; дихотомічного низхідного ділення; ділення схеми по її мінімальних розрізах та комбіновані.

Алгоритми послідовного виділення частин розбиття є найпростішими для реалізації, проте вони мають суттєвий недолік – в процесі вилучення

елементів та формування частин розбиття утворюються "хвости" – частини схеми, які слабо пов'язані між собою, та з яких формуються останні частини розв'язку. Перші виділені частини будуть доброї якості, а останні – поганої.

В алгоритмах паралельно-послідовного виділення частин розбиття дерево згортання формується повністю. Виділяються всі підсхеми, число елементів яких дорівнює бажаному значенню або перевищило це значення вперше в процесі нарощення дерева згортання. З них процедурою вилучення зайвих елементів формуємо частини розбиття. Для решти елементів будуємо нове дерево та виділення частин рекурсивно продовжуємо. У порівнянні з попереднім типом вказаний підхід є більш швидкодіючим, проте він може давати гірші результати, оскільки "хвости" формуються паралельно без врахування решти частини схеми. Для одного дерева згортання утворюється одночасно декілька "хвостів". У попередньому випадку після формування кожної окремої частини дерево згортання будувалось наново.

Ефективним підходом для усунення "хвостів" може бути "вимушене" призначення елементів у частини, яке одночасно враховує потреби рівномірного розподілу елементів серед усіх частин розбиття з врахуванням характеристик їх зв'язності. Такий підхід реалізується в алгоритмах *дихотомічного низхідного ділення*. Виділяються дві найвищі вершини дерева згортання та визначається число частин розбиття, що може бути утворене з

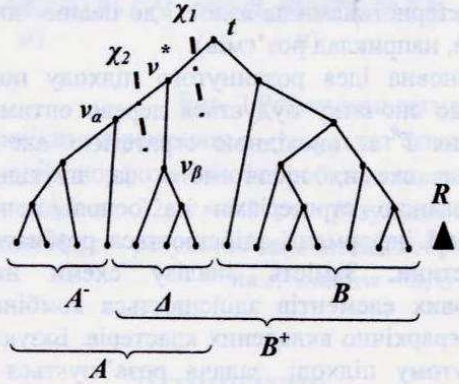


Рис. 1. Початкове виділення частин

кожної з них. Якщо числа їх елементів є кратними до бажаного значення елементів однієї частини (або близькими з врахуванням заданих обмежень на їх значення), то задача розбивається на дві ідентичні задачі менших розмірів, які продовжуємо дробити аналогічним чином. У більшості випадків ця умова не виконується. Тоді визначаємо, яке число елементів необхідно перенести з однієї частини в іншу, щоб задовольнити це значення. Виділяємо ту частину, з якої необхідно вилучити менше число елементів. Нехай це буде частина A, що відповідає вершині v^* (Рис. 1). Розглянемо *процедуру перенесення елементів з однієї частини в іншу*. Тут можливими є два варіанти – спрощеного або розширеного аналізу. При спрощеному аналізі порівнюємо за числом елементів обидві складові підсхеми вершини v^*: v_α та v_β, виділяємо меншу з них (нехай це буде v_β) та визначаємо серед її елементів бажані для перенесення. Можливими є три випадки. У першому з них число її елементів відповідає бажаному для перенесення значенню. Тоді приєднуємо цю підсхему до частини B, утворюючи дві нові частини: $A^-(A - \Delta)$ та $B^+(B + \Delta)$ з бажаним для наступного ділення числом елементів. У другому випадку це число може виявитися меншим за бажане значення. Тоді можливими є два варіанти. Для першого з них, який забезпечує більшу точність, приєднуємо елементи Δ до частини B, моделюємо частину $(B + \Delta)$ однією вершиною, а з елементів частинии $(A - \Delta)$ будуємо нове дерево та продовжуємо пошук необхідних для перенесення елементів аналогічним чином. Для другого, більш простого, йдемо вниз по дереву, розглядаючи наступні дві складові більшої підсхеми v_α і т.д. аж поки не отримаємо бажане для перенесення значення число елементів або перше більше за це значення. У третьому випадку це чило є більше бажаного значення. Тоді задача зводится до *розподілу елементів множини Δ між частинами A^- та B* таким чином, щоб задовольнити бажані умови на число елементів, що призначаються в обидві частини.

При розширеному аналізі пошук елементів для перенесення здійснюємо серед всіх елементів частини A. Тут можуть бути використані два

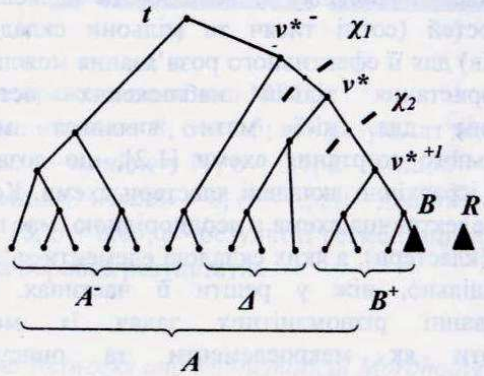


Рис. 2. Перерозподіл елементів частини A

підходи: на основі дерева згортання та на основі визначення ефективності перенесення окремих елементів та кластерів. Розглянемо спочатку перший підхід. Суть його полягає в побудові дерева згортання на основі всіх елементів частини A з розглядом частини B як єдиного цілого (Рис. 2). Решта частини схеми R також розглядається як єдине ціле, проте в процесі побудови дерева згортання до неї не дозволяється приєднувати елементи. Аналізуємо шлях від вершини B до вершини дерева t. Визначаємо на цьому шляху першу таку вершину v^*, що включення всіх елементів (включно з нею) на шляху від неї до B утворює вперше множину з

числом елементів, рівним або більшим від бажаного для перенесення. Нехай перетин χ_1 (між вершинами v^{*-1} та v^*) виділяє таку вершину. Якщо це число відповідає бажаному для перенесення значенню, то розрізаємо схему за цим перерізом, утворюючи нові частини. Якщо воно є більшим від бажаного, то утворюємо нові частини A^- та B^+, а елементи множини Δ розсипаємо. До частини A^- приєднуємо всі елементи, що знаходяться на рисунку лівіше перерізу χ_1, а до частини B^+ - всі елементи від перерізу χ_2 до вершини B включно з останньою.

Може трапитися частковий випадок, коли на виділеному вище шляху від вершини B до t немає належного числа елементів, або такі елементи взагалі є відсутніми. У першому разі до частини B приєднуємо всі ці елементи та задачу зводимо до нової аналогічної задачі зі збільшеним числом елементів частини B та зменшеним числом елементів частини A і продовжуємо задачу ітеративно розв'язувати аналогічним чином. Виродженим може трапитися випадок, коли на цьому шляху взагалі вершин немає. Раціональним тут слід вважати застосування підходу на основі визначення ефективності перенесення окремих елементів та кластерів.

Перейдемо до розгляду процедури *розподілу елементів множини Δ між частинами A^- та B (або B^+)*. Частини A^- та B (B^+) та решта частина схеми R розглядаються як окремі елементи, а елементи множини Δ розсипаються. Будується дерево

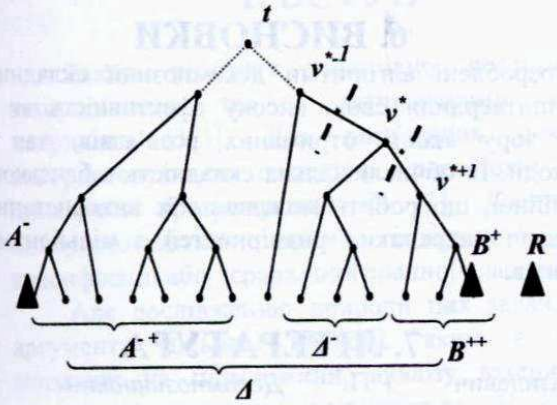


Рис. 3. Розподіл елементів множини Δ

згортання (Рис. 3). У процесі його побудови не дозволяється об'єднуватися піддеревам, що включають елементи A^- та B^+. До елемента R також не дозволяється приєднувати елементи, проте зв'язки з цим елементом беруться до уваги. Після завершення згортання вершин утворюються два піддерева, які умовно об'єднуємо вершиною t. На шляху від B^+ до A^- знаходимо вершину v^* таку, що включення її елементів та решти елементів цієї частини шляху буде відповідати бажаному значення або буде першим значення, більшим за задане число. Якщо це число відповідає бажаному значенню, то задачу вважаємо розв'язаною. В протилежному випадку всі елементи до цієї вершини зліва додаємо до частини A^-, а всі елементи справа – до частини B^+, а елементи самої вершини v^* розсипаємо на множину Δ та продовжуємо задачу ітеративно розв'язувати аналогічним чином. Як вироджений може трапитися випадок, коли всі елементи між вершинами A^- та B^+ об'єднуються в процесі згортання в одну вершину v^*, тобто не виникає умов для їх перерозподілення. Тоді необхідно застосувати підхід на основі визначення ефективності призначення окремих елементів та кластерів в кожну з підсхем зокрема.

4. АЛГОРИТМИ ОПТИМІЗАЦІЇ РОЗБИТТЯ

Доцільно виділити дві основні стратегії оптимізації розбиття: попарну та групову. Перша реалізується між окремими сформованими на попередньому етапі частинами розбиття. Групова оптимізація дозволяє переформовувати одночасно декілька груп. Початковими умовами для оптимізації є деяке розбиття $\widetilde{P} = \{P_1, ..., P_k\}$, яке оцінюється за певним значенням критерію якості $Q(\widetilde{P})$. Метою оптимізації є отримання розбиття $\widetilde{P}^* = \{P^*_1, ..., P^*_k\}$ з кращим значенням критерію якості: $Q(\widetilde{P}^*) \succ Q(\widetilde{P})$. Запропоновані алгоритми реалізують покроковий процес покращання розв'язку.

В алгоритмах попарної оптимізації заданими є частини P_i та P_j, які описуються критерієм якості $Q(P_i, P_j)$. Необхідно отримати оптимізовані частини P^*_i та P^*_j, для яких $Q(P^*_i, P^*_j) \succ Q(P_i, P_j)$. Мова йде про перерозподіл елементів між частинами, який може покращити значення критерію. Можливими є два шляхи: для випадку розбиття на декілька частин (більше двох) поновне розділення множин, утворених елементами всіх пар P_i та P_j, або ітераційна оптимізація кожної пари (P_i, P_j). Перший шлях передбачає об'єднання елементів обох частин в одну множину з повторним розв'язуванням задачі розбиття розглянутими вже методами початкового розбиття. Другий шлях передбачає ітераційну оптимізацію кожної пари (P_i, P_j). Можливими є декілька стратегій: поодинокого переносу кластерів та елементів з коригуванням показників ефективності після кожної такої процедури, що забезпечуватиме найкращу якість, проте вимагає суттєвих обчислювальних затрат, або групового одночасного перенесення всіх елементів та кластерів з найкращим значенням критерію, а також парного чи групового обміну множинами кластерів та елементів. Для виходу з локальних екстремумів досліджено декілька типів різних збурень отриманого розв'язку з наступним продовженням оптимізаційного процесу. Збурення реалізуються перенесенням одного або декількох кластерів з однієї частини в іншу з погіршенням значення куритерію, а також вилученням кластерів з найкращим значенням критерію для перенесення в обох частинах та

наступним перерозподілом їх складових елементів. Для цього можуть бути використані вже описані алгоритми початкового розподілу, які реалізуються тільки на виділеній частині елементів з початковими умовами, що визначаються залишеними частинами, або їх випадковий перерозподіл з наступною оптимізацією.

5. ЕКСПЕРИМЕНТАЛЬНІ ДОСЛІДЖЕННЯ

Ефективність запропонованих алгоритмів підтверджена рядом експериментів, виконаних на відомих тестах фірми IBM [3]. Графік часової залежності для базової процедури – побудови дерева згортання, при якому мінімізується кількість зовнішніх зв'язків кожного кластера. Для виходу з локальних екстремумів застосовувались збурення кількох типів: перенесення одного, двох або трьох кластерів одночасно, які мають оптимальне відношення оцінки ефективності перенесення до кількості елементів або найкраще значення оцінки ефективності перенесення, комбіновані збурення та зрізання заданого відсотку елементів "шапки" (кластери, що мають найкраще значення оцінки перенесення) в обох частинах з подальшим випадковим призначенням цих елементів частинам розбиття. Часові характеристики наведено для процесора Intel Pentium III 800 MHz.

Таблиця 1. Результати оптимізації

Стратегія	Зріз "шапки",%	Кількість кластерів збурення	Тип збурення	Кількість збурень	Макс. значення критерію	Мін. значення критерію	Оптимальні	≤1%	≤2%	≤5%	≤10%	Середнє арифметичне	Середня кількість ітерацій	Середній час, с.
1	0	1	оптимальний кластер	5	699	180	2	16	60	60	60	227	62	274
2	0	1	найкращий кластер	5	707	180	7	28	40	44	44	299	39	194
3	0	1	чергування 2 опт./ 2 найкр.	5	699	180	3	24	58	59	59	230	53	244
4	0	1	чергування 2 найкр./ 2 опт.	5	699	180	6	35	52	52	52	241	49	292
5	0	1+1	одночасно опт. + найкр.	5	699	180	2	16	61	61	61	227	62	252
6	0	2+2	одночасно опт. + найкр.	5	699	180	4	17	62	62	62	226	62	214
7	10	0	зрізання "шапки"	2	707	180	7	32	47	49	49	290	38	204
8	10	0	зрізання "шапки"	5	707	180	20	43	49	49	49	288	60	281
9	10	0	зрізання "шапки"	10	707	180	29	48	50	50	50	284	96	407
10	2	0	зрізання "шапки"	5	707	180	12	37	46	46	46	293	54	272

На пачці зі ста випадково згенерованих початкових розбиттів досліджено вплив ряду параметрів алгоритмів оптимізації, таких як тип критерію згортання, кількість та тип збурень, відсоток згортання дерева, відсоток зрізання "шапки" для збурення, на якість отримуваних розв'язків та кількість попадань в певний окіл оптимуму. Стратегії з високим відсотком попадань розв'язків у заданий окіл оптимуму доцільно застосовувати для подальшої оптимізації розв'язків у задачах розбиття.

У Таблиці 1 наведено результати досліджень для алгоритму оптимізації, що виявився при попередніх тестуваннях швидкодіючим та ефективнішим, у порівнянні з іншими розглянутими алгоритмами. Обраний алгоритм оптимізації передбачає почергове одностороннє перенесення з однієї частини в іншу кластерів/елементів з від'ємним та нульовим значенням оцінки ефективності перенесення q_i в кількості, що не перевищує заданого відхилення розміру схеми від точного розбиття на два. Завершується робота алгоритму при умові повторення значення критерію якості $Q(\tilde{P}^*)$.

Результати подано для відсотку згортання дерева на кожній ітерації у 20% від діапазону значень критерію, допустимому відхиленні розбиття схеми від точного розбиття на два - 10% та критерію згортання є близьким до лінійного [4].

6. ВИСНОВКИ

Розроблені алгоритми декомпозиції складних схем підтвердили свою високу ефективність як з точки зору якості отриманих розв'язків, так і швидкодії. Їх обчислювальна складність є близькою до лінійної, що робить можливим їх використання для задач надвеликих розмірностей з мільйонами елементів.

7. ЛІТЕРАТУРА

1. Базилевич Р.П. *Декомпозиционные и топологические методы автоматизированного конструирования электронных устройств.* Львів: Вища школа, 1981. –168 с.
2. Bazylevych R. *The Optimal Circuit Reduction Method as an Effective Tool to Solve Large and Very Large Size Intractable Combinatorial VLSI Physical Design Problems* // Proc. of 10th NASA Symposium on VLSI Design. - Albuquerque. - 2002. – P. 6.1.1-6.1.14.
3. Alpert C.J. *The ISPD-98 Circuit Benchmark Suit* // Proc. ACM/IEEE Intern. Symposium on Physical Design.- April 1998. – P. 80-85.
4. Базилевич Р.П., Подольський І.В. *Ієрархічна кластеризація складних схем.* Вісник Державного університету "Львівська політехніка", №. 392, 2000, с. 155-158

ДВІ ЗАДАЧІ РОЗБИТТЯ

Надія ТИМОФІЄВА
МННЦ ІТіС НАН та Міносвіти України 03022, Київ, просп. Ак. Глушкова, 40,
тел.: (044) 266 53 52, факс (044) 266 1570

ABSTRACT

Two subclasses of the problems of partitioning which are divided by a method of assigning of entry datas are considered. Mathematical models are elaborated for them, a structure of entry datas is investigated, an obective function is formulated. Differences of their mathematical models are investigated. For beforehand assigned order of the set partitioning into subsets, elements of which are arguments of the objective function, its values shanges are analyzed. It is proved, that for assigned order for both types of the problems of partitioning , its values changes follow the laws of piece-monotonons functions.

The results of these researches may be used for elaboration the practical problems of partitioning.

1. ВСТУП

Задачі розбиття грунтовно досліджені як закордонними так і вітчизняними вченими, наприклад [1-3]. Для розв'язання цих задач використовуються ітераційні, послідовні методи, динамічне програмування [1, 3]. Досить часто використовуються способи ієрархічної ростучої класифікації або ієрархічної спадної класифікації.

Але дослідженню природи цих задач, аналізу аргумента цільової функції, яким є розбиття множини на підмножини, аналізу вхідних даних достатньої уваги не приділяється. У статті на прикладі двох підзадач задач розбиття, які відрізняються між собою способом завдання вхідних даних (їх математичні постановки подано нижче), коротко висвітлено ці проблеми. Незважаючи на те, що цільова функція для них формулюється по різному, а для їх розв'язання необхідно розробляти різні підходи, в літературі ці задачі ототожнюються наприклад [1].

В літературі [2 3] оговорено, що значення цільової функції на всій множині змінюється за законами монотонних функцій, але строгого доведення не наводиться. Знання властивостей задач розбиття дозволяє строго довести, що вона змінюється за законами кусково-монотонних функцій.

2. ПРО АРГУМЕНТ ЦІЛЬОВОЇ ФУНКЦІЇ ЗАДАЧ РОЗБИТТЯ

Так як аргументом цільової функції в задачах розбиття є розбиття n елементної множини $A=\{a_1,\dots,a_n\}$ на підмножини, спочатку розглянемо характерні властивості цих об'єктів і структуру їх множини.

Розбиттям n елементної множини $A=\{a_1,\dots,a_n\}$ на η підмножин (блоків) назвемо множину підмножин $\rho=\{\rho_1,\dots,\rho_\eta\}$ таку, що $\rho_1\cup\dots\cup\rho_\eta=A$, $\rho_j\neq\varnothing$, $\rho_j\cap\rho_t=\varnothing$, $j\neq t$, $j,t\in\{1,\dots,\eta\}$,. Непуста підмножина $\rho_j=\{a_1,\dots,a_{\xi_j}\}$, $a_s\in A$, $s\in\{1,\dots,n\}$, може мати від 1 до n елементів, ξ_j - кількість a_s у підмножині ρ_j.

Два розбиття ρ і ρ^* назвемо тотожними, якщо кількість їх підмножин однакова і для будь-якої підмножини ρ_j можна знайти у множині ρ^* підмножину ρ_t^*, яка не відрізняється від ρ_j ні кількістю елементів ні самими елементами.

Множину розбиттів позначимо Θ. Вважатимемо, що Θ упорядкована у заданому порядку, а її елементи перенумеровані від 1 до q. В подальшому розбиття множини на підмножини позначимо ρ^k, де k порядковий їх номер у множині Θ.

Означення. Два нетотожні розбиття $\rho^k,\rho^l\in\Theta$ назвемо ізоморфними, якщо кількість їх підмножин однакова ($\eta^k=\eta^l$), і для будь-якої підмножини ρ_j^k у розбитті ρ^k знайдеться підмножина ρ_t^l для якої $\xi_j^k=\xi_t^l$.

Підмножину $L_{\eta^k}\subset\Theta$, яка складається з ізоморфних розбиттів, назвемо підмножиною ізоморфних розбиттів. Кількість ρ^k

у ній $\frac{n!}{\xi_1^k!\xi_2^k!\ldots\xi_{\eta^k}^k!\,m_s!}$, де m_s - число підмножин $\rho_j^k,\rho_t^k\subset\rho^k$, у яких $\xi_j^k=\xi_t^k$, $s\in\{1,\ldots,e\}$, e - кількість блоків, кожен із яких об'єднує ρ_j^k з однаковою кількістю ξ_j^k.

Отже пошук оптимального значення цільової функції в задачах розбиття проводиться або на всій множині Θ, або на підмножинах ізоморфних розбиттів L_{η^k}.

3. МАТЕМАТИЧНА МОДЕЛЬ ЗАДАЧ РОЗБИТТЯ

Математична модель задач розбиття, як правило, задається однією множиною $A=\{a_1,\ldots,a_n\}$. Але, в залежності від їх типу, вхідні дані у них задаються по-різному. Від цього залежить спосіб завдання цільової функції, способи і алгоритми розв'язання цих задач.

Отже за способом завдання вхідних даних в задачах розбиття можна виділити такі підкласи: а) задачі, у яких вхідними даними є значення вагів між заданими елементами $a_t,a_j\in A$ Це - кластеризація, таксономія, компоновка, розрізання графа на задану кількість частин, та ін.; б) задачі, у яких вхідні дані визначають властивість заданих елементів $a_j\in A$. До них належить задача про купу каміння, задача розбиття множини споживачів на зони обслуговування.

Побудуємо математичну модель для задачі а)

Нехай задано множину $A=\{a_1,\ldots,a_n\}$, між елементами яких визначено взаємозв'язок, кількісне значення яких подамо симетричною матрицею C. Назвемо вагою величину, що визначає залежність, яка існує між елементом a_t і елементом a_j.

Уведемо також (0,1)-симетричну комбінаторну матрицю $Q(\rho^k)\in Q_\Theta$, яка задає розподілення елементів a_j множини A по підмножинах ρ_j^k для k - го розв'язку задачі.. Елемент цієї матриці $g_{tj}(\rho^k)=1$, якщо елементи a_t і a_j відносяться до однієї підмножини, і $g_{tj}(\rho^k)=0$, якщо a_t і a_j відносяться до різних підмножин; Q_Θ упорядкована множина усіляко можливих таких матриць. Матриця $Q(\rho^k)$ утворюється із попередньої $Q(\rho^l)$ за правилами генерування усіляко можливої множини розбиттів.

Подамо елементи правого поля матриці $Q(\rho^k)$ комбінаторною функцією $\bar\beta(\bar f(\bar x)(\rho^k))=(\beta(f(x_1)(\rho^k)),\ldots$ $\ldots,\beta(f(x_m)(\rho^k))$, а елементи правого поля матриці C числовою функцією $\bar\varphi(\bar x)=(\varphi(x_1),\ldots,\varphi(x_m))$, де $m=\frac{n(n-1)}{2}$ кількість складових у $\bar\beta(\bar f(\bar x)(\rho^k))$ і $\bar\varphi(\bar x)$. Функція цілі для задачі а) набуде вигляду

$$F(\rho^k)=\sum_{j=1}^{m}\beta(f(x_j)(\rho^k))\,\varphi(x_j) \qquad (1)$$

Задача а) полягає у тому, щоб об'єднати однорідні елементи в одні і ті ж блоки. В залежності від критерію, який використовується для оцінки результату розв'язання задачі, однорідні елементи можуть мати між собою або найбільші ваги або найменші. Вважатимемо, що однорідні елементи мають між собою найбільші ваги і вони об'єднуються в одну підмножину. В цьому випадку задача розбиття полягає у знаходженні такого аргументу $\rho^k\in\Theta$, для якого функція цілі (1) при задоволенні заданих умов була б найбільшою.

Побудуємо математичну модель для задачі б). До них відноситься задача про купу каміння, яка формулюється так. Задано множину каменів $A=\{a_1,\ldots,a_n\}$, кожному із яких присвоєно додатнє число, його вагу $V(j)$. Необхідно розділити цю множину на задану кількість куп так, щоб вага самої важкої із них була б найменшою.

Математична модель для цієї задачі також задана однією множиною $A=\{a_1,\ldots,a_n\}$, але між $a_j\in A$ відсутні зв'язки. Натомість вхідні дані задано скінченною послідовністю (числовою функцією) $\bar\varphi(\bar x)$, складові якої $\varphi(x_j)$ визначають характерні властивості елемента a_j (в задачі про купу каміння це вага j - го каменя).

Для визначення розподілення елементів множини A по підмножинах ρ_j^k для k - го варіанту розв'язку задачі уведемо комбінаторну функцію $\bar\beta(\bar f(\bar x)(\rho^k))=(\beta(f(x_1)(\rho_t^k)),\ldots$

$\ldots,\beta(f(x_n)(\rho_t^k))$, де $\beta(f(x_j)(\rho_t^k))\in$ $\in\{0,1\}$ і $\beta(f(x_j)(\rho_t^k))=1$, якщо a_j - й елемент входить у ρ_t^k - й блок розбиття ρ^k, і $\beta(f(x_j)(\rho_t^k))=0$ в іншому разі. Для цього типу задач запишемо цільову функцію

$$F(\rho^k)=\max_{\substack{\rho_t^k\subset\rho^k\\ t=1,\eta^k}}\left(\sum_{j=1}^{n}\varphi(x_j)\beta(f(x_j)(\rho_1^k)),\ldots\right.$$

$$\left.\ldots,\sum_{j=1}^{n}\varphi(x_j)\beta(f(x_j)(\rho_{\eta^k}^k))\right)$$

Задача б) полягає у знаходженні такого ρ^{k*}, для якого цільова функція $F(\rho^k)$ на $L_{\eta^k}\subset\Theta$ набуває найменшого значення

$$F(\rho^{k*})=\min_{\rho^k\in L_{\eta^k}\subset\Theta}F(\rho^k)$$

4. ПРО ЗМІНУ ЗНАЧЕНЬ ЦІЛЬОВОЇ ФУНКЦІЇ

Для вибору способів розв'язання задач а) і б) необхідно спочатку визначити, яким чином змінюється цільова функція на усій множині розбиттів. Відомо [2,3], якщо $\rho^k\in\Theta$ упорядкувати так, що або $\eta^k<\eta^l$, або $\eta^k>\eta^l$, $k<l$, то цільова функція $F(\rho^k)$ змінюється за законами монотонних функцій.

Розглянемо, наскільки правильне це твердження. Спочатку наведемо такий приклад.

Задано множину $A=\{1,\ldots,5\}$. Для задачі а) числова функція $\varphi(x)=(1,2,3,4,5,6,7,8,9,10)$, а для задачі б) - $\varphi(x)=(1,2,3,4,5)$. Комбінаторна функція $\bar{\beta}(\bar{f}(\bar{x})(\rho^k))$ для задачі а) змінюється в залежності від значення $\rho^k\in\Theta$ і з урахуванням властивостей комбінаторних матриць. Для задачі б) функція змінюється за правилами, описаними вище.

Для цих параметрів підрахуємо значення цільової функції $F(\rho^k)$, $k=\overline{1,52}$. Результати обчислення наведено у табл. 1. Перша колонка містить кількість блоків η^k у розбитті ρ^k для підмножини $L_{\eta^k}\subset\Theta$. Друга - розбиття ρ^k, третя - значення комбінаторної функції для задачі а), четверта - значення цільової функції для задачі а), п'ята - значення цільової функції для задачі б). У зв'язку з обмеженням на об'єм статті для підмножин L_{η^k} у табл. 1 наведено лише ті ρ^k, для яких $F(\rho^k)$ набуває найбільшого і найменшого значень.

Таблиця 1. Зміна значень цільової функції $F(\rho^k)$ для задач а) і б)

	$\bar{\varphi}(\bar{x})$	1 2 3 4 5 6 7 8 9 10		
1	2	3	4	5
1	(1 2 3 4 5)	1 1 1 1 1 1 1 1 1 1	55	55
2	(1 2 3 4) (5)	1 1 1 0 1 1 0 1 0 0	25	10
	(5 2 3 4) (1)	0 0 0 0 1 1 1 1 1 1	45	14
2	(1 2 3) (4 5)	1 1 0 0 1 0 0 0 0 1	18	9
	(4 1 2) (3 5)	1 0 1 0 0 1 0 0 1 0	19	8
	(3 4 5) (1 2)	1 0 0 0 0 0 0 1 1 1	28	12
3	(1 2 3) (4) (5)	1 1 0 0 1 0 0 0 0 0	8	6
	(4 5 3) (1) (2)	0 0 0 0 0 0 0 1 1 1	27	12
3	(3 2) (1 4) (5)	0 0 1 0 1 0 0 0 0 0	8	5
	(2 3) (4 5) (1)	0 0 0 0 1 0 0 0 0 1	15	9
4	(1 2) (3) (4) (5)	1 0 0 0 0 0 0 0 0 0	1	5
	(5 4) (1) (2) (3)	0 0 0 0 0 0 0 0 0 1	10	9
5	(1) (2) (3) (4) (5)	0 0 0 0 0 0 0 0 0 0	0	5

Із результатів обчислень видно, що у підмножині ізоморфних розбиттів L_{η^k} по відношенню до L_{η^l}, $\eta^k<\eta^l$, і $\eta^l=\eta^k+1$, або $\eta^k=\eta^l$, а $\xi_1^k<\xi_1^l$, значення цільової функції $\min_{\rho^k\in l_{\eta^k}}F(\rho^k)<\max_{\rho^l\in L_{\eta^l}}F(\rho^l)$.

Отже, те, що значення цільової функції для оговореного вище порядку на множині Θ змінюється за законами монотонних функцій, не є справедливе.

Дослідимо зміну значень цільової функції на множині Θ для загального випадку. Складові $\varphi(x_j)$ функції $\bar{\varphi}(\bar{x})$ є випадкові змінні і для конкретних задач неоднакові. Комбінаторна функція змінюється лише в залежності від розбиття $\rho^k\in\Theta$ за законами комбінаторних матриць і для будь якої задачі значення її постійне. Розглянемо, яким чином вона змінюється на множині розбиттів Θ. Так як

значення $\beta(f(x_j)(\rho^k)) \in \{0,1\}$, визначимо, яким чином змінюється кількість одиниць у $\overline{\beta}(\overline{f}(\overline{x})(\rho^k))$ в залежності від ρ^k.

Наведемо такий приклад. Нехай задано множину $A = \{1,2,3,4,5,6,7,8)$, $n = 8$.

За формулою $\sum_{j=1}^{\eta^k} \frac{\xi_j^k(\xi_j^k - 1)}{2}$ [4] для підмножин L_{η^k} обчислимо кількість одиниць у комбінаторних функціях. Їх кількість для розбиттів $\rho^k \in L_{\eta^k}$ і $\rho^l \in L_{\eta^l}$ різна, а для розбиттів $\rho^k \in L_{\eta^k}$ однакова. Результати обчислень наведено у табл. 2. Перша колонка містить номер підмножини L_{η^k}, друга - розбиття числа n, яке задає кількість блоків η^k у ρ^k і кількість елементів у ρ_j^k, третя - кількість одиниць у функції $\overline{\beta}(\overline{f}(\overline{x})(\rho^k))$ для $\rho^k \in L_{\eta^k}$.

Таблиця 2. Кількість одиниць у комбінаторних функціях для підмножин $L_{\eta^k} \subset \Theta$.

1	2								3
	ξ_1^k	ξ_2^k	ξ_3^k	ξ_4^k	ξ_5^k	ξ_6^k	ξ_7^k	ξ_8^k	
1	8								28
2	7	1							21
	6	2							16
	5	3							13
	4	4							12
3	6	1	1						15
	5	2	1						11
	4	3	1						9
	4	2	2						8
	3	3	2						7
4	5	1	1	1					10
	4	2	1	1					7
	3	3	1	1					6
	3	2	2	1					5
	2	2	2	2					4
5	4	1	1	1	1				6
	3	2	1	1	1				4
	2	2	2	1	1				2
6	3	1	1	1	1	1			3
	2	2	1	1	1	1			2
7	2	1	1	1	1	1	1		1
8	1	1	1	1	1	1	1	1	0

Незалежно від вхідних даних (функція $\varphi(x)$) найменша кількість одиниць у комбінаторних функціях для $L_{\eta^k} \subset \Theta$ порівняно з $L_{\eta^l} \subset \Theta$, $\eta^k < \eta^l$, $\eta^l = \eta^k + 1$, або $\eta^k = \eta^l$, $\xi_1^k > \xi_1^l$, менша, ніж найбільша їх кількість для L_{η^l}. У [4] детально розглянуто цю властивість комбінаторних функцій для задачі а).

Сформулюємо теорему, яка справедлива як для задачі а) так і для задачі б).

Теорема. (Справедлива для $n > 6$ і $1 < \eta^k \le n - 3$). Якщо множина розбиттів Θ упорядкована ізоморфними підмножинами L_{η^k}, L_{η^l} таким чином, що $\eta^k \le \eta^l$ або $\eta^k \ge \eta^l$, $k \ne l$, то цільова функція змінюється за законами кусково-монотонних функцій, а в точці розриву її приріст $\Delta F(\rho^k)$ змінює знак на додатній у випадку кусково-монотонно зростаючої, або від'ємний для кусково-монотонно спадної функції.

5. ВИСНОВКИ

Отже, при розробленні алгоритмів і методів для розв'язання задач розбиття, слід ураховувати описані властивості цих задач. Якщо, наприклад, проводити пошук оптимального значення цільової функції на підмножинах L_{η^k}, L_{η^l}, для яких $\eta^k = \eta^l$, то $F(\rho^k)$ набуває найбільшого значення для $\rho^k \in L_{\eta^k}$, у якого $\xi_1^k > \xi_j^k$, а $\xi_j^k = 1$ для $j = \overline{1, \eta^k}$. Найменшого значення вона набуває для $\rho^l \in L_{\eta^l}$, у якого $\xi_1^k \le \sum_{j=2}^{\eta^k} \xi_j^k$. Якщо для цього випадку при розв'язанні задачі не увести додаткових обмежень, можна одержати некоректний результат.

ЛІТЕРАТУРА

1. Винцюк Т.К. Анализ, распознавание и интерпретация речевых сигналов. К.: Наукова думка, 1987 262 с.

2. Винцюк Т.К. Оптимальное разбиение последовательности элементов на подпоследовательности//Кибернетика. 1970, N 4. С. 128 - 133.

3. Жамбю М. Иерархический кластер анализ и соответствия. М: Финансы и статистика, 1988. 342 с.

4. Тимофієва Н.К. Про властивості комбінаторних функцій в задачах комбінаторної оптимізації//Комп'ютерна математика . Оптимізація обчислень. Зб. наукових праць. - Київ: Ін-т кібернетики ім. В.М.Глушкова НАН України - 2001- том 2. - С. 411 - 417.

ZERO-CROSSING-BASED TEMPORAL SEGMENTATION AND CLASSIFICATION OF AUDIO SIGNALS

Konstantin Biatov, Martha Larson, Stefan Eickeler

Fraunhofer Institute for Media Communication (IMK), Sankt Augustin, Germany
{biatov, larson, stefan.eickeler}@imk.fraunhofer.de

ABSTRACT

The task of audio classification is streamlined if the audio stream can be decomposed in real time into homogeneous segments larger than the individual frame. If segment morphology reflects the underlying audio type, classifier design and training becomes more transparent and can directly exploit real-world knowledge. The experiments presented here explore real-time zero-crossing-based audio stream segmentation and classification. We segment an un-edited audio stream from a German-language radio program into three classes: speech, speaker breathing and music. Features related to modulation of the amplitude envelop are exploited to segment the audio stream into homogeneous sub-syllable segments. The segments prove useful for classifying the audio-stream, especially the speech portions. We use features extracted from a zero-crossing-based pseudo-spectrum to train a Gaussian classifier that classifies audio content into the three categories. Our results demonstrate that our zero-crossing-based segmentation and classification method is a viable one, yielding satisfactory performance in real time.

1 INTRODUCTION

Progress of audio processing technologies has led to an expanding role for audio as an information source and a communication medium. The increasing importance of audio data has inspired the vision that that audio should be just as easily segmentable, indexable, searchable and even as translatable as text. Motivated by this vision is the move away from frequency-domain and other calculation intensive approaches, wherever there are not justified by the task or the medium.

This paper is an exploratory investigation in the direction of real-time segmentation and classification of audio data. This work was driven by two central insights. First, audio processing does not have to take place on the frame level, but rather that benefit is to be derived by matching the method to the medium. In our experiments we use modulations in the amplitude envelop to first decompose the audio stream into segments motivated by signal morphology, which in the case of speech turn out to be sub-syllables. Second, audio processing does not have to take place in the frequency domain. We introduce a pseudo-spectrum derived from the distribution of zero-crossings and show that features can be extracted from it that yield reliable classification into audio categories. We segment an un-edited audio stream from a German-language radio program into three classes: speech, speaker breathing and music.

Our experiments draw motivation from the literature on segmentation and classification based on larger-than-frame segments. A two pass classification was introduced in [1]. In the first pass, speech/non-speech boundaries were detected using a reduced phoneme inventory and gender-independent acoustic models. The second pass identified speaker changes, using only the boundaries detected in the first pass as potential candidates. Although this approach increased segmentation accuracy, it also increased segmentation time.

Several approaches have decomposed speech into syllable and syllable-like units. An automatic segmentation of speech into syllable units based on the convex hull of the loudness function in described in [2]. Relative loudness maximums are interpreted as potential syllabic peaks and relative loudness minimums as potential syllabic boundaries. A local loudness minimum separated from another local minimum by less than 100 ms may be insignificant, but a minimum of the same magnitude which had no other minimums within 500 ms would indicate a syllable boundary.

Automatic detection of syllable boundaries based on full-band energy contour and voicing detection using modified autocorrelation is described in [3]. The syllable boundary is determined to be the point of minimum energy contour within each portion of consecutive possibly unvoiced frames.

[4] describes a new representational format for speech, the modulation spectrogram, that represents amplitude modulation frequencies in the speech signal between 0 and 8 Hz, with the peak at 4 Hz corresponding to modulation spectrum of speech. The modulation spectrogram robustly extracts information related with the syllabic segmentation of speech.

The literature on zero-crossing rate is a growing corpus, and also served as an inspiration for our approach. Possibilities for spectral analysis based on zero-crossing are described in [5]. In [6] a spectral analyzer based on zero-crossing is demonstrated. Using zero-crossing interval measurement for formant frequency estimation in noise is presented in [7]. Experiments on speech recognition in a noisy, real–world environment in which the frequency information of the signal is obtained from zero-crossing intervals are described [8]. In [9] experiments which use gravity centers of energy as additional feature to the classical set of MFCC in automatic speech recognition system are described. These experiments demonstrate improved recognition performance when gravity centers are computed from zero-crossing intervals detected at the output of the filters of an ear model.

Using just zero-crossings for speech/music discrimination has been investigated previously in [10]. Other research on speech/music discrimination and classification has used temporal information from speech waveform and as well as spectral information. Spectral information has include such features as 4 Hz modulation energy, rolloff of the spectrum, spectral flux and spectral centroid. In this paper we would like to demonstrate that features like these can be extracted from a zero-crossing-based pseudo-spectrum and used to classify and audio stream with satisfactory accuracy.

In the next section of this paper, section 2, we describe our zero-crossing-based downsampling of the time-signal and present motivation for the approach. We introduce our method for extracting sub-syllable segments by looking at modulations in the amplitude envelop. We detail the calculation of the zero-crossing-based pseudo-spectrum and the features we derive from it. In section 4 we describe our classification experiments. Section 5 presents conclusions and outlook.

2 PSEUDO-SPECTRAL ANALYSIS

2.1 Zero-crossing-based Smoothing

Assuming that signal is normalized, i.e. the mean has been removed, we define zero crossings, t_n, $n = 1,2,3\ldots$, as the times at which the signal changes sign or is equal to zero. The successive zero-crossing intervals are defined as $z_n = t_n - t_{n-1}$. As discussed in [7] the successive zero-crossing intervals of sinusoidal signal exhibit high consistency and are inversely related to frequency. The signal resulting from the sum of more than one sinusoidal component, however, may not satisfy this principle. Usually the speech signal between two zero-crossings has more than one sinusoid. To be sure that the representation of speech signal as one sinusoid between two zero-crossings is valid for the purpose of discrimination we have implemented the method described in [12]. The audio signal is represented as a sequence of successive zero-crossing intervals and the maximal signal amplitudes for those intervals where amplitude is positive and minimal amplitude for the intervals where amplitude is negative. The signal is represented as a sequence (z_n, A_n), $n=1,2,3,\ldots$

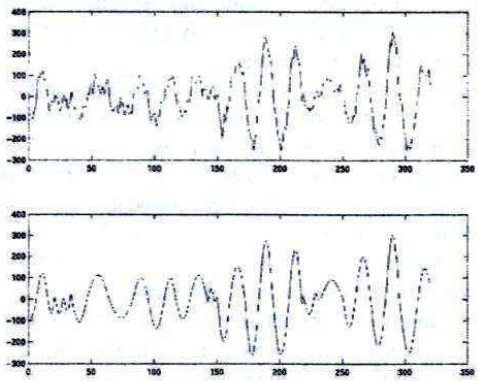


Figure 1. Comparison of original signal and recovered signals

An audio signal restored from this sequence retains sufficient quality to be understood by a human listener, and therefore holds good potential for discriminating speech, music, song, and other non-speech signals. Inconsistencies in the relationship between frequency and zero-crossing rate in the speech signal, remain well contained. From Figure 1 an impression can be gained of exactly how close the original speech signal is to the speech signal restored after zero-crossing-based downsampling.

According to Kedem [5] the zero-crossing rate of repeatedly differentiated series converges to its least upper bound. The restored signal has the same zero-crossing rate as its derivative. Since we are assured of an least upper bound, we are justified in considering that in each successive interval we have only one sinusoid and in using the inverse of the length of the interval as a frequency of the signal in this interval.

2.2 Extracting larger-than-frame segments

The changes that are relevant for segmenting and classifying a radio program, are not those that occur on a frame to frame, but rather those which occur between some higher-level content based units. Audio content processing can be streamlined if a higher-level segment can be found, which is identifiable in the signal, but whose morphology reflects underlying audio content.

Segments which are morphologically justified are also to be homogeneous. We want to exploit facts about the audio stream such as that a strain of music is never going to separate two halves of a syllable of speech. We are interested in morphological properties of the signal that correspond to detectable sections of the audio input. For breathing this might be individual breaths, and for music, the notes. Here we concentrate on extracting a basic unit for spoken audio, namely the syllable. We find that sub-syllables, units that are useful since they never straddle syllable boundaries and can be identified reliably using the minima in the amplitude envelop of the signal.

The signal is divided into frames, each 20 ms seconds long with an overlap of 10 ms. For each frame we extract the maximal non-negative amplitude from the set of amplitudes A_n, $n=1,2,3,\ldots$ corresponding to interval between successive zero-crossing, z_n, A_n $n=1,2,3,\ldots$ On a new signal composed from these maximum amplitudes we find the points which are local minima. These are the boundaries of our larger-than-frame units.

2.3 Deriving the Pseudo-spectrum

The zero-crossing-based pseudo-spectrum is calculated from the restored signal described in sub-section 2.1 using the following procedure. From each 20 ms frame (overlap 10 ms) we extract the lengths of the intervals between successive zero-crossings. Then extracted periods are converted into pseudo-frequencies, by $f = 1/z$, where z is the length of the interval between successive zero-crossings. For each interval between successive zero-crossings we extract the amplitude that is maximal for the interval with positive sign of the signal, and is minimal for the interval with negative sign of signal. Then using 24 frequency bands according to the Bark scale, we group frequencies that correspond to one band together and then normalize all groups using sum of all frequencies. As the result for each 20 ms frame we have smoothed a pseudo-spectrum which can be deployed for a range of discrimination tasks.

The resulting pseudo-spectrum is well smoothed, but retains the discriminative features of the original speech signal. Figure 3 illustrates the pseudo-spectrum corresponding to the 16 kHz waveform depicted in Figure 2. These figures motivate the use

of the pseudo-spectrum distinguish voiced and unvoiced parts of the signal and identify the syllabic modulation of the speech signal. Figure 4 is the pseudo-spectrogram of a waveform in which the speaker breathes between words or phrases. The duration of the breath is 100 ms. Once again, visual inspection confirms that the pseudo-spectrum contains enough information about the signal to allow a discrimination to be made.

Using the amplitudes which correspond to each zero-crossing interval we calculate a value related to energy.

$$S = \frac{abs(z_n * A_n)}{2} \qquad (1)$$

S approximates the square of the amplitude of the envelop corresponding to interval z_n. We group together all S values corresponding to the same Bark band and normalize these values by dividing by total sum for all intervals. In this way we produce a pseudo-spectrum representing the distribution of energy with respect to frequency bands.

We use this pseudo-energy spectrum to calculate an additional feature, the rolloff point. We define our rolloff point to be the frequency below which 85% of the energy in the spectrum is concentrated.

Additionally we use the centers of gravity of the Bark-spectrum frequency bands, proposed in [9].

$$CG = \frac{\sum_{i=1}^{n} A_i z_i}{\sum_{i=1}^{n} A_i} \qquad (2)$$

From the pseudo-spectrum we also calculate the frequency corresponding to the maximum zero-crossing interval in each frame. The final feature that we derive from the pseudo-spectrum is the Euclidean distance between the pseudo-spectral vectors of neighboring frames.

3 EXPERIMENTS

We experimented on data from our *Kalenderblatt* database, which contains radio programs recorded from the Deutsche Welle *Kalenderblatt* series and the accompanying transcripts. The programs are five minutes in length and each contain about 650 running words. They are liberally interspersed with music and other sound effects. Although each program is narrated by a single speaker, many other voices are present in the form of interviews and original sound footage. The radio programs were downloaded from the Internet (http://www.kalenderblatt.de) and resampled to 16 kHz. The corresponding text transcripts were also downloaded and normalized. The database is described in detail in [14]. Also in [14] is a description of the semi-automatic alignment visualization tool with which we create reference segmentations and segment labels for our experiments.

For the exploratory experiments presented here, we indiscriminately chose one 5-minute program for training and one for testing. We segmented and labeled these with the semi-automatic method down to the syllable level, also marking breathing and music. In this way we generated high-quality reference labels for training and testing. The duration of training data for the experiments presented here is 5 minutes; the duration of test data is also 5 minutes.

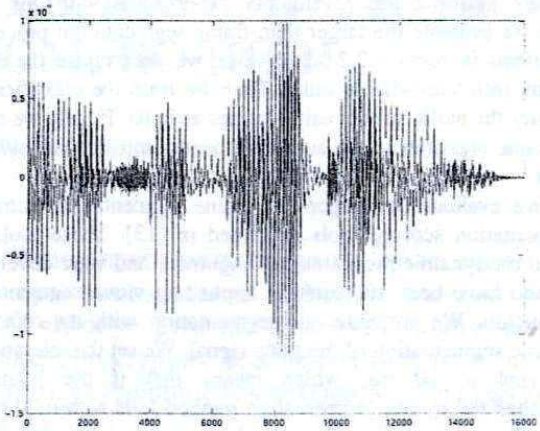


Figure 2. Waveform of speech signal. (20 ms frame with 10 ms step represented on x-axis)

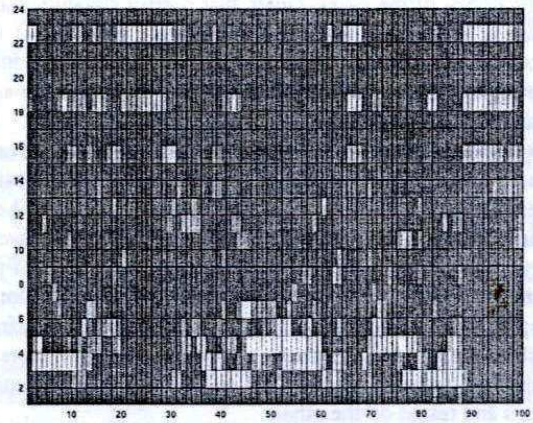


Figure 3: Pseudo spectrum for the waveform depicted in Figure 2. (Bark-scale pseudo-frequencies vs. frames)

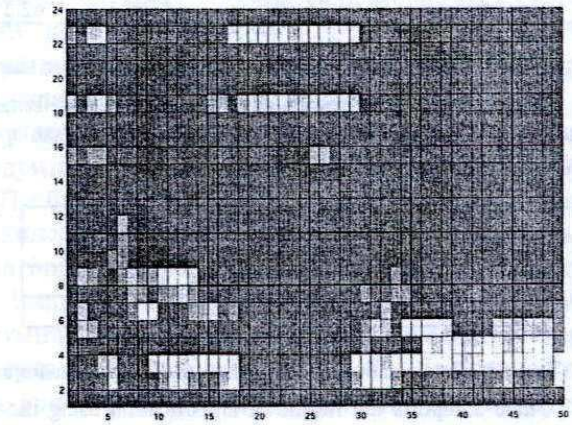


Figure 4: Pseudo spectrum for the waveform containing speaker breathing (Bark-scale pseudo-frequencies vs. frames)

We perform three evaluatory experiments with this data. First we evaluate the larger-than-frame segmentation procedure described in section 2.2 with which we decompose the audio stream into sub-syllable units. Then we train the classifier and classify the audio stream using frames as units. Finally we apply the same classifier to the audio segmented into the sub-syllable units.

We evaluate our larger-than-frame segmentation using the segmentation scoring tools described in [13]. These tools are based on dynamic programming alignment and were developed for and have been successfully applied to video segmentation evaluation. We compare our segmentation with the reference syllable segmentation of the same signal. We set the tolerance of the tool to 20 ms, which means that if the boundary hypothesized by our segmentation method falls within 20 ms of the reference boundary, the boundary is considered correct. 20 ms represents a conservatively narrow margin of tolerance, since an average syllable is 250 ms in length. Our segmentation scores 87.5% insertions and 12.5% deletions with respect to the reference segments. This means that our segments in most cases are proper sub-units of syllables and cross syllable boundaries in only 12.5% of the cases. Given that correct syllabification in German is partially a matter of convention, we feel that these results indicate that our larger-than-frame segments are indeed capturing a sub-unit that provides a useful link between signal structure and audio content.

For classification we use a maximum likelihood multivariate Gaussian classifier. The input vectors for the classifier consist of the features extracted from the pseudo-spectrum (described in section 2.3) plus their variances. These features are: average energy, gravity centers of pseudo-frequency bands, rolloff point of energy distribution, frequency corresponding to the longest interval between successive zero-crossings present in frame, Euclidean distance between pseudo-spectral vectors of neighboring frames. The classifier was trained on one 5 minute program and tested on the other.

	classified as music	classified as breath.	classified as speech
music data	**84.2%**	3.7%	11.1%
breathing data	8.2%	**89.2%**	2.6%
speech data	25.6%	12.2%	**62.2%**

Table 1: Classification on frames (20 ms, with 10 ms step)

Table 1 reports the results of classification using frames as the units. Music and breathing are classified better than speech using frame as the basic classification unit.

	classified as music	classified as speech
music data	**68%**	32%
speech data	6%	**94%**

Table 2: Classification on sub-syllables (larger-than-frame)

Table 2 reports the results of classification using the sub-syllable units. To score this classification, we look at which frames are contained in the sub-syllable and see which class label is associated with the majority of them in the reference segmentation. Speech is classified better than music using the sub-syllable units. We feel that this result reflects the utility of choosing a larger-than-frame unit that is related to the morphology of the audio content.

4 CONCLUSIONS

The zero-crossing pseudo-spectrum proposed here has proven itself to be a valuable source of features for real-time audio classification and has been shown to allow effective discrimination between speech, speaker breathing and music. Further results suggest that larger-than-frame units enhance classifier performance, but that they must be well-matched with the audio content.

Future work will focus on identifying additional features that can be derived from the pseudo-spectrum that might aid classification of audio, particularly into classes above and beyond the three we experiment with here. Finally, we hope that further work will allow us to develop a larger-than-frame segment that will be as useful for music and breath classification as the sub-syllable units are for speech classification.

5 REFERENCES

[1] Makhoul, J. et al. (2000). Speech and Language Technologies for Audio Indexing and Retrieval: Advances and Applications. Proceedings of the IEEE, 88:1338-1353.

[2] Mermelstein, P. (1975). Automatic segmentation of speech into syllabic units. Journal of the Acoustic Society of America, vol. 58, No. 4, pp.880-883.

[3] Sakaguchi, S., Arai, T. and Murahara, Y. (2000). The effect of polarity of speech on human perception and data hiding as an application. ICASSP 2000.

[4] Greenberg, S. and Kingsbury, B. (1997). The modulation spectrogram: in pursuit of an invariant representation of speech. ICASSP 1997, pp. 1647- 1650.

[5] Kedem, B. (1986). Spectral analysis and discrimination by zero-crossing. Proceedings of IEEE, VOL. 74, No. 11, pp. 1477-1493.

[6] Kay, S. A zero-crossing based spectrum analyzer. (1986). IEEE Transaction on Acoustic, Speech and Signal Processing, vol. ASSP-34, No. 1, pp. 96-104.

[7] Sreenivas, T. and Niederjohn, R. (1992). Zero-crossing based spectral analysis and SVD spectral analysis for formant frequency estimation in noise. IEEE Transaction on Signal Processing, vol. 40, No. 2, pp. 282-293.

[8] Kim, D.S., Lee, S.Y. and Kil, R.M. (1999). Auditory processing of speech signal for robust speech recognition in real-world noise environment. IEEE Transaction on Speech and Audio Processing, 7, No. 1, pp. 59-69.

[9] De Mori, R., Moisa, L., Gemello, R., Mana, F. and Albesano, D. (2001). Augmenting standard speech recognition features with energy gravity centres. Computer Speech and Language, 15, pp. 341-354.

[10] Saunders, J. (1996). Real-Time Discrimination of Broadcast Speech/Music", Proc. ICASSP 1996, pp. 993-996.

[11] Scheirer, E. and Slaney, M. Construction and evaluation of a robust multifeature speech/music discriminator. ICASSP 97, vol. 2, pp. 1331-1334.

[12] Vintsiuk, T., Kulias, A. and Dys, A. (1982). A. economic presentation of acoustic signal in computer. ARSO-12, Institute of Cybernetics of AS USSR.

[13] Eickeler, S. and Rigoll, G. (2000). A novel error measure for the evaluation of video indexing systems. ICASSP 2000.

[14] Eickeler, S., Larson, M., Rüter, W. and Köhler, J., (2002) Creation of an Annotated German Broadcast Speech Database for Spoken Document Retrieval," to appear, Proceedings of LREC2002.

ДЕТЕКТОР ЗВУКОВИХ ОБРАЗІВ, ЩО ЕМІТУЮТЬСЯ СЛУХОВИМ ОРГАНОМ ЛЮДИНИ

***В.Г. Грибенко**, інженер звукотехніки, м. Яготин Київської обл.*
abris 2001 @ hotmail. com.

Анотація.

Явище емісії звукових коливань периферійним слуховим органом людини залишається загадковим у нейрофізіологічному сенсі [1]. Дослідження Кемпа [2, 3] стимульованої та спонтанної отоакустичної емісії започаткували вивчення механізму цього явища.

Для кращого розуміння походження сигналів, генерованих слуховими центрами мозку людини в процесі мислення, та їх перетворення і трансляції, автор пропонує новий підхід у дослідженні механізму слуху людини на порозі чутності [4] та бачення слухового органу не тільки як пасивного приймача зовнішніх слухових коливань, але і активного органу емісії звукових образів, що відповідають думкам людини.

Пропонується ідея іноваційного проекту створення оптоелектронного отоендоскопа – приладу, за допомогою якого стане можливою селекція інформації з шуму коливань барабанної перетинки та ідентифікація її з мовними музичними та іншими звуковими образами, що відповідають певній розумовій діяльності.

Вступ.
Нейрофізіологічна версія розпізнавальної функції слухових структур мозку людини

В анатомії мозку тварин розрізняють декілька слухових нервових шляхів, що місцями перетинаються між собою [5].

Висхідні шляхи – це послідовна сукупність асоційованих нервових слухових утворень з тонотопічною організацією від органа Корті до височних ділянок кори.

Ці шляхи мають численні зв'язки з ретикулярною формацією мозку – сукупністю структур в центральних його відділах, які регулюють рівень збудження відповідних ділянок центральної нервової системи з корою великих півкуль включно.

Зворотні шляхи (на жаль маловивчені) починаються з відповідних ділянок кори і через нижчі відділи слухової системи з'єднуються з органом Корті [6].

Аналізуючи слухову систему мозку (ССМ) людини можна припустити наявність селективного зворотного зв'язку між органом Корті і відповідними відділами мозку.

Автор припускається думки, що на порозі чутності при розпізнаванні зовнішніх звукових образів ССМ виконує функцію багатоканального надрегенератора структурних фрагментів імпульсів, що відповідають внутрішнім усвідомленим слуховим образам (рис.1).

Ці фрагменти імпульсів після селекції та адресації в ССМ попадають в орган Корті, де перетворюються у відповідний спектр механічних коливань, які через внутрішній і середній відділи вуха передаються на поверхню барабанної перетинки (ПБП) у вигляді фрагментів внутрішніх, але вже звукових образів.

За відсутності зовнішніх звукових образів ПБП під дією вушного шуму постійно знаходиться у коливальному стані. Спектральні характеристики цього шуму контролюються ССМ [7].

ССМ, маніпулюючи довжиною фрагментів слухових імпульсів, їх інтенсивністю, формою, місцем та часом генерації через функцію органа Корті підлаштовують рух ПБП синхронно її коливанню, збудженому дією фрагмента зовнішнього звукового образу на порозі чутності, чим збільшують його амплітуду, полегшуючи розпізнавання.

При асинхронній модуляції руху ПБП фрагментами зовнішнього і внутрішнього звукових образів, інформація про невідповідність динаміки коливань ПБП динаміці фрагмента внутрішнього звукового, а значить і слухового образу, перетворюється і передається до ССМ органом Корті, чим викликається негайна корекція слухового фрагмента, що сприяє розпізнаванню та усвідомленню образа.

Насамкінець, можливо припустити, що в стані роздумів у "повній тиші" (Nзовн.шуму ≤ 0 дБ) на ПБП з боку ССМ надходять лише звукові коливання, обумовлені думкою та іншою життєдіяльною та патологічною емісією (т.ч. ми "слухаємо" свої думки).

Вищенаведені припущення дещо полегшують розуміння феноменально високої локальної чутливості вуха в діапазоні частот 1 – 5 кГц, бо чутливість до синхронного збурення динамічних систем суттєво вища, ніж статичних.

Принцип роботи детектора власних коливань барабанної перетинки та вимоги до його конструкції

Подумки поставимо себе перед необхідністю визначити у темному місці дотиком динаміку рухомого предмета і його вібрацію.

При випадковому короткочасному дотику результат напевне буде невизначеним. Задача спрощується, якщо рукою супроводжувати предмет, не порушуючи траєкторії і режиму його руху.

Подібну механічну аналогію можливо розповсюдити на принцип розпізнавання структури звукових коливань мембрани за допомогою рухомого на відповідній відстані від ПБП детектора.

Необхідність гранично високої чутливості детектора до амплітуди вібрації ПБП (атомного масштабу) і водночас достатності його динамічного діапазону для компенсації значних амплітуд неінформативних звукових збуджень від процесів дихання, серцево-судинної діяльності, стоматофонової та шлункової активності, а також інших спонтанних звукових реакцій ПБП обумовлена слідуючими головними засадами до проекту детектора:

перетворювання фізичних параметрів у датчику повинно бути механо-оптичним задля уникнення значного біоелектронного шуму в зоні розташування перетворюючого елемента детектора;

торець мікросвітловода необхідно покрити напівпрозорим дзеркальним шаром металу, мікрооб'єм тіла барабанної перетинки безпосередньо в зоні торця та проміжок між ними при вібрації ПБП повинні створювати модулююче промінь світла середовище, тобто мікрооб'єм тіла ПБП в її центрі повинен бути елементом конструкції перетворювача в сенсі оптимальності його використання як модулюючого відбите світло середовища.

Фізичний процес в оптичному мікроконтакті грунтується на природній модуляції енергії індукованого випромінювання коливанням ПБП у квазірезонаторі, створеному напівпрозорим дзеркалом торця мікросвітловода та відбиваючим частину світлового променя об'ємом на ПБП.

Коректне апаратурне втілення такого принципу зробить можливим створення мікрооптичного контакту з необхідним ступенем модуляції відбитого від ПБП світла.

Оптимальний режим роботи мікрооптичного контакту визначиться пороговим рівнем автоматичного утримання його параметрів, а також значенням інтенсивності опорного променя без ризику деструкції мікроконтактної зони на ПБП.

Слід зазначити, що запропонований до розгляду детектор повинен бути невід'ємною частиною мультипроцесорної системи оброблення та розпізнавання динамічних образів з граничною якістю специфікацій.

Ескіз пілот-конструкції детектора отоендоскопа

Пара детекторів власних коливань ПБП складає вхідну частину отоендоскопа і має бути симетрично розміщена на звукопоглинаючому оголів'ї згідно попередньо одержаних топограм зовнішніх слухових каналів (ЗСК) голови пацієнта.

Наступний процес юстирування полягає у приведенні детекторів в положення (див. рис.2) відносно зовнішньої поверхні барабанної перетинки 1, при якому між ними виникає мікрооптичний контакт.

В ЗСК входить гнучкий звукопоглинаючий штуцер 2 з трьома капілярами 2.1 і пневмопорожнинами 2.2 на їх кінцях, утвореними вклеєним в штуцер 2 платиноіридієвим розрізним з антифрикційною обробкою внутрішньої його поверхні циліндром 2.3 та неприклеєними до нього ділянками матеріалу штуцера 2 в кінці кожного капіляра 2.1.

Таким чином, на кінці штуцера 2 утворюються пневмопорожнини, які окремо регулюються по висоті, що дає можливість фіксувати штуцер 2 у необхідному положенні згідно параметрів ЗСК.

Всередині штуцера 2 вздовж його осі знаходиться рухомий зонд, який складається з гнучкого електропровідного і звукопоглинаючого штока 3, з'єднаного з платиноіридієвим наконечником 4; п'єзоелемента 5, закріпленого електропровідним цементом в фасонному пазі наконечника 4; та оптоволоконного трійника 6 світловода.

Шток 3 і наконечник 4 мають співпадаючі канавки, в яких залягає трійник 6 світловода [8], а також ізольований провідник 5.1 для подачі напруги на п'єзоелемент 5.

Закріплена на п'єзоелементі 5 частина 6.1 трійника має конічну форму і радіальний градієнт показника заломлення, а потім переходить у короткий відрізок одномодового волоконного світловоду, торець якого покритий напівпрозорим дзеркальним шаром металу. Мікрооптичний контакт створюється між торцем частини 6.1 трійника і ПБП.

Частина 6.2 трійника підключена до світловипромінюючого діода 7, закріпленого в канавці штока 3, а інформаційна частина 6.3 трійника – до мікроканального ФЕП – дінода 8 (далі по тексту дінода 8).

Шток 3 і світловипромінюючий діод 7 приклеєні до голчастого радіатора 9, який рухається трьома реверсивними лінійними п'єзоелектричними мікродвигунами 10.

Таким чином, зонд зміщується з робочої точки в неробочу і навпаки в штуцері 2 як в пеналі – запобіжнику від пошкодження частини 6.1 трійника.

Радіатор 9 виконаний як тепловідвід термоелектричного мінікріоблока 11, який охолоджує дінод 8, вмонтований в колектор 11.3 холодних спаїв.

На радіаторі 9 закріплена металічна плата 12 з двобічним монтажем екранованих мікропотужних НВІС – процесорів, обслуговуючих датчик.

Мікродвигуни 10 вклеєні в корпус-кондуктор 13 з трьома карпілярами 13.1 сумісними з капілярами 2.1.

Дінод 8 має в центральній своїй частині 8.1 металокерамічний спай 8.2 для її температурного обмеження.

Місце припайки керамічних анодної 8.3 та катодної 8.4 частин дінода 8 до центральної його

Рис. 1. Структурно-функціональна схема системи виявлення звукових сигналів, що емітуються органом Корті

Аналоговий зв'язок при Р зовн. збудж. > 0 дБ

Аналоговий зв'язок при Р зовн. збудж. < 0 дБ

Σ Багатоканальний дискретний зв'язок

Σ Зворотний зв'зок

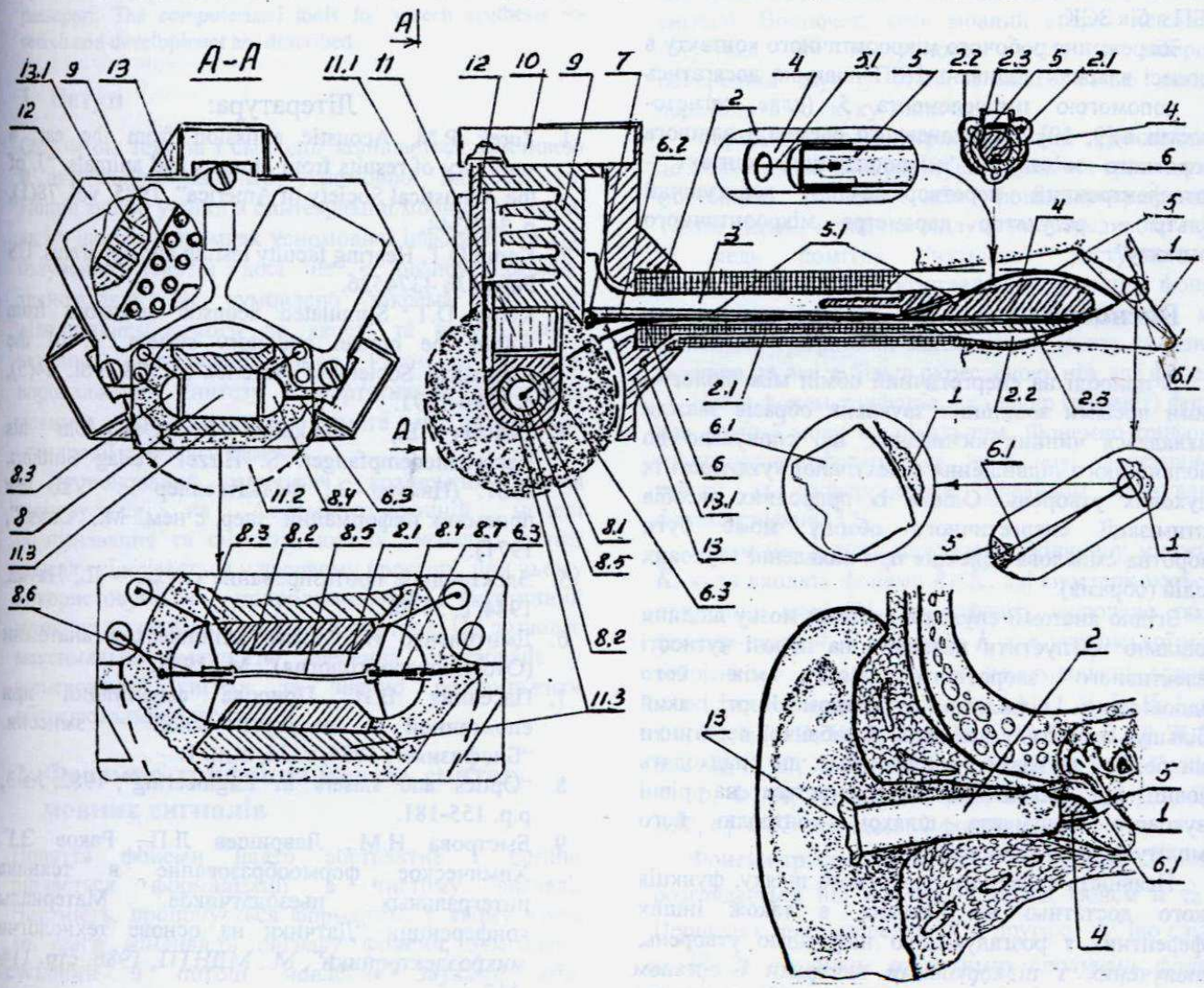


Рис. 2. Конструкція датчика отоендоскопа.

1 – барабанна перетинка; 2 – штуцер (2.1 – капіляр; 2.2 - пневмопорожнина; 2.3 – розрізний циліндр); 3 - шток; 4 - наконечник; 5 – п'єзоелемент (5.1 – провідник); 6 – трійник світловода (6.1 – приймальна частина; 6.2 – опорна частина; 6.3 – інформаційна частина); 7 – світловипромінюючий діод; 8 – дінод (8.1 – центральна частина; 8.2 – металокерамічний спай; 8.3 – анодна частина; 8.4 – катодна частина; 8.5 – мікроканал; 8.6 – колектор електронів); 9 – радіатор; 10 – п'єзоелектричний двигун; 11 – мінікріоблок (11.1 – комутаційна панелька; 11.2 – зовнішній теплоізолятор).

частини 8.1 є відповідно анодом і катодом, які електрично з'єднані між собою покриттям мікроканала 8.5.

Колектор електронів 8.6 електроізольований від анода 8.3.

В катодну частину 8.4 впаяний світловод 8.7.

Коефіцієнт посилення фотоструму дінода 8 не слід робити занадто високим в порівнянні із звичайними ФЕП. Він має бути достатнім для його роботи в режимі модуляції т.з. "жирного нуля" відповідного рівню шуму входу синхронного посилювача, підключеного до колектора 8.6.

Таким чином, радіатор 9 є авторухомою платформою, яка приводить частину 6.1 трійника до мікрооптичного з ПБП контакту і, обов'язково, за мить до максимального відхилення робочої зони ПБП в бік ЗСК.

Збереження робочого мікрооптичного контакту в процесі власних коливань ПБП повинно досягатись за допомогою п'єзоелемента 5 (клас сегнетоеластиків)[9, 10] як виконавчого елемента ланцюга зворотного зв'язку – "мікрооптичний контакт – фотоелектронний перетворювач – коригуючий фільтр – регулятор параметра мікрооптичного контакту".

Висновки.

В природі на енергетичний обмін між біологічними носіями зовнішніх звукових образів завжди накладався чинник виживання, що спонукало до еволюційного підвищення селективної чутливості їх слухових утворень. Одним із природних засобів оптимізації енергетичного обміну може бути зворотна смислова селекція при виявленні звукових подій (образів).

Згідно анатомії слухової системи мозку людини доцільно припустити наявність на порозі чутності селективного зворотного зв'язку між його відповідними відділами і органом Корті, який збільшує амплітуду коливань барабанної перетинки синхронно звуковим коливанням, що надходять ззовні. Це збільшення забезпечується на рівні звукового фрагмента шляхом контролю його амплітуди та швидкості її зміни.

Наявність складного зворотного шляху, функція якого достатньо не вивчена, а також інших аферентних з розгалуженою інервацією утворень, сполучених з підкорковими центрами і органом Корті, вказує на зворотний напрямок імпульсації і, не виключено, що вона пов'язана також із свідомою вербальною діяльністю мозку. Відповідь на це питання можуть дати тільки точні експерименти.

Виявлення спонтанної отоакустичної емісії стало можливим завдяки впровадженню електретних мікрофонів, але принципова наявність значних шумів стримує їх використання в сенсі підвищення чутливості відповідного експериментального обладнання. Тому був запропонований більш чутливий метод виявлення власних коливань зовнішньої поверхні барабанної перетинки, започаткований на способі модуляції нею інтенсивності світлового променя, тим більш, що сучасний рівень значної низки технологій дозволяє провести репрезентативне дослідження тонкого спектру власних коливань барабанної перетинки в діапазоні збуджень на порозі чутності.

* * *

Автор щиро вдячний шановному голові оргкомітету "УКРОБРАЗ-2002" п. доктору Вінцюку Т.К. за пропозицію оприлюднити цю неоднозначну ідею.

Література:

1. Zurek P.M. Acoustic emission from the ear: a summary of results from humans and animals. "J. of the Acoustical Society of America", 1985, vol. 78(1), p. 340-344.
2. Kemp D.T. Hearing faculty testing and apparatus. US Patent № 4374526.
3. Kemp D.T. Stimulated acoustic emissions from within the human audiotory system "J. of the Acoustical Society of America", 1978, vol. 64(5), p. 1386-1391.
4. Zwicker E., Feldtkeller R. Das Ohr als Nachrichtenempfanger. S. Hirzel Verlag Stuthart. 1967. (Цвикер Э., Фельдткеллер Р. "Ухо как приемник информации", пер. с нем.. М., "Связъ", 1971).
5. Электродное протезирование слуха. – Л., Наука, 1984, с. 53-80.
6. Дмитриенко И. "Атлас клинической анатомии (Оториноларингология)", М., 1998.
7. Пасечник В.И. Природа флуктуаций при спонтанной отоакустической эмиссии. "Биофизика", 1984, т. 34, вып. 1.
8. "Optics and Lasers in Engineering", 1982, № 3, p.p. 155-181.
9. Быстрова Н.М., Лаврищев Л.П., Раков Э.Г. Химическое формообразование в технике интегральных пьезодатчиков. Материалы конференции "Датчики на основе технологии микроэлектроники", М., МДНТП, 1986, стр. 114-117.
10. J. van Randaraat, Setterlington R. E. Piezoelectric Geramics. Eds. Eindhoven: N.V. Philips, 1974.

АВТОМАТИЧНИЙ ОЗВУЧУВАЧ УКРАЇНСЬКИХ ТЕКСТІВ НА ОСНОВІ ФОНЕМНО-ТРИФОННОЇ МОДЕЛІ З ВИКОРИСТАННЯМ ПРИРОДНОГО МОВНОГО СИГНАЛУ

Тарас Вінцюк, Микола Сажок, Тетяна Людовик, Руслан Селюх

Міжнародний науково-навчальний центр інформаційних технологій та систем
40 просп. Академіка Глушкова, Київ 03680
Електронна пошта: {vintsiuk, mykola, tetyana_lyudovyk, selyukh}@uasoiro.org.ua

Taras Vintsiuk, Mykola Sazhok, Tetyana Lyudovyk, Ruslan Selyukh. **Automatic Ukrainian Text-to-Speech System Based on Phoneme-Threephone Model Using Natural Spoken Signal.** The text-to-speech system in time domain for Ukrainian is described. The concatenated acoustic elements are chosen in accordance to phoneme-threephone model for speech synthesis. Acoustical data is taken from the speaker voice passeport. The computerized tools for speech synthesis research and development are described.

1 Вступ

Озвучення текстів і сьогодні залишається важливою й актуальною задачею усномовної інформатики. Попри значні успіхи в синтезуванні мовних сигналів, як і в інших напрямках усномовної інформатики [1], озвучення текстів досі не є розповсюдженою технологією. Це зумовлено зокрема тим, що підвищились вимоги до якості та натуральності звучання синтезованої мови. Разом з тим для впровадження синтезу в портативних пристроях повинна задовольнятися вимога на обмеження швидкодії та обсяги пам'яті.

Пропонований озвучувач українських текстів засновується на фонемно-трифонній моделі розпізнавання та синтезу мовних сигналів, синтез сигналу відбувається у часовому просторі, при цьому використовується природномовний акустичний матеріал з усномовного файлу диктора. Це дозволяє максимально зменшити внесення спотворень у згенерований сигнал та значно розвантажити обчислювальний модуль.

2 Фонемно-трифонна модель синтезу мовних сигналів

Поняття фонеми надто абстрактне і погано піддається формалізації в чистому вигляді. Натомість, пропонується формалізм, у якому взято до уваги мінливість сигналу фонеми, зумовлену сусідніми в потоці мовлення звуками або коартикуляцією. Отже, вводимо поняття фонеми-трифона, коли розглядається фонема в контексті з попередньою та наступною фонемами. Поняття фонеми-трифона покладається в основу акустичних моделей як розпізнавання, так і синтезу усної мови.

Фонемно-трифонна акустична модель усної мови дозволяє врахувати явище коартикуляції, що виникає при взаємодії звуків у потоці мовлення. Справді, при відтворюванні послідовності звуків, що відповідають певним фонемам, рухи мовного апарату людини відбуваються з певною інерційністю. Стан мовного апарату перед наступним рухом залежить від попереднього звуку (фонеми), а отже і динаміка рухів мовного апарату при переході до наступної фонеми різна в залежності від попередньої фонеми, а це, зрештою, якісно відбивається на акустичному сигналі. Водночас, весь мовний апарат немов би готується до наступного звуку, і завершує попередній звук у стані, з якого більш вигідно переходити до звуку, який слідує.

Формалізм фонеми-трифона, який пропонується, дозволяє також конкретизувати поняття границь між фонемами, точніше, між фонемами-трифонами. Таким чином, незначне відлуння попереднього звуку та ледь помітна "чутність" наступного є допустимими при розставлянні границь фонем-трифонів. Уцілому, розставляння границь між фонемами-трифонами залишається досить складною задачею, та все ж більш окресленою, ніж для фонем.

Набір фонем-трифонів, як і набір (алфавіт) фонем для кожної мови є унікальним. Фонемно-трифонна транскрипція формується на основі фонетичного тексту за універсальним правилом стикування фонем-трифонів [2].

Виходимо з того, що задано скінченну множину K, куди входять фонеми $k \in K$, які спостерігаються в природній мові [3]. До алфавіту включено також фонему-паузу #. У множині K для української мови розрізняємо наголошені та ненаголошені голосні, м'які та тверді приголосні: $k \in$ {А, О, У, Е, И, І, А1, О1, У1, Е1, И1, І1, Б, Б', В, В', Г, Г', Ґ, Ґ', Д, Д', Ж, Ж', З, З', Й, К, К', Л, Л', М, М', Н, Н', П, П', Р, Р', С, С', Т, Т', Ф, Ф', Х, Х', Ц, Ц', Ч, Ч', Ш, Ш', ДЗ, ДЗ', ДЖ, ДЖ', #} $\equiv K$ – загалом 57 фонем.

Фонема-трифон $t=uWv$ є фонемою W, яка розглядається під впливом сусідніх фонем u та v. Першою є u, що передує W, а другою – v, що слідує за W. За правилом допустимих сполучень фонем-трифонів [2] допустимими для з'єднання є лише фонеми-трифони вигляду $t_1=uWv$ і $t_2=wVz$ через Wv та wV.

Загальна кількість фонем-трифонів у алфавіті теоретично дорівнює кількості базових фонем у степені три (125000 для алфавіту з 50 фонем). Практично ж можна обмежитися декількома тисячами фонем-трифонів. Відсутні фонеми-трифони

замінюються найближчою згідно з фонемно-трифонною ієрархією.

З природи усномовного сигналу випливає, що дзвінкі фонеми можна описувати (транскрибувати) як послідовність одно-квазіперіодичних мікросегментів, що мають певну форму звукової хвилі та довжину (Рис. 1). Такий опис поширюємо також і на глухі (шумні) фонеми і називатимемо акустичною транскрипцією фонеми-трифона. Послідовність хвиль одноквазіперіодичних мікросегментів утворює акустичний образ (прототип) фонеми-трифона. Сукупність всіх акустичних образів усіх фонем-трифонів, властивих певній людині, складає усномовний файл диктора [2].

Розглянемо процес автоматичного озвучення тексту. Спочатку, проводиться розбиття тексту за схемою: абзац—речення—синтагма (інтонаційна група)—ритмогрупа (група наголосу)—фонетичне слово. З абзаців виділяються речення, кожне речення розбивається на синтагми, ті, в свою чергу, – на ритмогрупи. І вже всередині ритмогрупи визначаються фонетичні слова, які або збігаються з орфографічними словами або містять їх декілька, включно зі службовими словами. Потім з використанням фонетичних знань про усну мову орфографічний текст перетворюється на фонетичну та одночасно й на фонемно-трифонну транскрипції за універсальним правилом. Далі згідно розпізнаних типів синтагм будується інтонаційний контур, розраховуються тривалості поточних одноквазіперіодичних мікросегментів та фонем уцілому. При цьому враховується, що кожна синтагма містить лише один основний наголос, який відповідає ядерній ритмогрупі. Решта ритмогруп – початкова, перед'ядерні та післяядерні.

На підставі розрахунків приступаємо до власне формування (синтезу) усномовного сигналу. Для кожної фонеми-трифона з фонемно-трифонної транскрипції обирається акустичний образ (прототип) з бази даних. Цей прототип піддається перетворенням згідно розрахованої тривалості фонеми-трифона та інтонаційного контуру. В результаті цих перетворень маємо отримати визначену перед цим кількість квазіперіодів розрахованої довжини. Перетворені прототипи квазіперіодів і загалом фонем-трифонів об'єднуються, і отриманий в результаті сигнал подається на засоби озвучення.

Визначальною рисою кожної технології синтезу усномовного сигналу є алгоритм змінювання просодичних характеристик прототипів елементів компіляції. Основною метою при зміні просодики, тобто інтонації та темпу, є досягнення якомога кращої якості синтезу за найменших обчислювальних витрат на кожну дискрету синтезованого сигналу.

Як видно з аналізу відповідних алгоритмів [4], у часово-амплітудній області досягається і те й інше. Так, вельми прийнятна якість синтезу за технологією *PSOLA* досягається за досить скромних витрат на обчислення при зміні інтонаційних характеристик прототипу елемента компіляції – до 9 арифметичних дій на одну дискрету. У технології *MBROLA* обчислювальні витрати ще скромніші – 6 арифметичних дій на одну дискрету, і це при кращих показниках якості синтезу, ніж у *PSOLA*. Технології *Unit-Selection* взагалі не передбачають яких-небудь інтонаційних змін прототипу сигналу, хоч при цьому і виникають надмірні витрати пам'яті на зберігання всіх можливих інтонаційних проявів кожного елементу компіляції, що конкатенуються без жодних перетворень сигналу. Це в свою чергу позбавляє певної гнучкості саму систему синтезу усної мови.

Пропонується ще один спосіб компіляції одно-квазіперіодичних мікросегментів у амплітудно-часовій області. В основі цього методу закладено модель лінійного прогнозування сигналу, що дозволяє за певною кількістю попередніх відкліків сигналу спрогнозувати наступні з достатньо високою точністю апроксимації:

$$\tilde{f}_n = -\sum_{s=1}^{s=m} a_s f_{n-s} + \varepsilon_n , \qquad (1)$$

де $\tilde{f}_n$ — відліки прогнозованого сигналу, f_n — відліки спостережуваного сигналу, a_s, $s = 1{:}m$ — параметри передбачення, кількість яких m обирається в межах від 10 до 20, ε_n — похибка прогнозування. Параметри передбачення оцінюються на інтервалі аналізу рівному одному або двом квазіперіодам шляхом, наприклад, мінімізації суми квадратів похибки прогнозу.

Отже, нехай розрахуок тривалостей фонем та довжин квазіперіодів або мікросегментів і їх кількості в кожній фонемі вже виконано. Тоді відповідний прототип фонеми-трифону необхідно піддати темпоральним змінам, тобто привести до розрахованої довжини. Це досягається шляхом доведення кількості квазіперіодів до розрахованої. Таким чином, темпоральні зміни прототипу фонеми-трифону здійснюються шляхом викидання або повторення певних мікросегментів. У свою чергу, інтонаційні зміни сигналу здійснюються внаслідок скорочення або збільшення довжин відповідних квазіперіодів.

При темпоральних змінах, які вимагають подовження або скорочення тривалості прототипу фонеми-трифона до заданої довжини, обчислюється кількість мікрофонем (квазіперіодів), на яку їх необхідно збільшити або зменшити. Збільшення кількості мікрофонем відбувається шляхом повторення деяких квазіперіодів прототипу фонеми-трифону певну кількість разів. Зменшення кількості квазіперіодів здійснюється за рахунок викидання певних квазіперіодів.

Визначення квазіперіодів (мікрофонем), які повторюємо або викидаємо при темпоральних змінах, не є однозначним і потребує додаткових досліджень. Тому маємо право обрати одне з

простіших рішень, а саме повторюємо (викидаємо) переважно центральні квазіперіоди.

Як зазначалося, інтонація сигналу регулюється шляхом задання певної довжини квазіперіодів на вокалізованих ділянках синтезованого сигналу. Для того, щоб подовжити окремо взятий квазіперіод згідно моделі лінійного прогнозування, достатньо знати певну кількість попередніх відліків та значення коефіцієнтів прогнозування на відповідному інтервалі аналізу. Решту відліків, що доповнюють квазіперіод до потрібної довжини, обчислюємо за алгоритмом (1). Скорочення довжини квазіперіоду до заданої здійснюється шляхом видкидання відліків квазіперіоду понад задану довжину.

3 Комп'ютерні засоби дослідження та розроблення фонемно-трифонної моделі синтезу мовлення

Розроблено комп'ютеризовані засоби формування бази даних і знань, які використовуються як для синтезу, так і для пофонемного розпізнавання усної мови. З їх допомогою також проводяться експериментальні дослідження розпізнавання та синтезу.

При формуванні баз даних і знань, опрацюванні методів та алгоритмів, що стосуються усномовної інформатики, розроблені програмні засоби забезпечують виконання таких дій:

- введення фонетичних специфікацій, що включають опис базових фонем та перелік можливих інтонацій для обраної природної мови;
- накопичення навчальної вибірки за заданим текстом;
- автоматична сегментація навчальної вибірки на одноквазіперіодичні мікросегменти та на квазіперіодичні й неперіодичні сегменти;
- автоматизована сегментація навчальної вибірки на фонеми-трифони;
- дослідження та розроблення методів та алгоритмів синтезу усної мови.

Початкова інформація про природну мову складається з алфавіту базових фонем та переліку інтонацій. Задаємо її в режимі діалогу. Спочатку вводимо потрібну мову до списку підтримуваних *мов або активізуємо відповідну мову, якщо вона вже* є в реєстрі. В окремій секції задається алфавіт базових фонем та опис кожної з них. Передбачено підтримку паралельних алфавітів у кодуванні як кирилицею, так і латинськими літерами. Окреме поле відведене для задання символу, який розділяє фонеми. Нарешті, передбачено секцію, в якій описуються можливі інтонації та відповідні їм символи.

Таку початкову інформацію задаємо як для окремої мови, так і для групи мов одночасно у випадку, якщо в поставленій задачі є елементи багатомовності.

Навчальною вибіркою називаємо сукупність наговорених диктором звукових файлів разом з відповідними орфографічними або фонетичними *текстами, які є окремими словами або фразами.* Звукові файли з навчальної вибірки ще називатимемо реалізаціями слова або фрази.

Накопичення навчальної вибірки за заданим текстом виконується за допомогою вікна накопичення навчальних вибірок. На початку задається ім'я (ідентифікатор) диктора та вводиться текст, орфографічний або фонетичний, за яким проводиться запис навчальної вибірки. Передбачено можливість запису реалізацій слів або фраз в окремі файли для зручності при подальшому обробленні даних.

Запис навчальної вибірки відбувається за такою схемою:

1. Мишкою активізується текст реалізації, з якої розпочинаємо запис.
2. Диктор натискає кнопку запису *Record* і, утримуючи її, промовляє відповідний текст.
3. Після промовляння тексту диктор відпускає кнопку запису.
4. Автоматично активізується наступна реалізація, якщо список не вичерпано. Вручну можна вибрати будь-який інший елемент.
5. Для запису наступної реалізації переходимо до кроку 2. Якщо записуваний текст вичерпано, завершуємо роботу.

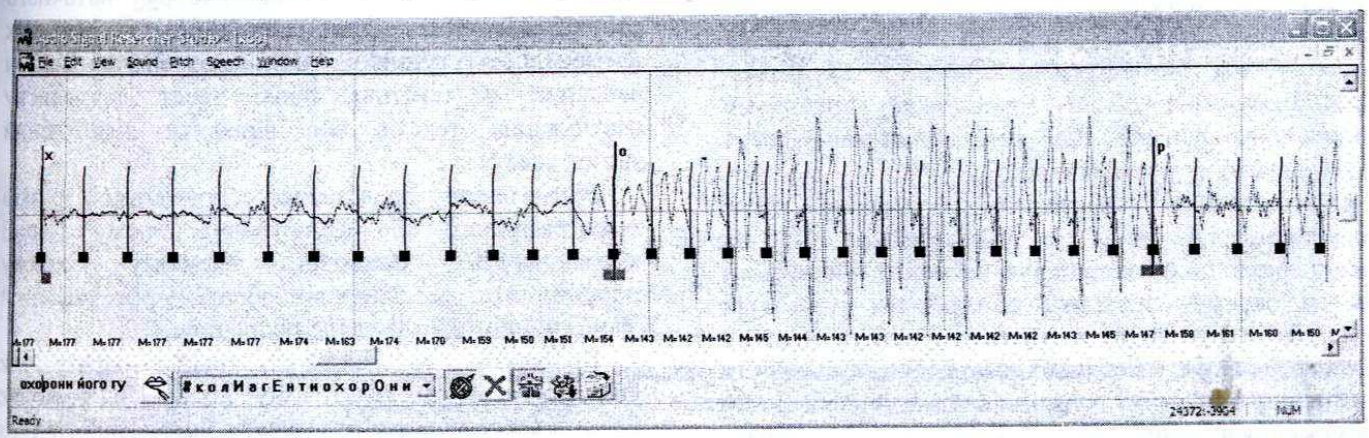


Рис. 1. Приклад зображення ділянки відсегментованої реалізації навчальної вибірки з чоловічого голосу. Сигнал розглядається крізь одноквазіперіодичну сітку, "натягнуту" як на періодичні сегмети, що відповідають фонемам-сонорантам, так і на неперіодичні сегменти, що відповідають глухим та паузі. Цифри знизу сигналу позначають тривалість квазіперіода в дискретах.

Записаний усномовний сигнал прослуховує експерт з метою виявлення бракованих або невиразних реалізацій, які позначаються як "погані" (*bad*). Остаточно, "погані" реалізації перезаписуються за вищенаведеною схемою з тією лише різницею, що автоматично активізуватиметься наступна відбракована реалізація.

В результаті процедури запису навчальної вибірки сформуються файли, що відповідають реалізаціям навчальної вибірки, яким співставлено ім'я диктора та текст, що було вимовлено. Далі проводимо сегментацію реалізацій навчальної вибірки на фонеми-трифони, а вибірки окремих фонем-трифонів – на одноквазіперіодичні мікросегменти та квазіперіодичні й неперіодичні сегменти. Причому, порядок сегментування може бути і зворотній.

На Рис. 1 показано приклад результату сегментування реалізації з навчальної вибірки. Ділянці усномовного сигналу співставлено відповідні фонеми-трифони, та відображено границі одноквазіперіодичних мікросегментів.

Сегментація навчальної вибірки на фонеми-трифони проводиться за участю експерта. Програмне забезпечення автоматично розставляє границі фонем лише дуже приблизно. Далі експерт, керуючись звуковою та візуальною підказкою уточнює межі фонем.

Пересування границь фонем здійснюється у двох режимах. Перший режим передбачає незалежне пересування границь сусідніх фонем, тобто можливі ділянки сигналу, на яких фонеми "накладаються" одна на одну, або ділянки, які не належать жодній фонемі. Коли активований другий режим, шляхом переведення кнопки *Linked Marks* у натиснутий стан, границі сусідніх фонем пересуваються узгоджено: початок поточної фонеми збігається із закінченням попередньої. В обох випадках за наявності одноквазіперіодичної розмітки границі фонем автоматично синхронізуються з границями квазіперіодів, тобто відбувається автоматичне прив'язування границь фонем-трифонів до одноквазіперіодичної сітки.

При уточнюванні границь фонем експерт використовує режим візуалізації сигналу, в якому зображується авторегресійний спектр, та режим прослуховування (*Sound->Sound Phone*), при якому озвучується фонема, на яку користувач-експерт вказує лівою кнопкою мишки.

Сегментація навчальної вибірки на одноквазіперіодичні мікросегменти та квазіперіодичні й неперіодичні сегменти виконується як на окремих звукових файлах, так і на всій навчальній вибірці в автоматичному режимі. Експерту лише необхідно вказати найбільшу та найменшу допустимі довжини квазіперіодів, а також обмеження на приріст тривалості одноквазіперіодів.

Границі квазіперіодів у вікні дослідження звукового сигналу зображаються у вигляді вертикальних рисок з кольоровими квадратиками-ручками в нижній частині риски. Ручки призначені для пересування границь квазіперіодів (мікрофонем).

Як тільки проведено сегментацію навчальної вибірки, інформація про фонеми-трифони та їх структуру вноситься до реєстру фонем-трифонів, формуючи таким чином усномовний файл диктора. Для цього слід натиснути кнопку *Update Collection*, попередньо виділивши сигнал тих фонем-трифонів, які заносяться до реєстру. Якщо не виділено жоної фонеми-трифона, до бази даних заносяться всі фонеми-трифони з реалізації.

Знання про індивідуальні інтонаційні контури отримуються шляхом оброблення відсегментованих реалізацій фраз, в яких представлені всі типи інтонації.

Сформований усномовний файл диктора зберігається у вигляді файлу, який містить не сам сигнал, а лише посилання на фрагменти звукового файлу, що відповідають фонемам-трифонам. Таким чином, зберігається зв'язок з навчальною вибіркою, і зміни сегментації, проведені на навчальній вибірці, автоматично відображаються в базу даних шляхом виклику команди *Update Collection*.

4 Дослідження синтезованого сигналу

В рамках Студії дослідника усномовного сигналу розроблено стенди дослідження синтезу та розпізнавання усного мовлення [5]. З використанням першого стенду виконуються експериментальні дослідження розбірливості та якості звучання синтезованого усномовного сигналу. На другому стенді проводяться дослідження пофонемного розпізнавання.

Стенд дослідження усномовного синтезу являє собою вікно, представлене діалоговою панеллю з багатьма елементами регулювання та командними елементами, які логічно розбиті на частини, як це показано на Рис. 2.

Поле задання орфографічного або фонетичного тексту, що буде подано на озвучення, дозволяє маніпулювати зі зразками тексту або транскрипції, озвучений сигнал яких планується дослідити. Це поле складається з вікна вибору поточного озвучуваного тексту, текстового вікна та командних елементів – ґудзиків, що спонукають додати введений у текстове вікно текст до списку озвучуваних текстів або видалити непотрібний зразок тексту.

Вікно вибору представляє собою список зразків орфографічного або фонетичного тексту. Перші кільканадцять символів елементу списку допомагають дослідникові обрати орфографічний /фонетичний текст із уведених раніше.

Вибраний зі списку текст одразу ж відображається у текстовому вікні. Це дозволяє бачити текст цілком, редагувати його. Щоб додати новий зразок тексту, орфографічного або фонетичного, слід увести зразок тексту до текстового вікна або відредагувати текст попередньо доданого зразка і виконати команду додання нового зразка

тексту шляхом активації відповідного командного гудзика. Також передбачена можливість видалити зразок тексту.

Щойно додано новий зразок тексту або вибрано з уведених раніше зразків, поле виділення фонем-трифонів відображає фонемно-трифонні транскрипції, що відповідають поточному зразкові тексту. У поточній фонемно-трифонній транскрипції передбачено можливість виділяти окремі фонеми-трифони, які з метою досліджень можна замінювати на інші та/або задавати їм різні просодичні характеристики за допомогою описаних далі полів.

Поле вибору фонем-трифонів дозволяє "підсаджувати" різні фонеми-трифони з індивідуального усномовного файлу диктора у фонемно-трифонну послідовність озвучуваного тексту.

Основний елемент поля вибору фонем-трифонів є таблиця фонем-трифонів, доступних у індивідуальному усномовному файлі диктора, що містить база даних і знань озвучення текстів. Таблиця показує лише фонеми-трифони-претенденти, тобто ті фонеми-трифони, чия термінальна фонема збігається з термінальною фонемою виділеної фонеми-трифона у полі фонемно-трифонної транскрипції.

У таблиці коротко характеризуються фонеми-трифони-претенденти, показано їх оригінальне фонемне оточення. Щоб замінити виділену фонему-трифон у поточній фонемно-трифонній послідовності озвучуваного тексту, необхідно вибрати фонему-трифон з таблиці та викликати команду *Assign*.

Поле регулювання просодичних характеристик сигналу дозволяє змінювати інтонацію, темп і гучність синтезованого сигналу. Це поле містить список всіх одноквазіперіодичних сегментів (мікросегментів) виділеної фонеми-трифона поточної фонемно-трифонної транскрипції. У сусідніх текстових віконечках відображаються значення довжини вибраного одноквазіперіоду та його гучності. Передбачена можливість розмножувати окремі квазіперіоди.

Шляхом зміни довжини квазіперіодів фонеми-трифона в допустимих межах відбувається зміна інтонаційного контуру фонеми на низькому рівні. Викидання окремих мікросегментів або їх множення призводить відповідно до скорочення або подовження фонеми-тифона вцілому, а отже до зміни його темпоральних характеристик.

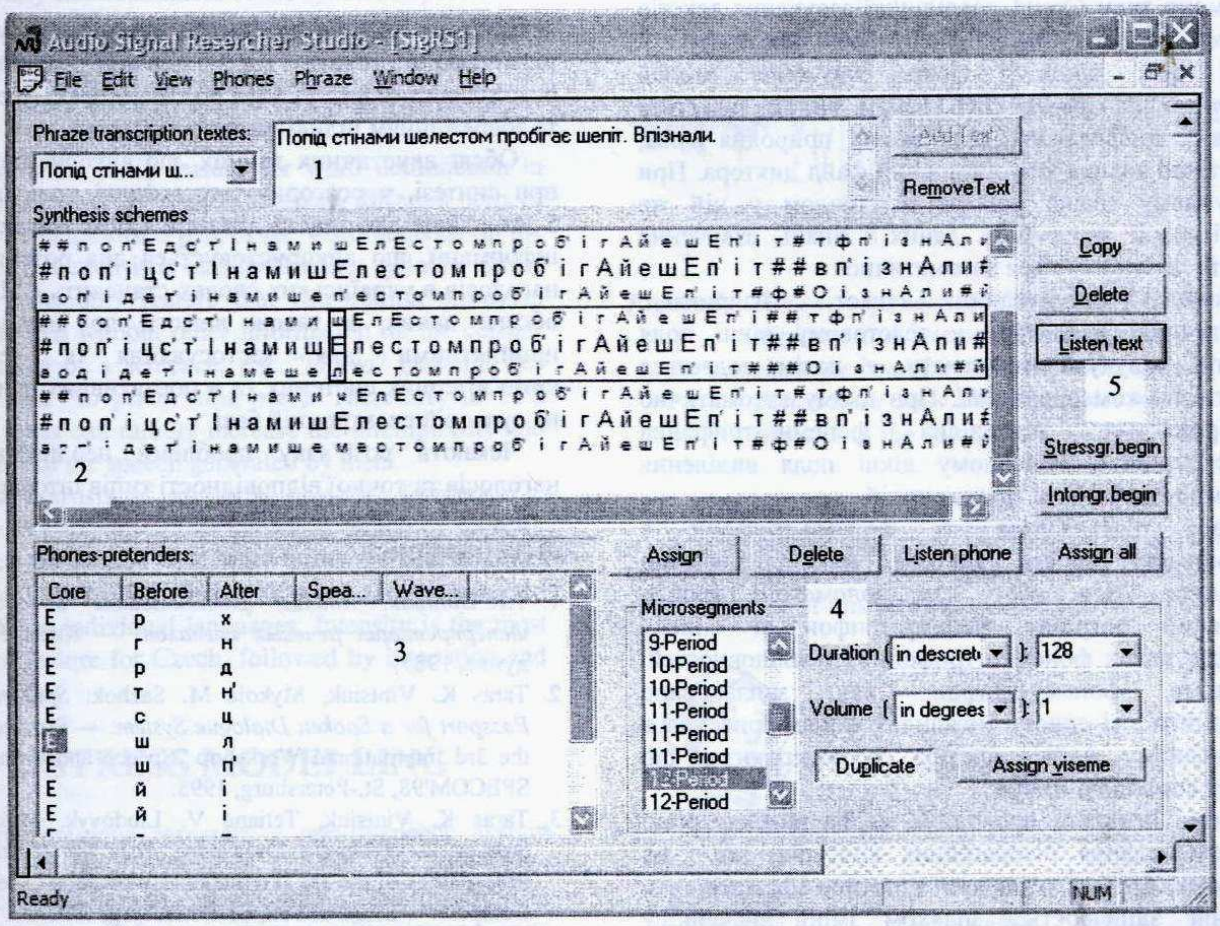


Рис. 2. Вікно дослідження усномовного синтезу представляє собою діалогову панель з елементами регулювання та командними елементами, які підрозділяються на такі розділи:

1) секція задання озвучуваного тексту, орфографічного або фонетичного;

2) інтерактивний список фонемно-трифонних транскрипцій заданого орфографічного (фонетичного) тексту;

3) секція вибору фонем-трифонів з бази даних і знань;

4) секція регулювання просодичних характеристик сигналу;

5) загальні командні елементи.

Передбачена також можливість зміни інтонаційних харктеристик на вищому рівні шляхом задання інтонаційних контурів синтезованого тексту. Відповідна команда у меню (*Phraze->Apply intonation*) викликає діалогове вікно, в якому за спеціальною схемою описуються інтонаційні контури ритмогрупи або синтагматичні. Після введення опису інтонаційних контурів і підтвердження операції, просодичні параметри керування синтезатором можуть бути обчислені за одним з уведених контурів.

Загальні командні елементи містяться як на стенді, окремо від описаних вище секцій, так і в меню.

Після заповнення належним чином полів стенду досліджень усномовного сигналу подається загальна команда озвучення тексту. На самому початку роботи зі стендом необхідно завантажити базу даних і знань для озвучення українських текстів загальною командою з меню. Одночасно можливо працювати як з декількома базами даних, так і з різними стендами дослідження синтезу усної мови.

Загальний порядок роботи зі стендом дослідника усномовного синтезу виглядає наступним чином. Щоб започаткувати нові дослідження, у Студії дослідника усномовного сигналу створюємо новий документ типу Стенд дослідника озвучення текстів (*Speech Synthesis Master*). Далі необхідно викликати файл бази даних і знань озвучення текстів (*Phones->Add speech DB*). Таким чином задається основна конфігурація досліджень: природня мова, фонетичні знання та усномовний файл диктора. При наступному сеансі роботи зі стендом у цій же конфігурації файл бази даних і знань озвучення текстів завантажується автоматично.

Новий озвучуваний текст, наприклад, фонетичний, вводиться у текстовому вікні поля задання озвучуваного тексту, а потім додається супутньою командою *Add*. При цьому автоматично відображається відповідна фонемно-трифонна транскрипція у головному вікні поля виділення фонемно-трифонних транскрипцій.

Далі підсаджуємо ті фонеми-трифони з усномовного файлу диктора, які заплановано дослідити. Для цього за допомогою мишки виділяємо потрібну фонему-трифон зі списку досліджуваних фонемно-трифонних послідовностей, обираємо фонему-трифон, яку заплановано "підсадити", зі списку можливих фонем-трифонів в усномовному файлі диктора, і закріплюємо цю заміну командою *Assign*.

Зміни інтонації проводимо як на вищому рівні шляхом задання інтонаційних контурів, так і на нижчому, вручну змінюючи довжини квазіперіодів. Останнє здійснюється шляхом зміни значень у віконечку довжини виділеного квазіперіоду. Щойно точка введення перейде до іншого елемента стенду (мишкою виділяємо наступний квазіперіод), змінені значення довжини квазіперіоду запам'ятаються і будуть використані при наступній процедурі синтезу сигналу.

Темпоральні зміни прототипу на низькому рівні здійснюються за допомогою маніпуляцій з мікросегментами фонем-трифонів. Множення виділених мікросегментів за допомогою команди *Duplicate* призводить до подовження окремої фонеми-трифона, а отже і до загального сповільнення темпу. Викидання мікросегментів, тобто задання їм нульової довжини, призводить до скорочення довжини фонеми-трифона, а тому – до прискорення темпу.

Нарешті, командою *Listen text* запускаємо процедуру синтезу, яка моделює оригінальний алгоритм синтезу усномовного сигналу в часово-амплітудній області. При цьому відкривається новий документ типу *Усномовний сигнал*, куди автоматично вставляється синтезований сигнал, який тепер є доступним для прослуховування, аналізу та подальшого дослідження (Рис. 1).

При прослуховуванні експертами синтезованих як окремих слів, так і злитого мовлення, було з'ясовано словесну розбірливість синтезованого сигналу, що виявилася на рівні не менше 90% на перших сенсах, зростаючи за мірою "звикання" до синтезованого голосу.

5 Висновки

Запропонований метод синтезу дозволяє озвучувати українські тексти з доволі прийнятною розбірливістю, натуральністю синтезованого сигналу та збереженням індивідуальності мовця.

Обсяг акустичних данних, що використовуються при синтезі, у розгорнутому вигляді становить від 5 МБ і вище для одного диктора. Обсяг лінгвістичної інформації, що використовується для розставлення наголосів в українських словах становить 4 МБ. Такі обсяги даних, а також вимоги до швидкодії є прийнятними для застосування в реальних комп'ютерних системах та в портативних пристроях на сучасній електронній базі.

Чекають розв'язку проблеми неоднозначності наголосів та точної відповідності типів інтонації.

Література

1. Т.К. Винцюк. *Анализ, распознавание и смысловая интерпретация речевых сигналов.* — Киев: Наукова думка, 1987.
2. Taras K. Vintsiuk, Mykola M. Sazhok: *Speaker Voice Passport for a Spoken Dialogue System.* — Proceedings of the 3rd International Workshop "Speech and Computer" - SPECOM'98, St.-Petersburg, 1998.
3. Taras K. Vintsiuk, Tetiana V. Liudovyk, Mykola M. Sazhok. — *Phonetic Knowledge Base for Ukrainian.*— Proceedings of the 3rd International Workshop "Speech and Computer" - SPECOM'98, St.-Petersburg, 1998.
4. T. Dutoit, H. Leich. — A Comparison of Four Candidate Algorithms in the Context of High Quality Text to Speech Synthesis. — ICASSP'94.
5. Микола Сажок. *Комп'ютерні засоби експериментальних досліджень усномовного сигналу.* — Праці 4-ї Всеукраїнської міжнародної конференції "УкрОбраз-98", Київ, 1998.

Modelling Word Stress for Use in Speech Synthesis

Daniel Tihelka, Jindřich Matoušek, Martin Vlach

University of West Bohemia in Pilsen, Department of Cybernetics,
Univerzitní 22, 306 14 Pilsen, Czech Republic
dtihelka@kky.zcu.cz, jmatouse@kky.zcu.cz

ABSTRACT

This paper deals with word stress modelling for use in text-to-speech systems. It describes one of the ways of stress modelling. Several experiments were carried out for different phonemes containing stress and other prosodic features. Listening tests with specially designed words were set up for the comparison of speech generated from each experiment. The results show that the examined way of stress modelling can increase intelligibility and naturalness of the synthetic speech.

1. INTRODUCTION

Stress is the basic feature for word delimitation in continuous spoken speech, emphasising different parts of words. The emphasised part of the word depends on the language; in Czech it is usually the first syllable. Stress is mostly the only feature that can delimit words in spoken speech without other information on the context and meaning of a sentence, e.g. "*topivo*" (material for burning) and "*to pivo*" (this beer). Therefore, in TTS systems stress can rapidly increase the intelligibility and naturalness of the speech generated by them.

Stress consists of all three prosodic features, i.e. *duration* of phonemes (also known as tempo of speech), *intonation* (frequency F_0 of voice base tone) and *intensity* (loudness). Each of these features is emphasised differently in individual languages. Intensity is the most important feature for Czech, followed by intonation and duration [3].

2. STRESS MODELLING

In our previous research, a text-to-speech system ARTIC (ARtificial Talker in Czech) was built. This system is based on unit concatenation in the time domain [2] and was used for experiments with stressed units presented in this paper.

A new speech corpus recorded by a female speaker, comprising several hours of speech, was used for building the speech segment database (SSD) for this system. A fully automatic process based on hidden Markov models (HMM) was used in a segmentation process, and the speech corpus was segmented into *fenemes* (each feneme corresponds to one state of HMM; each HMM is used for the modelling of a triphone, see Figure 1) [2].

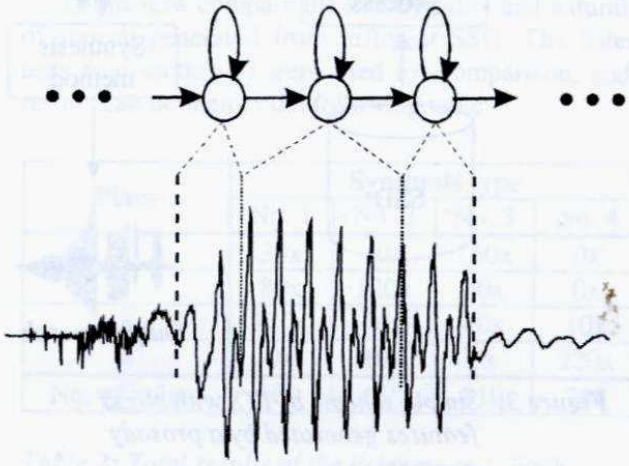


Figure 1: *The correspondence of three-state HMM, a triphone and fenemes. The dashed lines show the boundary of a triphone, the dotted lines show the boundaries of the fenemes of the vowel "a".*

Representative segments were stored in the speech segment database after the segmentation process. A very simple scheme of this process can bee seen in Figure 2.

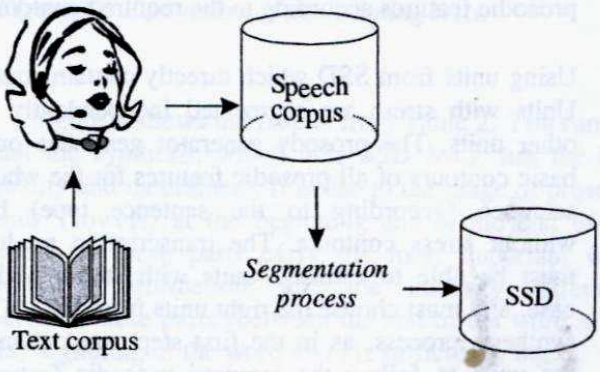


Figure 2: *Simply scheme of segmentation process*

ARTIC can work with prosodic features. All these are represented as a contour of changes from the base value, and the concatenation method can modify concatenated units according to contour requirements.

There are two basic ways of modelling the stress features in TTS systems:

- Using the contours from prosodic features for stress modelling, especially the contour of intensity and intonation for Czech speech. For a simple scheme of this system see Figure 3. In this case, the prosody

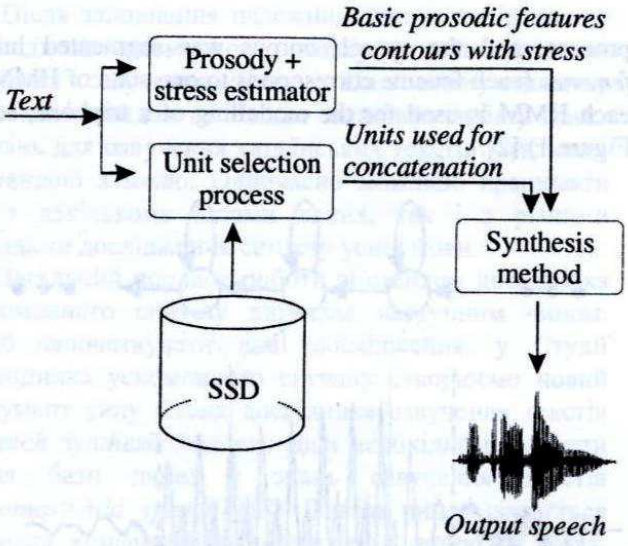


Figure 3: *Simple scheme of TTS with stress features generated by a prosody generator.*

generator estimates the basic contours of all prosodic features for the whole sentence (these features depend on the sentence type). The stress features of all stressed phonemes must be added to the basic contours. This means that the generator must be able to estimate phonemes with stress. The transcription module provides transcription to the concatenated units (fenemes in our case) stored in SSD, which are independent of stress. The synthesis module joins units together and changes their prosodic features according to the required contours.

- Using units from SSD which directly contain stress. Units with stress are segmented independently of other units. The prosody generator generates only basic contours of all prosodic features for the whole sentence (according to the sentence type) but without stress contours. The transcription module must be able to estimate units with stress in this case, and must choose the right units from SSD. The synthesis process, as in the first step, must change the units to follow the required prosodic features contours. A simple system scheme can be seen in Figure 4. In our research we focused on the second approach, as described below.

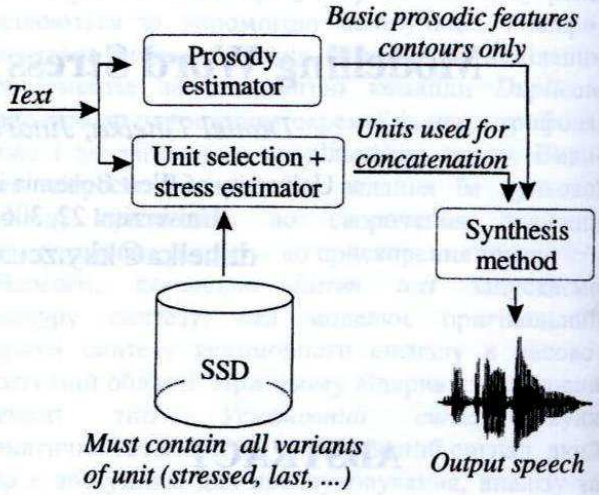


Figure 4: *Simple scheme of TTS using units with stress features*

3. PHONEMES USED AS PROSODIC UNITS

The statistical approach based on hidden Markov models was used, as mentioned above. The following text describes the positions and types of phonemes chosen for prosody modelling:

1. At first, the vowel in the first syllable of each word with one or more syllables was modelled as *stressed*, which means that this unit contains stress. All other vowels and consonants were modelled as unstressed, because, as we mentioned, stress lies on the first syllable in Czech, and intensity change is usually bigger in a vowel than in a consonant. It can be expected that this simple approach will increase the intelligibility of speech without a large increase in the size of SSD.

2. The vowel in the first syllable (as in the previous case) was modelled as stressed, and the vowel in the last syllable in the words with two and more syllables was modelled as *final*. This means that it lies at the end of the word, and it is possible that the next syllable will be stressed. All the other units were modelled as unstressed. A bigger intensity and intonation contrast between the stressed syllable and the final syllable can be expected, increasing the delimiting feature of stress in Czech. Generated speech could be more intelligible and the size of SSD could still be reasonable.

3. All vowels and consonants in the first syllable were modelled as stressed and all vowels and consonants in the last syllable were modelled as final. All other triphones (i.e. triphones in the middle syllables) were modelled as unstressed. Syllables are modelled

with a greater precision in this approach, as every triphone of the word contains information about its position in the word. We expected that this approach would bring the highest intelligibility and naturalness. Using this approach, the size of SSD will be the largest.

Three-state left-to-right models were used for all described units. These models had the same initialisation of the state mean vectors, covariance matrices and transition matrices before the training process.

3. THE METHOD FOR RESULTS COMPARISON

We set up small listening tests for the comparison of results obtained with different prosodic units. These tests were designed especially for Czech phenomena on words boundaries.

Here are some examples of word pairs which are differentiated from each other by stress positions. The individual words were used in the sentences for the listening tests:

topivo	(fuel)	*to pivo*	(that beer)
tabulka	(a chart)	*ta bulka*	(that roll)
prsten	(a ring)	*prs ten*	(that breast)
jak oběžné	(as circular)	*jako běžné*	(as usual)

Etc.

There were 15 sentences containing one of these words. Each of these 15 sentences was synthesised from:

- SSD without stressed units at first (synthesis output No.1 in the following tables).
- SSD containing units for vowels in the first syllable of the word (synthesis output No.2).
- SSD with units for vowels in the first and in the last syllables of the word (synthesis output No.3).
- SSD with units for all phonemes in the first and in the last syllables of the word (synthesis output No.4).

Naturally, for each synthesis output mentioned, ARTIC had to be able to select the right units for each used SSD.

23 people took part in the listening tests (14 female and 9 male listeners). Every listener heard 10 sentences in all their versions (4 versions per sentence) and had to select the sequence of types of each sentence from the best to the worst. The best type was evaluated by 4 points, the worst by 1 point.

4. RESULTS

Let us compare the sizes of the Speech Segment Database first.

As can bee seen in Table 1, the size of SSD was very similar for syntheses No. 1 and 2, as synthesis No.2 has

SSD used	Number of units in SSD	Size of SSD [MB]
No. 1	9097	17.501
No. 2	10108	20.808
No. 3	10635	31.165
No. 4	11877	47.566

***Table.1** SSD sizes for different prosodic units and number of units in the SSD.*

only 1011 stressed units (triphones, corresponding to vowels at the beginning of the word). The size of SSD for synthesis No.3 is still not very big, as 527 units were added for modelling vowels at the end of the word. On the other hand, the size of SSD used for synthesis No.4 is a little bigger.

Let us now compare the intelligibility and naturalness of speech generated from different SSD. The listening tests (see section 3) were used for comparison, and the results can be seen in the following table.

Place	Synthesis type			
	No. 1	No. 2	No. 3	No. 4
1	30x	40x	160x	0x
2	80x	120x	30x	0x
3	120x	60x	40x	10x
4	0x	10x	0x	220x
No. of points	600	650	810	240

***Table 2:** Total results of the listening test. Each number represents the number of occurrences at a given place.*

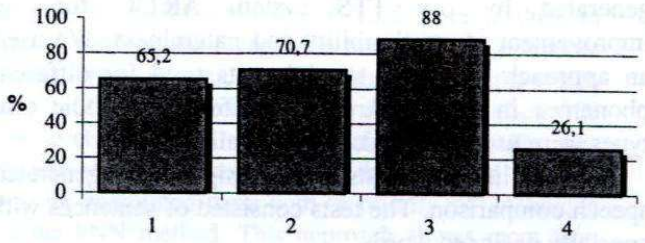


***Figure 5:** Evaluation of the listening tests.*

Figure 5 shows the results from Table 2. You can see that the synthesis, which uses SSD No.3, has the best quality and naturalness. It is due to the usage of prosodic units (vowels) at the beginning and at the end of the word, as these parts carry the most important word prosodic information, and the intensity difference between these parts (between the end of the word *n* and the beginning of the word *n+1*) is sufficiently big.

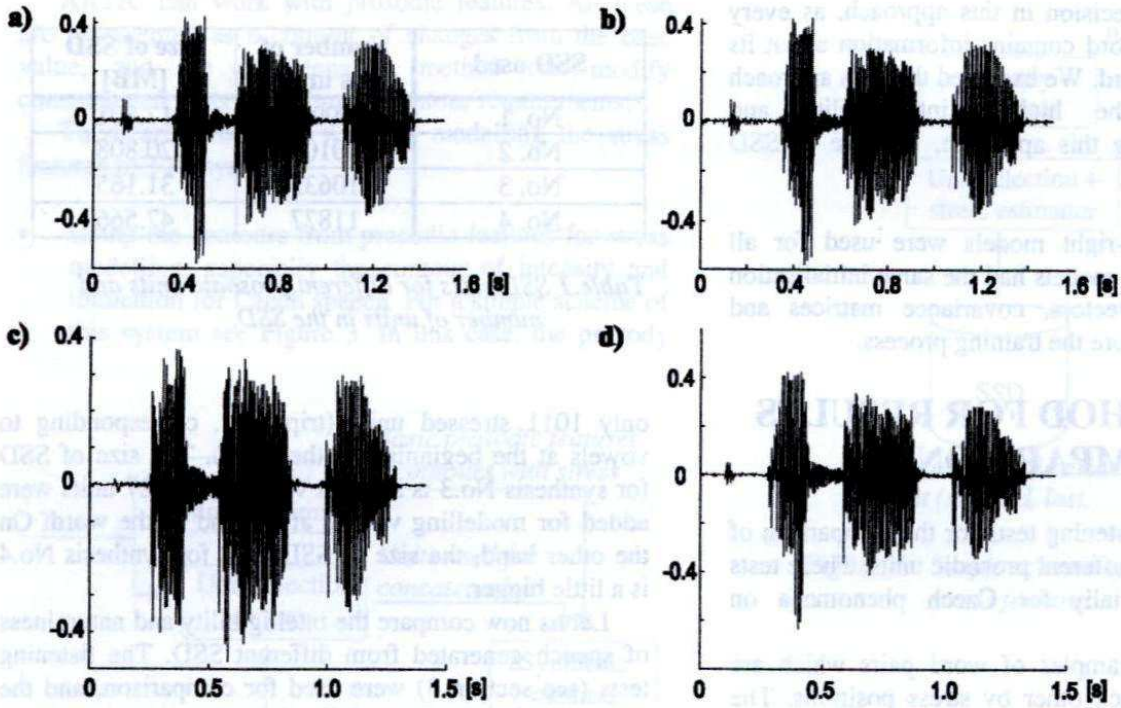


Figure 6: *Speech signals for the words "tabulka" (a, c) and "ta bulka" (b, d) from SSD without prosodic units (a, b) and with stressed and final vowels as prosodic units (c, d).*

Using SSD No.4 for speech generation, the best results were expected. However, this speech turned out to be the worst. It is due to the small number of units used for training the HMMs, as every unit from SSD No.1 had 3 variants in the SSD No.4.

7. CONCLUSION

We tried to add the stress of the words to the speech generated by our TTS system ARTIC for the improvement of intelligibility and naturalness. We used an approach based on special units used for different phonemes in the word. Three different prosodic unit types were used: stressed, final and all others.

Special listening tests were designed for generated speech comparison. The tests consisted of sentences with specially designed words.

The listening tests showed that this approach brought higher intelligibility and naturalness, especially when vowels in stressed and final syllables of the word were modelled independently of other phonemes.

Stress modelling directly by a prosody generator is planned as our future work.

AKNOWLEDGEMENT

This research has been supported by the Grant Agency of the Czech Republic No. 102/02/P134 and by the Ministry of Education of the Czech Republic No. MSM 235200004.

REFERENCES

1. Matoušek J., Psutka J. and Krůta J.: "*Design of Speech Corpus for Text-to-Speech Synthesis.*" -In: Proceedings of the EUROSPEECH 2001, vol. 3. Aalborg, Denmark, 2001, pp. 2047-2050.

2. Matoušek J. and Psutka J.: "*ARTIC: A New Czech Text-to-Speech System Using Statistical Approach to Speech Segment Database Construction.*" -In: The Proceedings of ICSLP2000, vol. IV. Beijing, China, 2000, pp. 612-615.

3. Palková Z.: "*Fonetika a fonologie češtiny (Phonetics and phonology of Czech).*", Karolinum, Prague 1994.

4. Donovan R.E., Woodland P.C.: "*A Hidden Markov-Model-Based Trainable Speech Synthesiser.*", Computer Speech and Language, 1999.

Automatic phonetic baseforms for person names in the Czech dialogue system

Pavel Brodský, Luděk Müller
University of West Bohemia in Pilsen, Department of Cybernetics,
Univerzitní 22, 306 14 Pilsen, Czech Republic
pbrodsky@kky.zcu.cz, muller@kky.zcu.cz

ABSTRACT

Phonetic baseforms are the basic recognition units in most speech recognition systems. These baseforms are usually determined by linguists once a vocabulary is chosen and not modified thereafter. However, several applications, such as name dialling require the user to be able to add new words to the vocabulary. These new words are often names or task-specific jargons that have user dependent pronunciation.

In this paper, we describe a method how to generate a phonetic baseform from acoustic pronunciation of a name without a prior knowledge of the name spelling. We used a language model based on bigram statistics.

1. INTRODUCTION

There has been considerable interest in *telecommunications and embedded speech recognition* application that provide personalized vocabularies. Name dialling is one such example of a telephonic application where it is necessary to have ability to provide speaker dependent vocabularies for repertory dialling. This feature enables the user to add even such words to the personalized vocabulary for which a spelling or acoustic representation does not exist in the speech recognition lexicon, and associate these words to a phone number to be dialled. We will show how speaker dependent baseforms could be derived from one or two speech utterances by using speaker independent acoustic model and a language model. We use the bigram probabilities to constrain the transition between phonemes.

The structure of this paper is as follows. In Section 2 we present our recognition system and its components: Speech recognition engine, acoustic modelling, front-end, labeller and decoder. In Section 3 we describe baseform generating algorithm. The mumble model is described in Section 4. In Section 5 language model for names and surnames of Czech Republic inhabitants is described. Experimental results are contained in Section 6 and Section 7 is conclusion.

2. SYSTEM OVERVIEW

The speech recognition engine is based on a statistical approach. It comprises a front-end, an acoustic model, a language model and a decoding block [1].

Acoustic Modelling: As a basic speech unit of the recognition system a triphone is used. Each triphone is represented by a 3-state left-to-right HMM with a continuous output probability density function assigned to each state. Each density is expressed as a mixture of multivariate Gaussians with a diagonal covariance matrix. The Czech phonetic decision trees were used to tie states of Czech triphones.

Front-end: The speech signal is digitized through a telephone board at 8 kHz sample rate and converted to the mu-law 8-bit resolution format. The parametrization process used in our system is as follows: Firstly the pre-emphasized acoustic waveform is segmented into 25 millisecond frames every 10ms. A Hamming window is applied to each frame and 13 MFCCs (including the energy coefficient c_0) are computed. The first-order and second-order derivates of MFCCs are computed and appended to the static MFCCs each speech frame.

Labeller: The recognition algorithm uses 2510 different tie states, each of which represented by a mixture of 8 Gaussian distributions in the 39-dimensional space. Thus during a decoding it is necessary to compute a large number of log-likelihood scores (LLSs) every 10ms. In order to perform the recognition in real time the number of calculations is reduced by applying a technique which seeks to find and precisely determinate only first 150 most probable LLSs. This technique efficiently uses relevant statistical properties of the Gaussian mixture densities combining them with "a priority hit" technique and the kNN method. This approach allows more than 90% reduction of a computation cost without substantial decrease recognition accuracy.

Decoder: The decoder uses a crossword context dependent HMM state network generated by a Net generator. The input of the Net generator is a text grammar format represented by an extended BNF with respect of JSGF. The whole net consists of one or more in run-time connected regular grammars. A considerable part of the net is usually generated before the decoder starts but every part of the net can be generated on

demand in run-time. The decoder utilizes a Viterbi search with a beam pruning.

3. PHONETIC BASEFORMS

Phonetic baseforms are the basic recognition units in most speech recognition systems. These baseforms are usually determined by linguists once a vocabulary is chosen and not modified thereafter. However, several applications, such as name dialling, require the user be able to add new words to the vocabulary. These new words are often names or task-specific jargons that have user dependent pronunciations.

The phonetic baseform is sequence of phonemes which represents a given utterance (name). We can create it manually (by phonetic transcription rules) or automatically.

Example of baseform (for the Czech phonetic alphabet):

Name:	Luděk Müller
Manually:	sil l u dj e k sil m i l e r sil
Automatic:	sil r i d e p t sil m e l aa r sil

Some utterances can be pronounced by several ways depending on speaker, speaking style, and other conditions. Therefore, generally more than one baseform can be constructed and stored for an utterance (e.g. a person name). In this work we used only one phonetic baseform for each person name.

Each baseform can be easily added manually. In the case of automatic baseform generation the user for example can be required to say the given person names twice and for each utterance a baseform should be automatically generated.

The baseform generation algorithm is based on Viterbi algorithm [3] that searches the best phoneme path through the HMM net consist of all Czech phonemes and a set of phoneme transitions. The net structure is dependent on the language (phoneme) model and can be interpreted also as so-called mumble model described in more detail in the Section 4.

Also N-best hypotheses instead the first best hypothesis can be considered and in this case the decoder should produce a list of the most probable phoneme sequences. The recognition results can be stored as well in a phoneme graph which is an analogy to the word graph in the case of N-best word sequences decoding problem.

4. MUMBLE MODEL

The mumble model is constructed as a set of HMM [4] models connected in a parallel fashion. Each HMM model is 3-state left-to-right and represents one context-independent phone. The structure of the mumble model is depicted in Figure 1. Actually the probabilities of emission of an observation vector in a given state are evaluated as the maximal emission probability of all corresponding states of context-dependent triphones. Thus neither additional HMM models nor additional training is required. The value of the backward loop probability *BPr* causes a various length of the phone sequence recognized by the network in Figure 1. While the higher value produces more insertions, the small value induces more deletions in the resulting phone sequence.

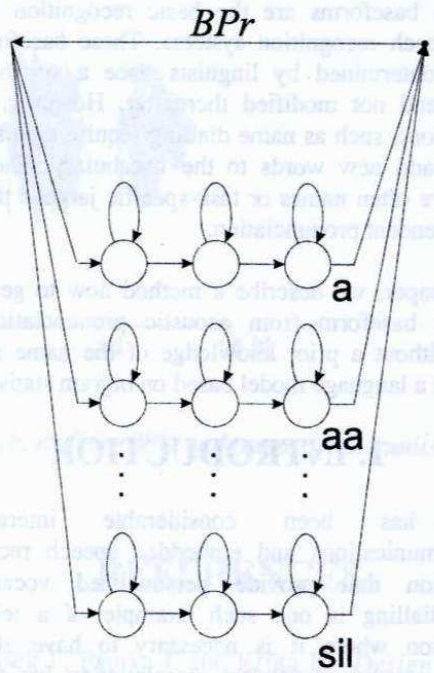


Figure 1.

In Figure 2. is mumble model with language model basis on bigrams with full matrix of transition probabilities. Language model is described in Section 5.

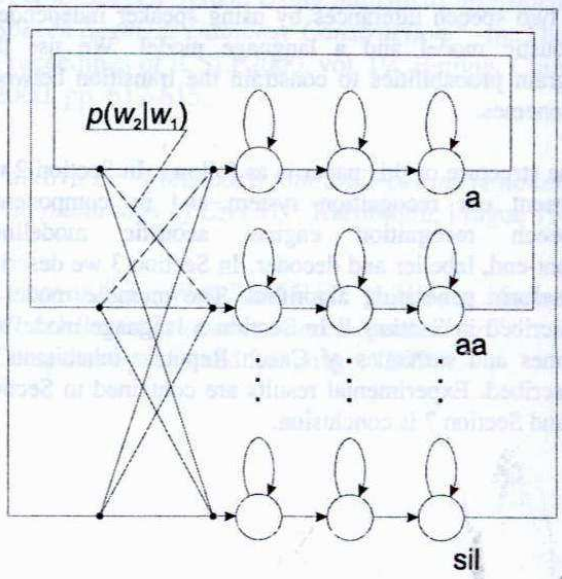


Figure 2.

5. LANGUAGE MODEL

The language model generally restricts variety of words sequences W and consequently also phone sequences. This restriction may be either deterministic (i. e. some words or phones sequence are not allowable) or "softer" stochastic (some word sequence are less probable). In our case we chose a probabilistic approach and an absolute discounting language model with backing-off for conditional probabilities method.

The basic idea of the absolute discounting language model with backing-off for conditional probabilities is to keep a high number of joined events (h, w), a word history h and a word w, almost unmodified. We suppose that the number of occurrences of joined events in training text will not change probably too much, if we select another training text of the similar size (from the same problem area). To consider possible variability of the number $N(h, w)$ of occurrences (h, w) in text we introduce parameter of permanent deviation b_h, so-called absolute discounting parameter that decreases the number of seen events $N(h, w)$. Furthermore we suppose that b_h is not dependent directly on the value $N(h, w)$, nevertheless it is dependent on the history h. This deviation must remain negative because unseen events in the text requires nonzero (thus positive) probabilities. By means of absolute discounting parameter b_h the part of probability mass is redistributed from the seen events to unseen events. The resulted formulae for the absolute discounting language model is:

$$\bar{p}(w|h) = \begin{cases} \dfrac{N(h,w) - b_h}{N(h,.)}, & \text{for } N(h,w) > 0 \\ b_h \dfrac{V - n_0(h,.)}{N(h,.)} \cdot \dfrac{\beta(w|\bar{h})}{\sum\limits_{w':N(h,w')=0} \beta(w'|\bar{h})}, & \text{for } N(h,w) = 0 \end{cases} \tag{1}$$

where $\beta(w|\bar{h})$ is a conditional probability of observing the word w given the generalized history $\bar{h}$. The generalized (also reduced) history $\bar{h}$ is defined as:

If (h, w) n-gram $(w_1, w_2, ..., w_n)$, then $(\bar{h}, w)$ is $(n\text{-}1)$-gram $(w_2, ..., w_n)$.

Our language model was created for names and surnames occurring in Czech Republic. We had at disposal 4 137 different names and 236 769 different surnames. In total we had 10 282 470 names and 10 296 459 surnames. As a basic unit of the language model we chose a phoneme which means that before computing individual probabilities we had to perform a phonetic transcription. Foreign names and surnames that are not subjected to Czech phonetic transcriptions rules were transcribed manually and saved into vocabulary of exceptions. After the phonetic transcription we computed probability of unigrams, bigrams and trigrams according to equation (1). Results are shown in Table 1. including the task perplexity and entropy.

Characteristics of training and test corpus			
		training	test
	number of names	20 579 070	9 488
	number of phonemes in vocabulary (V)	44	44
unigrams	unigrams (N)	174 692 140	77 100
	different unigrams (nr(.,.))	44	43
	unseen unigrams (n0(.,.))	0	1
	singletons (n1(.,.))	0	0
	doubletons (n2(.,.))	0	0
	perplexity	19.22	19.03
	entropy	4.26	4.25
bigrams	bigrams (N)	154 112 095	67 612
	different bigrams (nr(.,.))	1 392	852
	unseen bigrams (n0(.,.))	544	1 084
	singletons (n1(.,.))	35	101
	doubletons (n2(.,.))	13	60
	perplexity	9.95	10.11
	entropy	3.32	3.34
trigrams	trigrams (N)	133 532 050	58 124
	different trigrams (nr(.,.))	18 904	4 999
	unseen trigrams (n0(.,.))	66 280	80 185
	singletons (n1(.,.))	1 198	1 725
	doubletons (n2(.,.))	645	757
	perplexity	4.70	4.54
	entropy	2.23	2.18

Table 1.

6. EXPERIMENTAL RESULT

In the speech recognition system equipped by an acoustic and a language model it is practically advantageous to use different weights of the language and the acoustic model. These weights can be defined by two variables p a s. The word insertion penalty p is a fixed value added to each token when it transits from the end of one word to the start of the next. The grammar scale factor s is the amount by which the language model probability is scaled before being added to each hypothesis as it transits from the end of the word to the start of the next. These parameters can have a significant effect on recognition performance and hence, some tuning on development test data is well worthwhile.
Formulae for computing with parameters s and p:

$$\log(P(O, w_k|w_l)) = \log(P(O|w_k)) + s \cdot \log(P(w_k|w_l)) + p \tag{2}$$

where O is observation vector sequence generated by Hidden Markov Model; w_k and w_l are phonemes and P is a likelihood.

After the language model had been created we performed several tests. The first test was performed using 718 utterances. Each utterance consists of person name and surname. From 718 utterances two sets A, B were randomly chosen. Each of them contained 25 different utterances. Two training sets were constructed. The first one (T1) is equal to the set A and the second one (T2) is composes both sets A and B. The test set contained all 718 utterances. From sets A and B the baseforms were generated. The test set was recognized on basis these baseforms. Results are shown in Table 2. for the grammar scale factor $s = 1$ and the word insertion penalty $p = 30\,000$.

Number of baseforms	25 (T1)	2x25 (T2)
Number of utterances	718	718
Correctly recognized	554	601
Incorrectly recognized	164	117
Correctly recognized [%]	77.16	83.70
Incorrectly recognized [%]	22.84	16.30

Table 2.

Following series of tests were performed with the same set of 718 utterances. 25 utterances were always randomly chosen (one for every name) and all the 718 utterances were subsequently recognized with various values s and p. General results shown in Table 3. are arithmetic mean of all individual tests.

s/p	20000	25000	30000	35000	40000	50000
-1	75.81	75.07	72.28	68.71	64.02	59.52
0	79.53	79.20	75.91	74.74	70.98	63.37
1	79.02	78.46	77.21	76.42	74.47	68.48
2	76.04	78.55	79.06	77.11	76.93	74.65

Table 3.

In the final test we tried how the system works for one user. We recorded 200 utterances by one user (4 utterances for each full name from the test A. From 200 utterances we randomly chose 25 utterances of different names and for each utterance a baseform was generated. The rest (175) utterances were recognized with these baseforms. We executed two tests with different training utterances. Results shown in Table 4. are arithmetic mean of both the tests.

s/p	20000	25000	30000	35000	40000	50000
0	94	93	92	90,5	88	84
1	93.5	92.5	93	95.5	91.5	87.5
2	99.5	97.5	91.5	89	89	90.5

Table 4.

7. CONCLUSION

We have presented a system for recognition and automatic phonetic baseform generation. We used an acoustic model and a language model with bigram probabilities to constrain the transitions between phonemes.

In conclusion, we believe that we have a viable technique for automatic generation of phonetic baseforms that give a good decoding accuracy with our speech recognition system. This is particularly useful for our telephony dialogue system where personalized vocabularies are a must.

Experimental results show that the recognition based on acoustic baseforms has a good accuracy. The highest achieved accuracy ($s = 2$, $p = 20\,000$) for system working with one user is greater then 99 %.

In the future we want to provide further improvements in accuracy.

REFERENCES

[1] Müller L., Psutka J., and Šmídl L. "*Design of Speech Recognition Engine*", TSD 2000 – Third International Workshop on TEXT, SPEECH and DIALOGUE, Brno, Czech republic, September 13-16, 2000.

[2] Ramabhadran B., Bahl L. R., deSouza P. V., Padmanabhan M., "*Acoustic-Only Based Automatic Phonetic Baseform Generation*", ICASSP 1998

[3] Forney, G. D., Jr. „*The Viterbi Algorithm*", In Proceedings of the IEEE, vol.61, no.3, pp.268-278 1973

[4] Young, S., Evermann, G., Odell, J., Ollason, D., Woodland, P., Kershaw, D., Moore, G., Valtchev, V. „*The HTK Book (for HTK Version 3.1)*", Cambridge University 2001

Prosody Model and its Application to Czech TTS System

Jan Romportl, Jindřich Matoušek, Daniel Tihelka
University of West Bohemia in Pilsen, Department of Cybernetics,
Univerzitní 22, 306 14 Plzeň, Czech Republic
rompi@students.zcu.cz, jmatouse@kky.zcu.cz, dtihelka@kky.zcu.cz

ABSTRACT

The first part of this paper[1] proposes a formal theoretical framework for prosody description. This framework is based on empirically acquired axioms (following the linguistic structuralism) and in terms of the mathematical set theory it presents prosody as a relation between abstract sentence underlaying structures and intonation (together with timing). On the basis of this framework one can utilise various system description and analysis methods as well as pattern processing techniques to model the aforementioned relation. The second part of this text introduces the application of this framework to the Czech text-to-speech system ARTIC. It uses rules to place abstract intonation schemes (melodemes and cadences) depending on the position of an intonation centre of an utterance.

1. INTRODUCTION

Probably all text-to-speech (TTS) concerned papers agree that naturalness and also intelligibility of synthetic speech strongly depends on its prosodic quality. However, if one asks what such "prosodic quality" means one usually gets an answer vaguely summarising the goal of all prosody research to make TTS sound "human-like", no matter methods involved and long-term perspective offered. Indeed, this is under certain circumstances true but we feel that more comprehensive insight to this problematics is needed if the aforementioned dream of all TTS designers should be fulfilled. In this paper we present results of the initial stage of our research on prosody which tries to employ some results achieved by the functional approach of Prague Linguistic School. Couple of our results were utilised and applied to the state-of-art Czech TTS system being developed at Cybernetics department of the University of West Bohemia in Pilsen.

2. PROSODY PROPERTIES

Although "prosody" is generally known as a sort of synonym for suprasegmental features of human speech, we will try to give more formal conception. First of all we must cope with some empirical observations as well as set up what we want to achieve. The latter is explicated by an assumption, that "acceptably natural prosody" means such suprasegmental properties of synthetic speech that would be produced by a human speaker in a clear and intelligible utterance without excessive emotional concern. The former can be briefly summarised by following:

Axiom 1 (based on [1]):

I. Every continuous speech is divisible into smaller units (we will call them "phonemic clauses").
II. These units have their own specific intonation (e.g. melody and intensity, or contour of fundamental frequency and volume) and timing. The word "intonation" is used for the attribute constituted by melody and intensity.
III. These units can be separated by pauses.

Axiom 2:

"Prosodic quality" of every utterance can be fully described by intonation and timing.

Axiom 3 (for discussion see [1], [2], [3]):

Intonation and timing are "functionally involved"; they are constituted by elements where some of them have linguistic function(s) meanwhile some of them do not. Most relevant functions are delimitative and semantical.

The semantical function is (at least) of two kinds: it helps a listener create a notion of an utterance's *meaning* (as a linguistic concept) and participates in creating an utterance's (ontological) *content* (or factual knowledge) which can poorly be derived from the text itself.

Axiom 4 (see [1]):

The principle of tolerance and relevance. Each element of a certain functional layer can realise itself freely within boundaries given by a system.

We do not have enough space to discuss the above axioms and we are aware they might be somehow modified when facing future research. Further in the text we will use this notation: $<x_1, x_2,... x_n>$ is an ordered n-tuple of objects $x_1... x_n$ (in this order); relation R is a set of all 2-tuples <y, x> such that objects y and x (in this order) are in a relation R (e.g. yRx); for a relation R symbol *dom*(R) is a set of all x such that $\langle y, x\rangle \in R$; for a relation R symbol *rng*(R) is a set of all y such that $\langle y, x\rangle \in R$; for a set A symbol *pot*(A) is a set of all

1 This research is supported by the Grant Agency of Czech Republic no. 102/02/0124 and the Ministry of Education of Czech Republic, project no. MSM235200004.

subsets of A; for sets A and B operator A | B produces a set of <y, x> such that $\langle y,x\rangle \in A|B \Leftrightarrow \langle y,x\rangle \in A \wedge y \in B$ (notation is based on [10] and slightly modified).

Now we can propose the following definition:

Definition 1:

Be relation P', <IT_{SJ}, <TR_S, MR_S, A>> ∈ P', called *prosody*. TR_S is the tectogrammatical representation of a sentence S, MR_S is the morphonological representation of a sentence S, A stands for stochastic attributes (perhaps properties that cannot be modeled otherwise), IT_{SJ} is the set whose elements adequately describe intonation and timing of sentence S realised as an uttered sentence token J.

Tectogrammatical representation is such a (non-linear) representation of a sentence which describes both its meaning and its syntax in terms of so-called *tectogrammatical layer* of language description. In detail it is elaborated and described in [4], [5]. It can be also called a *semantical structure*. Morphonological form is understood as a form which represents a sentence as it appears in its very surface structure (more specially for our purposes we can call this a graphemical form). "Adequate description of intonation and timing" is a very vague term but this way we leave the question of suitability of various techniques of intonation and timing description open.

The "real" nature of the aforementioned relation is somehow *infinite* (note the original meaning of this word is closely related to *indeterminate*) thus it would be of great benefit to utilise some mathematical theory for indeterminacy description (see very promising [10]). So far we must settle for the following (maybe contra-intuitive) assumption which helps us establish some sort of "discourse universe":

Axiom 5:

For each 3-tuple <TR_S, MR_S, A> all possible uttered sentence tokens have been brought to existence.

Theorem 1:

$\forall S \in dom(P')\, \exists IT_{SI}, IT_{SJ} \in rng(P'):$

$\langle S, IT_{SI}\rangle \in P' \wedge \langle S, IT_{SJ}\rangle \in P' \Rightarrow IT_{SI} \neq IT_{SJ}$

Proof is a direct consequence of axioms 4 and 5. In short it formalises the fact that one sentence can be uttered in more ways (concerning intonation and timing).

Definition 2:

Acceptably natural prosody is the relation P = P' | Q where

$Q \in pot(rng(P'))$ such that

$\forall S \in dom(P')\, \forall q \in rng(P'):$

$$q \in Q \Leftrightarrow q = \underset{q_i:\langle S,q_i\rangle \in P'}{argmin}\ J(S, q_i)$$

where J(S, q_i) is a criterial function such that (in case of suitable indexing) J(S, q_1) < J(S, q_2) < ... < J(S, q_n).

Theorem 2:

P is a function.

Proof results from the definition 2 and from the requirements posed on the criterial function J(S, q_i) because for each $S \in dom(P)$ exactly one IT_S is given.

The responsibility for constituting functional (not in terms of linguistics but in terms of mathematics) dependency between prosody and text is thus put on the criterial function that can be represented for example by some subjective perceptional tests. Basically its goal is to choose one realization of intonation and timing that is best in terms of given criteria. Technically for the sake of artificial prosody generation (in TTS systems for instance) we may outline the criterial function so as to select such realizations which are as close to specific real data as possible.

Through the above approach we postulated and formalised functional dependency between intonation (plus timing) and abstract linguistic representation of a sentence which can be derived from graphemical (i.e. generally morphonological) representation of this sentence and its context. The reason we considered the tectogrammatical level of representation as underlying instead of some "more surface" level is straightforward: the tectogrammatical representation (TR) is able to describe contextual information with regard to the relevance of *topic-focus articulation* (TFA) which reflects communicative function of a sentence (this can deal for instance with communicational aspects of word ordering which has a significant relevance in Slavic languages). For details see [4], [5], [6], [8]. We are convinced that ignoring all these aspects of text could bring some short-term advantages but seems to be shortsighted when facing the long-term goal of cognitive sciences.

Our framework allows us to understand prosody in terms of the system theory and thus model it using methods based on system description and analysis (in practical applications a lot of work can be done by pattern processing techniques). For the importance of language structures formalising see for example [7].

3. APPLICATION FOR CZECH TTS

Concerning the above mentioned we distinguish two levels of prosody description (see [1], [2], [3], following the Prague linguistic structuralism): functionally motivated and acoustically motivated (though this term might not be best fitting). We are far from thinking some great progress can be achieved without co-operation of these two constituents so we adopted and slightly changed terminology as it is presented in [3].

Continuous text - in written form - is segmented into sentences which we understand to be particular utterances (to make the problem easier) - in spoken form. Each utterance is divided into major and (optionally) minor phonemic clauses (also called prosodic phrases) which are then rhythmically segmented into phonemic words (a phonemic word is one or more words subordinated to one word-stress). We can actually say it

is almost a rule in Czech to place a word stress in the beginning of a phonemic word and for the sake of TTS we can postulate it (in this case we must cope with sometimes occurring "pre-phonemic words"). An example:

'V posledních **'dnech** $/_1$ **'naší 'do**volené $/_2$ jsem **'ko**nečně **'po**chopil $//_2$ že **'ne**snáším **'ces**tování..
In the last days $/_1$ of our holiday $/_2$ I finally realised $//_2$ that I hate traveling.

Major phonemic clauses are divided by $/_2$, minor phonemic clauses by $/_1$ (these are quite optional and strongly depends on speech rate) and // signs pauses. Bold characters show stressed syllables and ' stands for phonemic word beginnings. So far we have implemented rather simple rule-based method of phonemic words and word stresses placement but it has proved to be efficient enough because of relative simplicity of this area in Czech language (word stress has only a delimitative function and does not change the meaning anyway). Phonemic clauses detection is being developed and has not been implemented sufficiently yet.

Further in this text we will speak mostly about intonation (and more specifically about melody) since for Czech it is the most important attribute of prosody (see [1], [3]) and the only one implemented in the TTS system ARTIC so far (see [9]). There is also no tectogrammatical parser available for us which means that at this stage of research and development we cannot fulfil the goals appointed by the theoretical part of this paper. However, they (together with the above formalization) should be kept in mind as well as the fact we want to develop a prosody model for TTS, not describe and formalise language system.

Prosody of an utterance is described (as it was already mentioned) in two levels. The first level consists of so called *melodemes*. Melodeme is an abstract intonational pattern established in certain function within a language system. The second level is characterised purely by acoustic attributes irrespective to their function. If we stay under "protective wings" of *acceptably natural prosody* we can describe this level in terms of *cadences*. We can understand cadence to be a model generating elements of IT_{SI} set, but - an important note - the input of such model must be a structure much more simple than $<TR_S, MR_S, A>$. In short - cadence describes "real" acoustic properties of intonation without regard to a meaning or content of an utterance.

So far we use simple rule-based model of cadences of this inventory (in terms of the F0 contour slope): flat, ascending, descending, ascending-descending, descending-ascending. Each cadence can be enhanced by the attribute "stressed" (e.g. it models F0 contour with a word stress) and is presented (except for the flat cadence) in three forms (simply distinguished by numbers 1, 2, 3) which vaguely express "how quick F0 moves up or down" (we will refer to it as the tendency of a cadence). Each cadence places a number of F0 values (in Hertz) at certain timestamps and all such values are then linearly interpolated producing the final F0 contour. These cadences are intended to underlay phonemic words. It means that a cadence is quite a simple unit covering only one phonemic word and thus concerning just morphonological form of a phonemic word (e.g. a subset of MR_S). For example the flat stressed cadence places three values: first at the beginning of the first phoneme of a phonemic word, then local maximum at the beginning of the nucleus of the stressed syllable, then local minimum at the end of the last phoneme of a phonemic word. The concrete numbers are calculated using referential F0 value (supplied by the superordinate layer - melodemes, as will be seen later) and coefficients specific to particular cadence (indeed the coefficients are dependent on a speaker and are set up experimentally).

Melodeme placement depends on the position of so called *intonation centre*. Normal position of an intonation centre in Czech is on the last phonemic word of an utterance. This position is automatically understood to designate emotionally neutral utterance. However, contextual factors of TFA (especially word order) may change the position and this can be expressed by TR_S (for details see [4], [6], [8]). Since we do not have a tectogrammatical parser so far (as it was already mentioned) we must actually omit the structure given by TR_S and we place intonation centre always on its "normal" position. There are also phonemic clause intonation centers and these are automatically placed on the last phonemic word of phonemic clauses.

Intonation centre - because of its function - is a border between two neighbouring melodemes (indeed it is just a conceptual point of view - the function and the intonation centre itself is actually realised by particular melodemes or cadences respectively). This means that an intonation centre starts a new melodeme of the type dependent on a sentence modality (declarative, interrogative, imperative, etc.) and other aspects (not realised in this phase yet). We currently use following melodemes (based on [2], [3]):

M0 - null melodeme
M1 - terminating descending melodeme
  M1-1 - neutral
M2 - terminating ascending melodeme
  M2-1 - neutral
M3 - nonterminating melodeme
  M3-1 - neutral
  M3-2 - neutral, pause preceding

Indeed we should distinguish more melodemes but those written above are quite sufficient for emotionally neutral utterance description (when almost omitting TR_S, and accounting only for a sentence modality as it is our case at this stage of research and prosody application). M0 is understood as a formal description of such a segment of a phonemic clause where actually no melodeme is realised. It ranges from the beginning of a phonemic clause to the phonemic word right before the intonation centre of the phonemic clause and it describes slightly descending tendency of phonemic clause melody. M1 is used for an utterance terminating in case the sentence is declarative, imperative or interrogative-

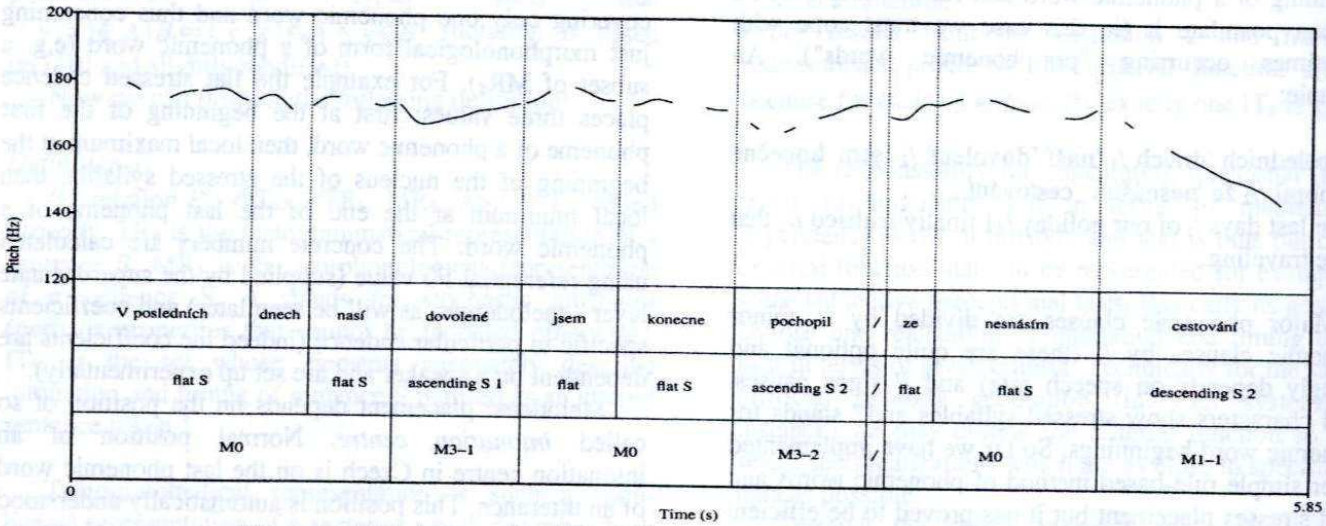


Figure 1: Melodeme and cadence placement in the resulting synthesised speech

supplementary (like English "wh-" questions). It starts at the intonation centre and reaches the end of the utterance. M2 is very similar to M1 with the difference that it is used to terminate interrogative-inquiring sentence (e.g. questions with a "yes/no" answer). M3 is used in those phonemic clauses which are not last in particular utterances.

Melodemes are chosen to fulfil particular function given by the communicative intention and cadences can be understood as tools intended for their concrete realization. Figure 1 transparently illustrates the usage of cadences and melodemes on the example of the above introduced sentence (without minor phonemic clauses).

4. CONCLUSION

The formalization of prosody in the part 2 should not be taken linguistically rigorous. It is certainly purpose build, however, it tries to establish solid ground on which one can base prosody models applied in many systems such as TTS. Moreover, it implies the dependency between the semantical structure and a context of a sentence which is necessary to take into account to build as much as possible human-like sounding TTS system.

The presented rule-based method of prosody description is a very simplified realization of this formalization. However, the results - as can be evaluated with Czech TTS ARTIC system - are quite satisfactory, taking the simplicity of so far used method into account. The intonation is quite far from being perfectly human-like but this can be much improved by the use of more sophisticated cadence models (for example stochastic or neural network based). Moreover, since the prosody description is separated into two autonomous parts (functional and non-functional), changes made to one part almost do not influence the other one and the formalization sets up a solid area for a further research.

The future work will focus on more elaborate cadence models (based on the stochastic system description) and will enhance timing properties, especially syllable duration. More thorough insight into the intensity attribute of intonation will be undertaken too (this means modelling intensity contour not only at stressed syllables).

5. REFERENCES

1. Daneš, F. "Intonace a věta ve spisovné češtině (Sentence Intonation in Present-Day Standard Czech)" , Nakladatelství Československé akademie věd, Prague, 1957
2. Romportl, M. "Základy fonetiky (Basics of Phonetics)", Prague, 1973
3. Palková, Z. "Fonetika a fonologie češtiny (Phonetics and Phonology of Czech)", Karolinum, Prague, 1994
4. Sgall, P. - Hajičová, E. - Panevová, J. "The Meaning of the Sentence in Its Semantic and Pragmatic Aspects", Reider, Dordrecht, 1986
5. Panevová, J. "Formy a funkce ve stavbě české věty (Forms and Function in Czech Syntax)", Academia, Prague, 1980
6. Sgall, P. - Hajičová, E. - Buráňová, E. "Aktuální členění věty v češtině (Topic and Focus in Czech)", Academia, Prague, 1980
7. Hajič, J. - Hajičová, E. - Rosen, A. "Formal Representation of Language Structures", In TELRI Newsletter No.3, pp. 12-19, June 1996
8. Hajičová, E. "The Prague Dependency Treebank: Crossing the Sentence Boundary", In Proceedings of the Second Workshop on Text, Speech, Dialogue, pp. 20-27, Marianske Lazne, Czech Republic, 1999
9. Matoušek, J. - Psutka, J. "ARTIC: A New Czech Text-to-Speech System Using Statistical Approach to Speech Segment Database Construction", In The Proceedings of the 6th International Conference on Spoken Language Processing ICSLP2000, vol. IV. Beijing, China, 2000, pp. 612-615.
10. Vopěnka, P. "Úvod do matematiky v alternatívnej teórii množín (Introduction to Mathematics in Alternative Set Theory)", Alfa, Bratislava, 1989

АНАЛІЗ ТЕМПОРАЛЬНИХ ПЕРЕТВОРЕНЬ МОВНИХ ЕЛЕМЕНТІВ ДЛЯ ЗАДАЧ ЧАСОВОГО МАСШТАБУВАННЯ ГОЛОСОВИХ ПОВІДОМЛЕНЬ

Зореслава Шпак, Юрій Рашкевич

Національний університет "Львівська політехніка"
вул. Ст. Бандери, 12, м. Львів-13, 79646; тел.: (38-0322) 398-793; факс: 744-143
Електронна пошта: zshpak@polynet.lviv.ua; rashkev@polynet.lviv.ua

ABSTRACT

The characteristics of the speech units (sounds, syllables, words, etc.) duration have been specified for different pronunciation types. Time-scale transformation regularity of the Ukrainian voice phonemes sounding at different speech rates has been investigated. Six groups of oral speech elements are determined: stressed vowels, unstressed vowels, voiced consonants, unvoiced consonants, plosive stops and word-spacing pauses. The elements of specified groups are characterized by similar temporal transformations. Non-linear analytic function presenting the relative temporal changes of sound duration have been plotted for every group.

1. ВСТУП

Підвищення ефективності усномовної комунікації в людино-машинних системах ставить задачу керування темпом надходження голосової інформації. Щоб забезпечити розбірливість і натуральність звучання мови, відтвореної зі зміненою швидкістю, у процесах часового масштабування голосових записів треба враховувати природні темпоральні властивості мовних елементів [1].

У роботі наведено часові параметри структурних одиниць мовного потоку для різних темпів мовлення. Проаналізовано закономірності темпоральних змін звуків української мови і виділено класи мовних елементів, що характеризуються спільністю перетворень, викликаних зміною швидкості мовлення. Побудовано графіки, що відображають залежність відносної зміни тривалості елементів виділених класів від загального коефіцієнта зміни темпу.

2. ЧАСОВІ ХАРАКТЕРИСТИКИ РІЗНИХ ТЕМПІВ МОВЛЕННЯ

Для визначення темпоральних параметрів мовлення використано набір текстів, які відображають сучасні тенденції у структурі української розмовної, ділової та наукової мови. Кожен з текстів (загальний обсяг мовного матеріалу – 167 речень та відповідно: 2138 слів, 5331 складів і 12315 фонем) промовлявся групою з семи мовців жіночої та чоловічої статі в чотирьох темпах: швидкому, звичайному, повільному та протяжному. У табл.1 наведено темпоральні дані двох мовців (M1 і M2), які відрізняються стилями мовлення та усереднені значення для всієї групи мовців.

Основні підсумки отриманих результатів:

- найбільш стабільною і незалежною від тексту характеристикою швидкості мовлення є параметр "складів за секунду" – його доцільно застосовувати як базову темпоральну характеристику;

Таблиця 1

Часові характеристики основних стилів мовлення

Параметр		Темп мовлення								
		швидкий			звичайний			повільний		
		M1	M2	сер.	M1	M2	сер.	M1	M2	сер.
Середня тривалість	речення (с)	8,16	5,29	5,79	12,43	5,10	7,57	20,34	9,69	12,08
	синтагми (с)	2,37	1,47	1,68	3,53	2,34	2,47	5,64	2,44	3,45
	слова (мс)	552	341	415	734	441	540	1106	556	667
	складу (мс)	221	131	159	302	173	206	419	213	267
	звуку (мс)	94	56	68	129	74	88	179	91	114
Швидкість мовлення	слів/хвилину	96,5	135,8	130,1	64,6	119,7	89,1	39,2	86,1	62,3
	складів/секунду	4,1	6,7	5,3	2,6	4,7	3,7	1,6	3,4	2,5
	звуків/секунду	9,2	16,3	13,2	6,3	11,5	9,0	3,7	8,2	6,2
Коефіцієнт прискорення/сповільнення		1,51	1,43	1,46	1,0	1,0	1,0	1,64	1,39	1,43

• варіативність темпоральних показників мовців є достатньо широкою, особливо у випадках швидкої та повільної вимови. Так, М1 володіє достатньо широким діапазоном зміни темпу і помірною швидкістю мовлення, а М2 може бути віднесений до групи швидких мовців – властивий для нього звичайний темп є близьким до швидкого мовлення М1;

• незважаючи на відмінності в абсолютних параметрах швидкості мовлення, всім мовцям властиві достатньо близькі значення відносної зміни темпу. Середні по всій групі коефіцієнти максимального прискорення та сповільнення складали відповідно 1,46 та 1,48 (коефіцієнти сповільнення характеризувались значно ширшим розкидом значень, ніж коефіцієнти прискорення);

• часові параметри мовця не залежать від його статі, а визначаються властивим даній людині стилем мовлення. Усім мовцям, які не мали спеціальної дикторської практики, важко було довготривало відхилятись від звичайної швидкості вимови – чим довшим був текст, тим меншим ставав загальний коефіцієнт зміни темпу.

3. ТЕМПОРАЛЬНІ ПЕРЕТВОРЕННЯ ПАУЗ І ЗВУКІВ

3.1. Темпоральні перетворення пауз.

Зміна темпу мовлення пов'язана з перерозподілом структури мовного потоку (рис.1). Перехід від звичайної мови до швидкої чи повільної насамперед відображається в зміні кількості та тривалості пауз: кількість пауз зменшується/зростає усереднено в 1,5 разів, а їх сумарна тривалість змінюється відповідно приблизно в 2,3 рази (табл.2). Водночас загальна тривалість звукової частини мовного потоку скорочується/збільшується тільки в 1,3 рази. Найбільший відсоток пауз у повільній мові – біля 30%. Відзначено, що при подальшому сповільненні темпу дещо зростає кількість пауз, але їх середня тривалість майже не змінюється. Як результат – у протяжній мові частка пауз зменшується до 20%.

Довгі синтагматичні паузи загалом характеризуються меншими і стабільнішими темпоральними змінами, ніж короткі міжслівні, які вирізняються найвищим коефіцієнтом варіації та є найбільш

Рис.1. Відсоток пауз і звуків у загальній тривалості мови

залежними від просодичних особливостей мовців. Середнє значення тривалості звукової ділянки між паузами, яке для експериментальних даних було близьким до 0,96 с, зазнає найменших темпоральних змін . Тільки в протяжному, близькому до співочого мовленні цей параметр зростає в 1,5 рази.

3.2. Зміни голосних звуків.

Для аналізу темпоральних перетворень звуків виконано аудіо-візуальний поділ мовних сигналів кожного із записів. Границі звуків контролювались та уточнювались за динамічними спектрограмами, що дозволило виділити складові частини звуків: початкові та кінцеві переходи і серединні квазістаціонарні ділянки.

Основна різниця відзначена в темпоральній поведінці голосних і приголосних (передусім шумних) звуків. Сповільнення темпу завжди проявлялось у зростанні частки голосних звуків – співвідношення голосні:приголосні для швидкого темпу складало 0,42:0,58; для звичайного – 0,45:0,55; для повільного – 0,51:0,49; а для протяжного – 0,59:0,41. У випадках переходу до швидкого чи повільного мовлення найбільше змінювались тривалості наголошених голосних (табл. 3), але в протяжній вимові спостерігалось невелике зменшення інтонаційного співвідношення між наголошеними і ненаголошеними звуками (рис.2). Місце голосного звуку в слові та

Таблиця 2

Темпоральні характеристики роздільчих пауз

Характеристика	Темп мовлення			
	швидкий	звичайний	повільний	протяжний
Тривалість пауз у 1с мовного потоку (мс)	95	198	293	214
Кількість пауз на слово	0,41	0,56	0,81	0,97
Середня тривалість паузи (мс)	141,6	238,1	383,4	405,7
СКВ тривалості пауз (мс)	129,1	188,7	215,3	273,7
Середня тривалість звукової ділянки між паузами (мс)	1012	664	905	1420
Відносне збільшення загальної тривалості пауз	1,0	2,47	5,73	6,41
Відносне збільшення середньої тривалості паузи	1,0	1,67	2,72	2,84

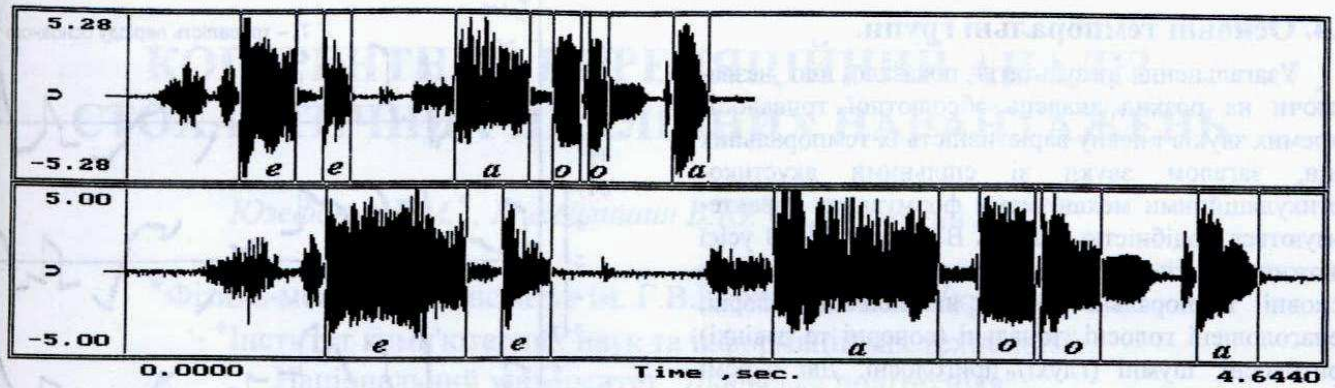


Рис.2. Мовні сигнали словосполучення *щебет жайворонка*, вимовленого в звичайному та протяжному темпах

просодика цього слова мали певний вплив на абсолютну тривалість звуку, проте відносні зміни тривалостей (ВЗТ) всіх голосних, викликані зміною темпу мовлення, були загалом достатньо однорідними.

3.3. Зміни приголосних звуків.

Відмінність у темпоральних перетвореннях голосних і приголосних звуків набуває принципового характеру у випадках значного сповільнення темпу (рис.2). Так, перехід від швидкого до протяжного мовлення пов'язаний із загальним збільшенням тривалості голосних у 4,4 рази, а приголосних – тільки в 2,2 рази. Водночас темпоральна поведінка різних приголосних звуків є теж достатньо неоднорідною (табл. 3). Проведено аналіз темпоральних змін для кожного звуку зокрема та узагальнено для основних класифікаційних груп приголосних [2]. В межах кожної групи враховувався вплив факторів пом'якшення, дзвінкості, назалізації та інших. Зроблено такі основні висновки:

- у випадках невеликих змін темпу (коефіцієнт прискорення/сповільнення не перевищує 1,4) всі приголосні змінюють свою тривалість приблизно однаково; при більших змінах темпу проявляються відмінності в перетвореннях приголосних різної артикуляційної природи;
- перетворення сонорних звуків при невеликих змінах темпу є близькими до перетворень ненаголошених голосних, а при значних змінах темпу вони є ближчими до перетворень дзвінких приголосних;
- темпоральні зміни зімкнених звуків проявляються в неоднаковій зміні їх складових: паузи зімкнення та наступної короткої звукової ділянки (табл. 3);
- як для сонорних, так і для щілинних та зімкнених звуків не встановлено істотних відмінностей у темпоральних змінах тривалостей м'яких та відповідних твердих звуків, за винятком випадків, коли м'які приголосні виступали як подовжені;
- приголосні з тональною складовою (дзвінкі звуки) дещо відрізняються за параметром тривалості, а також за значеннями ВЗТ від відповідних глухих приголосних (табл. 3), особливо у випадках протяжного мовлення;
- збільшення тривалості початково довгих звуків (зокрема подовжених приголосних) і скорочення початково коротких (зокрема вибухових) відбувається в меншій мірі, ніж для інших звуків даної групи.

Таблиця 3

Характеристики темпоральної тривалості звуків базових класів

Класи звуків				Темп мовлення											
				швидкий			звичайний			повільний			протяжний		
				ПВ, %	СТ, мс	ВЗТ	ПВ, %	СТ, мс	ВЗТ	ПВ, %	СТ, мс	ВЗТ	ПВ, %	СТ, мс	ВЗТ
Голосні	ненаголошені			22,6	59,5	1,32	22,8	78,8	1,0	25,0	105,8	1,34	29,6	239,5	3,04
	наголошені			19,5	91,1	1,42	21,9	129,6	1,0	25,2	181,5	1,40	29,8	408,5	3,15
Приголосні	сонорні			20,6	49,6	1,30	19,9	64,7	1,0	17,5	84,6	1,31	15,1	119,9	1,85
	шумні	щілинні	дзвінкі	4,8	61,7	1,27	4,6	78,6	1,0	4,1	98,9	1,26	3,4	136,8	1,74
			глухі	8,9	99,4	1,26	8,7	125,6	1,0	7,7	146,5	1,17	5,3	168,1	1,34
		зімкнені	дзвінкі	4,8	36,5	1,32	4,7	48,0	1,0	4,0	60,9	1,27	3,2	79,5	1,66
			глухі	2,6	24,6	1,25	2,5	30,7	1,0	2,1	36,6	1,19	1,2	42,3	1,38
			змикання	10,2	71,0	1,27	10,6	90,1	1,0	10,8	109,1	1,21	8,3	146,4	1,62
		африкати		2,8	77,1	1,28	2,6	98,4	1,0	2,4	110,8	1,13	1,7	127,6	1,30
		подовжені		1,7	108,7	1,31	1,7	142,4	1,0	1,6	178,3	1,25	1,3	216,8	1,52

Примітка: ПВ - процентний вміст; СТ - середня тривалість; ВЗТ - відносна зміна (збільшення/зменшення) тривалості.

3.4. Основні темпоральні групи.

Узагальнення результатів показало, що незважаючи на розкид значень абсолютної тривалості окремих звуків і певну варіативність їх темпоральних змін, загалом звуки зі спільними акустико-артикуляційними механізмами формування характеризуються подібністю значень ВЗТ (табл. 3). З усієї множини звуків української мови виділено чотири основні темпоральні групи: наголошені голосні, ненаголошені голосні, тональні (сонорні та дзвінкі) приголосні, шумні (глухі) приголосні. Дві окремі групи складають паузи зімкнення та роздільчі (міжслівні) паузи, темпоральні модифікації яких є принципово різними. На рис.3 наведено графіки ВЗТ мовних елементів основних темпоральних груп відносно значень, властивих для швидкого темпу (k – загальний коефіцієнт сповільнення вимови).

Зміна швидкості мовлення викликає неоднакові перетворення стаціонарної і перехідних ділянок звуків. Якщо зміни тривалості стаціонарів є достатньо регулярними, то трансформації перехідних ділянок характеризуються великою варіативністю. Загалом змінюються тривалості обох переходів, але усереднено в 1,4 (а в разі значних сповільнень – в 1,7) рази менше, ніж стаціонарних частин звуків.

В процесі досліджень проаналізовано вплив темпу мовлення на значення тривалості періодів основного тону (ПОТ) звуків, а також на форму мелодичного контуру фраз. Встановлено, що середні значення тривалості ПОТ тональних звуків і мелодика мови є практично незалежними від швидкості мовлення (рис.4). Зафіксовано тільки невелике зростання (біля 4%) середньої тривалості ПОТ у протяжному темпі та модуляції значень ПОТ для деяких довгих голосних. Темпорально незалежними виявилися також розподіли формантних частот і спектральні структури шумних звуків.

4. ВИСНОВКИ

Природна зміна швидкості мовлення пов'язана зі зміною тривалості всіх елементів мовного потоку. У разі невеликих змін темпу основних трансформацій зазнають паузи, а модифікації всіх звуків є достатньо подібними і можуть бути описані лінійною залежністю від коефіцієнта зміни темпу. У випадках сповільнення/прискорення більше, ніж у 1,5 рази проявляються розбіжності у темпоральних перетвореннях звуків з різними механізмами формування. Ці відмінності стають принциповими для протяжної мови.

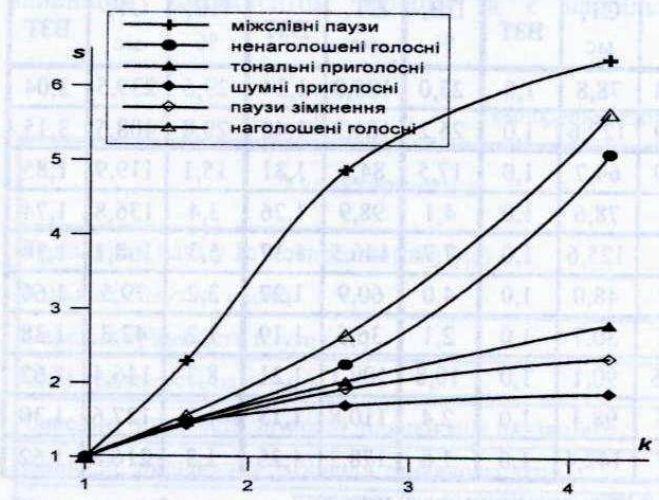


Рис.3. Графіки ВЗТ елементів основних темпоральних груп

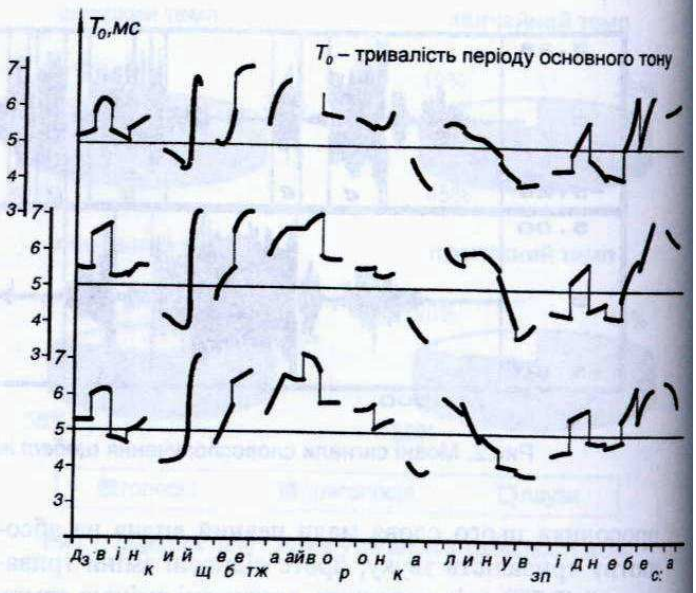


Рис.4. Мелодичні контури тестової фрази у звичайному (нижній рисунок), повільному (середній) і протяжному (верхній) темпах

За близькістю значень ВЗТ і подібністю перетворення структур сигналів, викликаних зміною швидкості мовлення, всі звуки української мови можна об'єднати в чотири основні темпоральні групи; ще дві окремі групи формують паузи. Залежність ВЗТ елементів кожної з цих груп від величини загального коефіцієнта зміни темпу носить нелінійний характер. Особливості зміни тривалості надмірно довгих чи коротких звуків вимагають додатковго введення функції нормування [3]. Підтверджено, що основні спектральні характеристики звуків є темпорально стабільними.

Таким чином, реалізацію зміни швидкості відтворення мовних записів за умови натуральності звучання голосового повідомлення необхідно здійснювати через диференційовану зміну тривалості мовних елементів відповідно до властивих їм значень ВЗТ, зберігаючи загальну структуру звукових сигналів, передусім тональних.

ЛІТЕРАТУРА

1. Рашкевич Ю.М. *Перетворення часового масштабу мовних сигналів*. - Львів: Академічний експрес, 1997. - 140 с.
2. Тоцька Н.І. *Сучасна українська літературна мова*. – Київ: Вища школа, 1981. – 183 с.
3. З. Шпак. *Нормування тривалостей звуків у процесі сповільненого відтворення мовних повідомлень* // Вісник Націон. ун-ту "Львівська політехніка": Львів, 2001. - №433. - С. 266-271.

КОГЕРЕНТНИЙ КОРЕЛЯЦІЙНИЙ АНАЛІЗ СТОХАСТИЧНИХ ЦИКЛІЧНИХ НАВАНТАЖЕНЬ

Юзефович Р.М.[♥], *Михайлишин В.Ю.*[♥♦], *Яворський І.М.*[♥♠]

[♥]Фізико-механічний інститут ім. Г.В.Карпенка НАН України, Львів
[♦]Інститут комп'ютерних наук та інформаційних технологій
Національний університет "Львівська політехніка"
[♠]Інститут телекомунікації Технічно-сільськогосподарської академії, Бидгощ, Польща

МСП, вул. Наукова 5, м. Львів, 79601
Тел: 8 (0322) 65-48-80, факс: 8 (0322) 64-94-27
E-mail: abzac@ipm.lviv.ua, mykhai@ipm.lviv.ua, iavor@ipm.lviv.ua

Розглянуто властивості когерентних оцінок ймовірнісних характеристик періодично корельованих випадкових процесів (ПКВП). Досліджено ймовірнісну структуру циклічних навантажень при вібрації підшипникового вузла з нерівномірним радіальним натягом.

1. ВСТУП

Різноманітні конструкції в процесі їх експлуатації зазнають навантажень, які здебільшого змінюються в часі стохастично. Такі навантаження зумовлені як зовнішними чинниками (пориви вітру, акустичний шум, морське хвилювання), так і внутрішніми (обертовий рух підшипників, валів, коліс трибових передач, приводних ременів, поступово-зворотний рух поршнів). Втома конструкційних матеріалів, яка виникає під впливом стохастичних навантажень, теж має стохастичний характер. Її можна розглядати як неспадний випадковий процес $\xi(t)$, характеристики якого залежать від характеристик навантажень. Для оцінки ступеня втоми конструкційних матеріалів необхідно знати ті характеристики навантажень, які найбільш істотні для розвитку втоми. Такі характеристики можуть бути обгрунтовані тільки на основі певних математичних моделей навантажень. Побудова математичних моделей навантажень є одним з основних етапів при дослідженні втомних пошкоджень матеріалів [1].

Особливу роль у виникненні втомних пошкоджень відіграють так звані циклічні навантаження, під впливом яких напруження в певні моменти часу змінюють свій знак на протилежний. При традиційному підході до аналізу втоми циклічні навантаження описуються за допомогою детермінованих періодичних функцій часу. За основну характеристику ступеня втоми тоді приймається число циклів при заданій амплітуді гармонічної зміни напружень. Детерміновані періодичні моделі в ідеалізованій формі описують повторюваність навантаження, але не відображають їх стохастичності. Однією з найпростіших моделей стохастичної циклічності є періодично корельовані випадкові процеси (ПКВП) [2]. Властивості циклічних навантажень в рамках такого підходу описуються за допогою імовірнісних характеристик даного класу нестаціонарних випадкових процесів. Верифікація такої моделі може бути проведена тільки на основі методів статистики ПКВП. Один з таких методів, що грунтується на усередненні відліків через період корельованості ПКВП, аналізується в даній праці.

2. ВЛАСТИВОСТІ КОГЕРЕНТНИХ ОЦІНОК ЙМОВІРНІСНИХ ХАРАКТЕРИСТИК

Математичне сподівання ПКВП $m(t) = E\xi(t)$, де E – оператор усереднення, і кореляційна функція $b(t,u) = E\,\mathring{\xi}(t)\mathring{\xi}(t+u)$, $\mathring{\xi}(t) = \xi(t) - n(t)$ є періодичними функціями часу і можуть бути подані у вигляді рядів Фур'є:

$$m(t) = \sum_{k\in Z} m_k e^{ik\omega_0 t} = m_0 + \sum_{k\in Z}\begin{pmatrix} m_k^c \cos k\omega_0 t + \\ + m_k^s \sin k\omega_0 t \end{pmatrix},$$

де $\omega_0 = \dfrac{2\pi}{T}$,

$$b(t,u) = \sum_{k\in Z} B_k(u) e^{ik\omega_0 t} =$$

$$= B_0(u) + \sum_{k\in Z}\left[B_k^c(u)\cos k\omega_0 t + B_k^s(u)\sin k\omega_0 t\right]. \quad (1)$$

Математичне сподівання описує регулярні цикли навантажень, дисперсія $d(t)=b(t,0)$ – зміну потужності флуктуацій, а кореляційна функція – зміну кореляційних зв'язків між значеннями флуктуаційних навантажень в моменти часу t і $t+u$.

Когерентний кореляційний аналіз ПКВП містить наступні завдання. По-перше, це оцінювання періодичного в часі математичного сподівання $m(t)$, $t\in[0,T]$, а також коефіцієнтів Фур'є m_k, які є параметрами форми регулярної складової циклу. По-друге, це оцінювання кореляційної функції $b(t,u)$, яка залежить від двох аргументів: часу t і зсуву u. Особливості зміни кореляційної функції відносно t і u зумовлюють різні підходи до оцінювання характеристик цих змін. Періодичність кореляційної функції, а також математичного сподівання відносно часу t забезпечує можливість використання такого методу оцінювання, який грунтується на усередненні відліків реалізації через період T. Оскільки для періодичної функції таке усереднення дає її значення в моменти часу, що відповідають точкам відліків, то є підстава сподіватися, що у випадку ПКВП, характеристики яких задовільняють певні умови, аналогічні дії будуть зменшувати вплив флуктуаційних складових. Такий метод оцінювання називають когерентним [2]. Знання значень характеристик для $t\in[0,T]$ дає можливість оцінити компоненти m_k і $B_k(u)$ на основі перетворень Фур'є. При такому оцінюванні зсув u вважається параметром, який можна змінювати. Вирази для оцінок компонентів тоді подібні до виразів для оцінок кореляційної функції стаціонарного випадкового процесу з тією різницею, що у випадку ПКВП при усередненні застосовують вагові функції $\varphi_n(t)=e^{-in\omega_0 t}$, $n\in Z$. Можливості когерентного методу оцінювання, його ефективність, очевидно, можуть бути досліджені тільки при проведені аналізу стохастичних властивостей оцінок з врахуванням специфічних особливостей певного типу циклічних навантажень. Припустимо, що усереднення виконується з використанням відрізка реалізації довжиною $\theta=NT$, N – натуральне число. Когерентна оцінка математичного сподівання має вигляд:

$$\hat{m}(t)=\frac{1}{N}\sum_{n=0}^{N-1}\xi(t+nT),\qquad t\in[0,T]. \tag{2}$$

Така оцінка є незміщеною $E\hat{m}(t)=m(t)$, а її дисперсія визначається виразом

$$D[\hat{m}(t)]=\frac{1}{N}\sum_{n=-N+1}^{N-1}\left(1-\frac{|n|}{N}\right)b(t,nT). \tag{3}$$

Якщо кореляції загасають зі збільшенням зсуву u

$$\lim_{|u|\to\infty} b(t,u)=0 \tag{4}$$

то $D[\hat{m}(t)]\to 0$ при $N\to\infty$. Це означає, що оцінка (2) є слушною. Дисперсія $D[\hat{m}(t)]$, як видно з (3), є періодичною функцією часу. Тому її можна подати у вигляді:

$$D[\hat{m}(t)]=\gamma_0+\sum_{l\in N}(\gamma_l^c\cos l\omega_0 t+\gamma_l^s\sin l\omega_0 t). \tag{5}$$

Для коефіцієнтів цього розкладу, використовуючи подане (1), отримуємо

$$\gamma_0=\frac{1}{N}\left[B_0(0)+2\sum_{n=1}^{N-1}\left(1-\frac{n}{N}\right)B_0(nT)\right], \tag{6}$$

$$\gamma_l^c=\frac{1}{N}\left[B_l^c(0)+2\sum_{n=1}^{N-1}\left(1-\frac{n}{N}\right)B_l^c(nT)\right], \tag{7}$$

$$\gamma_l^s=\frac{1}{N}\left[B_l^s(0)+2\sum_{n=1}^{N-1}\left(1-\frac{n}{N}\right)B_l^s(nT)\right]. \tag{8}$$

Як бачимо, середнє значення дисперсії (3) визначається нульовим кореляційним компонентом, а амплітуда L–их складових залежить тільки від кореляційних компонентів тих самих номерів.

На основі оцінки математичного сподівання (2) можуть бути знайдені оцінки коефіцієнтів $\hat{m}_0$, $\hat{m}_k^c$, $\hat{m}_k^s$:

$$\hat{m}_0=\frac{1}{T}\int_0^T\hat{m}(t)dt,\qquad \hat{m}_k^c=\frac{2}{T}\int_0^T\hat{m}(t)\cos k\omega_0 t\,dt, \tag{9}$$

$$\hat{m}_k^s=\frac{2}{T}\int_0^T\hat{m}(t)\sin k\omega_0 t\,dt. \tag{10}$$

Оцінки (9)-(10) є незміщеними, а їх дисперсії виражаються через відповідні кореляційні компоненти:

$$D[\hat{m}_0]=\frac{2}{\theta}\int_0^\theta\left(1-\frac{4}{\theta}\right)B_0(u)du, \tag{11}$$

$$D[\hat{m}_l^c]=\frac{2}{\theta}\int_0^\theta\left(1-\frac{4}{\theta}\right)\begin{bmatrix}2B_0(u)\cos l\omega_0 u+\\+B_{2l}^s(u)\sin l\omega_0 u-\\-B_{2l}^c(u)\cos l\omega_0 u\end{bmatrix}du, \tag{12}$$

$$D[\hat{m}_l^s]=\frac{2}{\theta}\int_0^\theta\left(1-\frac{4}{\theta}\right)\begin{bmatrix}2B_0(u)\cos l\omega_0 u-\\-B_{2l}^s(u)\sin l\omega_0 u+\\+B_{2l}^c(u)\cos l\omega_0 u\end{bmatrix}du. \tag{13}$$

Якщо кореляційні зв'язки ПКВП загасають зі збільшенням зсуву, то величини (11)-(13) прямують до нуля при $\theta\to\infty$.

Когерентну оцінку кореляційної функції запишемо у вигляді:

$$\hat{b}(t,u)=\frac{1}{N}\sum_{n=0}^{N-1}\xi(t+u+nT)\xi(t+nT)-$$
$$-\hat{m}(t)\hat{m}(t+u). \tag{14}$$

Усереднюючи (14), для зміщення знаходимо

$$\varepsilon\left[\hat{b}(t,u)\right] = E\hat{b}(t,u) - b(t,u) =$$

$$= -\frac{1}{N}\sum_{n=-N}^{N}\left(1-\frac{|n|}{N}\right)b(t,u+nT).$$

Цей вираз легко перетворюється до вигляду:

$$\varepsilon\left[\hat{b}(t,u)\right] = \varepsilon_0(u) + \sum_{l\in N}\begin{bmatrix}\varepsilon_l^c(u)\cos l\omega_0 t + \\ + \varepsilon_l^s(u)\sin l\omega_0 t\end{bmatrix}, \quad (15)$$

де $\varepsilon_0(u) = -\frac{1}{N}\sum_{n=-N+1}^{N-1}\left(1-\frac{|n|}{N}\right)B_0(u+nT),$

$$\varepsilon_l^c(u) = -\frac{1}{N}\sum_{n=-N+1}^{N-1}\left(1-\frac{|n|}{N}\right)B_l^c(u+nT), \quad (16)$$

$$\varepsilon_l^s(u) = -\frac{1}{N}\sum_{n=-N+1}^{N-1}\left(1-\frac{|n|}{N}\right)B_l^s(u+nT). \quad (17)$$

З отриманих формул випливає, що при виконанні умови (4) оцінка (14) є асимптотично незміщеною. Для її дисперсії, припускаючи, що процес є гаусовим, в першому наближені маємо:

$$D\left[\hat{b}(t,u)\right] = \alpha_0(u) + \sum_{l\in N}\begin{bmatrix}\alpha_l^c(u)\cos l\omega_0 t + \\ + \alpha_l^s(u)\sin l\omega_0 t\end{bmatrix}, \quad (18)$$

при цьому

$$\alpha_0(u) = \frac{1}{N}\left[\widetilde{B}_0(0,u) + 2\sum_{n=1}^{N-1}\left(1-\frac{n}{N}\right)\widetilde{B}_0(nT,u)\right], \quad (19)$$

$$\alpha_l^c(u) = \frac{1}{N}\left[\widetilde{B}_l^c(0,u) + 2\sum_{n=1}^{N-1}\left(1-\frac{n}{N}\right)\widetilde{B}_l^c(nT,u)\right], \quad (20)$$

$$\alpha_l^s(u) = \frac{1}{N}\left[\widetilde{B}_l^s(0,u) + 2\sum_{n=1}^{N-1}\left(1-\frac{n}{N}\right)\widetilde{B}_l^s(nT,u)\right]. \quad (21)$$

Величини $\widetilde{B}_0(u_1,u)$ і $\widetilde{B}_l^{c,s}(u_1,u)$ є коефіцієнтами Фур'є кореляційної функції процесу $\eta(t,u) = \overset{\circ}{\xi}(t)\overset{\circ}{\xi}(t+u)$ і виражаються через кореляційні компоненти ПКВП, тому при виконанні умови (4) $D\left[\hat{b}(t,u)\right] \to 0$ при $N \to \infty$.

Оцінки кореляційних компонентів знаходяться за формулами

$$\hat{B}_0(u) = \frac{1}{T}\int_0^T \hat{b}(t,u)dt, \quad \hat{B}_l^c(u) = \frac{2}{T}\int_0^T \hat{b}(t,u)\cos l\omega_0 t dt,$$

$$\hat{B}_l^s(u) = \frac{2}{T}\int_0^T \hat{b}(t,u)\sin l\omega_0 t dt.$$

Зміщення цих оцінок визначаються відповідно правими частинами виразів (16)-(18), а їх дисперсії мають такий самий вигляд як вирази для дисперсій оцінок компонентів математичного сподівання з тією різницею, що кореляційні компоненти ПКВП в них замінені на коефіцієнти Фур'є кореляційної функції $b_\eta(t,u_1,u)$. Нульовий компонент $\widetilde{B}_0(u_1,u)$ залежить від усіх кореляційних компонентів ПКВП, тому вірогідність знаходження нульового кореляційного компонента ПКВП аж ніяк не може бути оцінена без знання вищих кореляційних компонентів. Цей факт підкреслює ту важливу роль, яку відіграє вибір моделі реального процесу при її статистичному аналізі.

3. ІМОВІРНІСНА СТРУКТУРА ЦИКЛІЧНИХ НАВАНТАЖЕНЬ

Провести прямі вимірювання напружень, що виникають під час циклічних навантажень, практично неможливо. Однак, припустивши, що елементи машини є пружними та ізотропними, можна прийняти, що між вібраційним сигналом $\zeta(t)$, що вимірюється в деякій зовнішній точці, і напруженням $\xi(t)$ у певному елементі всередині машини існує лінійна залежність $\xi(t) = \alpha\zeta(t)$. Коефіцієнт α враховує вплив характеристик матеріалу і системи, а також геометрії, що формується деякою точкою вимірювань [2].

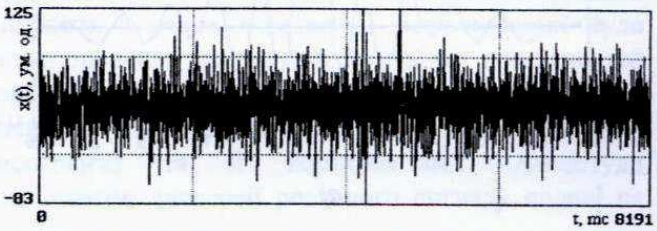
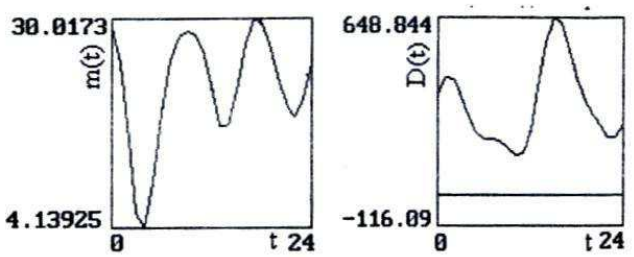


Рис.1. Сигнал осьової вібрації підшипника.

Застосувавши такий підхід, проаналізуємо структуру напружень, що виникають у підшипнику швидкохідного шпинделя машини для навивання вольфрамової спіралі електроламп, що працює в усталеному режимі (рис.1). Підшипниковий вузол має дефект нерівномірного радіального натягу. Період T оцінимо застосувавши компонентний метод оцінювання [3]. Оскільки оцінка періоду $T = 25{,}37$ не є цілим числом, то для оцінки імовірнісних характеристик когерентним методом необхідно змінити крок дискретизації. Для вирішення цієї задачі використаємо алгоритм інтерполяції поліномом Лагранжа 2-ого порядку [4].

Когерентні оцінки математичного сподівання і дисперсії, а також коефіцієнти Фур'є математичного сподівання подані на рис.2 та на рис.3.



а) б)

Рис.2. а) оцінка математичного сподівання;
б) оцінка дисперсії.

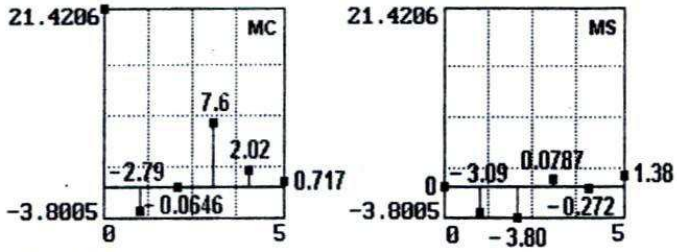
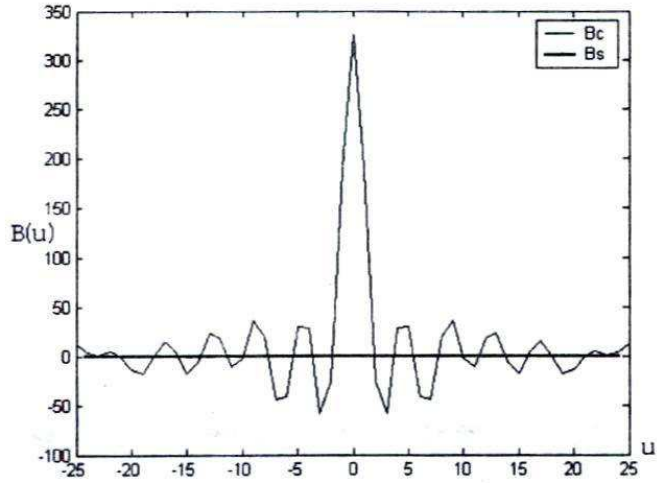
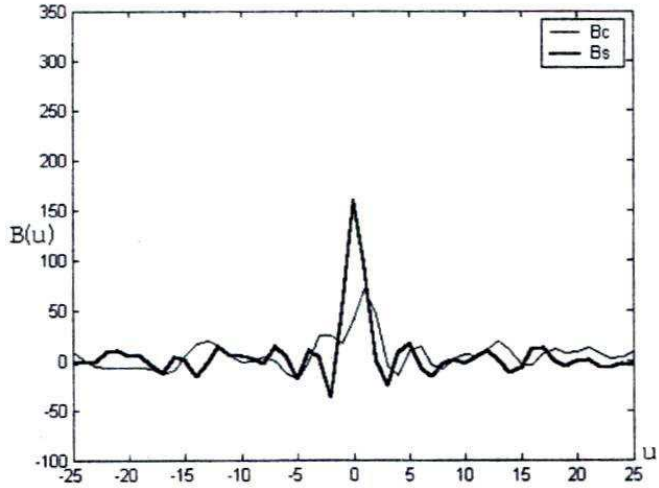


Рис.3. Компоненти математичного сподівання.



а)



б)

Рис.4. Когерентні оцінки кореляційних компонентів:
а) оцінка нульового кореляційного компонента;
б) оцінка другого кореляційного компонента.

На рис.4 показані графічні залежності від зсуву найбільш суттєвих за величиною кореляційних компонентів. Ці залежності можуть бути апроксимовані за допомогою формул:

$$B_0(u) = \frac{1}{2}\left(D_c e^{-\alpha_c|u|} + D_s e^{-\alpha_s|u|}\right) \cos\beta u \cos\omega_0 u\,,$$

$$B_2^c(u) = \frac{1}{2}\left(D_c e^{-\alpha_c|u|} - D_s e^{-\alpha_s|u|}\right) \cos\beta u \cos\omega_0 u + \\ + D_{cs} e^{-\alpha_{cs}|u|} \cos\beta u \sin\omega_0 u,$$

$$B_2^s(u) = \frac{1}{2}\left(D_s e^{-\alpha_s|u|} - D_c e^{-\alpha_c|u|}\right) \cos\beta u \sin\omega_0 u + \\ + D_{cs} e^{-\alpha_{cs}|u|} \cos\beta u \cos\omega_0 u.$$

З використанням цих формул та співвідношень (5)-(8), (11)-(13), (15)-(17) і (18)-(21) можуть бути обчислені конкретні значення показників якості оцінювання. Також може бути вирішена зворотня задача – вибрані такі параметри статистичної обробки, які забезпечують наперед задані показники якості оцінювання.

4. ВИСНОВКИ

Розглянуто та досліджено властивості когерентних оцінок ймовірнісних характеристик періодично корельованих випадкових прпоцесів.

Методом когерентного усереднення проаналізовано структуру напружень, що виникають у підшипниковому вузлі з нерівномірним радіальним натягом.

5. ЛІТЕРАТУРА

1. *Ксистрис, Санкар, Остигай. Оценка надежности машин по степени усталостного повреждения, накопленого вследствие случайных вибраций. // Конструирование и технология машиностроения. — 1978. — №4. — С.14-21.*
2. *Драган Я.П., Рожков В.А., Яворский И.Н. Методы вероятностного анализа ритмики океанологических процессов. — Л.:Гидрометеоиздат, 1987. — 320 с.*
3. *Михайлишин В.Ю., Яворский И.Н. Компонентное оценивание периода стохастической повторяемости ритмических сигналов // Автоматика. — 1993. — №4. — С. 44-48.*
4. *Юзефович Роман. Зміна кроку дискретизації ритмічних сигналів вібрації підшипника кочення // Матеріали відкритої науково-технічної конференції молодих науковців і спеціалістів Фізико-механічного інституту ім.Г.В.Карпенка НАН України. — Львів, 2002. — С.118-120.*

ВЛАСТИВОСТІ КОМПОНЕНТІВ ДИСПЕРСІЇ ІМОВІРНІСНОЇ МОДЕЛІ СИГНАЛУ ЕЛЕКТРОМАГНІТНОГО ЗБУРЕННЯ РЕЙОК ЗАЛІЗНИЧНОЇ КОЛІЇ

Ігор Ісаєв[1] Георгій Трохим[1], Ігор Яворський[1,2]

[1]Фізико-механічний інститут ім. Г. В. Карпенка НАН України,
[2]Інститут Телекомунікації, Технічно-сільськогосподарська академія,
м. Бидгощ, Польща

79601, Україна, Львів, МСП вул. Наукова, 5
телефон: 8 (0322) 63-33-55
факс: 8 (0322) 64-94-27
електронна пошта: isayev@ipm.lviv.ua, george@ipm.lviv.ua, iavor@ipm.lviv.ua

В роботі досліджуються енергетичні властивості дефектоскопічного сигналу від залізничних рейок на основі імовірнісної моделі сигналу у вигляді періодично корельованого випадкового процесу. Аналізуються вирази для дисперсії та її компонентів розкладу в ряди Фур'є методами нестаціонарної статистичної обробки.

ВСТУП

Сигнал дефектоскопії формується е.р.с., наведеною в індукційному давачі при скануванні підмагніченої рейки. Значення її як функції шляху виводяться на екран монітора персонального комп'ютера для візуального аналізу оператором. Типовий сигнал-відгук подано на рис.1.

Його періодична структура відповідає відстані між центрами шпал. В роботах [1-2] авторами були запропоновані методи аналізу такого сигналу на базі моделі нестаціонарних періодично корельованих випадкових процесів (ПКВП) [3]. Перевага нестаціонарного підходу полягає в можливості опису характеру мінливості процесу не тільки в середньому, а й в межах періоду.

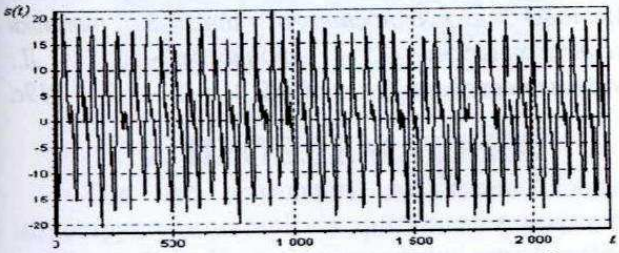


Рис. 1. Фрагмент осцилограми сигналу відгуку «бездефектної» рейки

1. СТОХАСТИЧНІ ВЛАСТИВОСТІ СИГНАЛУ

В роботі [1] було показано, що такий сигнал в ПКВП поданні містить 2 компоненти математичного сподівання та 4 компоненти кореляційної функції. Виходячи з задач швидкісної дефектоскопії було запропоновано замість використання властивостей кореляційної функції та її компонентів обмежитись лише їх значеннями при нульовому зсуві - дисперсією та її компонентами. Амплітуди компонентів дисперсії реального сигналу подані на рис.2.

При наявності в рейці дефекта росте кількість компонентів дисперсії (рис.3). Оскільки сигнали від різних дефектів різняться між собою, різного вигляду набуватимуть амплітуди компонент дисперсії сигналу.

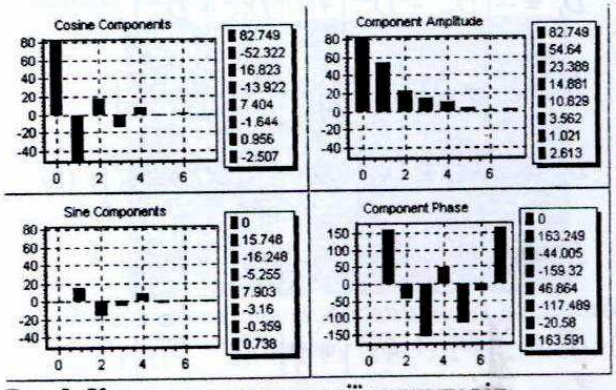


Рис.2. Компоненти дисперсії сигналу від бездефектної рейки.

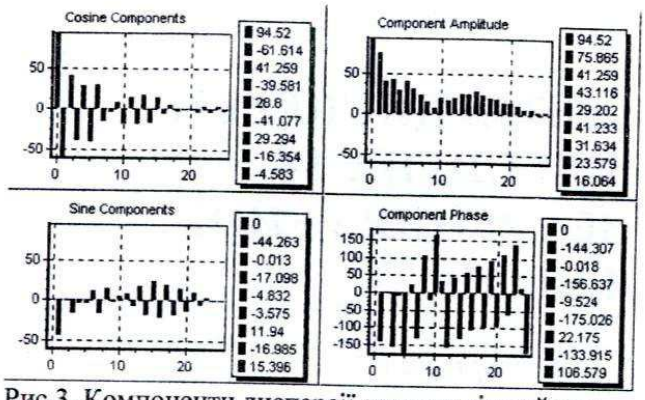


Рис.3. Компоненти дисперсії сигналу від рейки з двома дефектами

2. ІМОВІРНІСНА МОДЕЛЬ СИГНАЛУ

В роботі [2] на основі аналізу стохастичних властивостей сигналу та розкладу ПКВП на стаціонарні стаціонарно зв'язані процеси була побудована імовірнісна модель

$$\xi(t) = \eta_1(t)e^{i\omega_0 t} + \eta_2(t)e^{i2\omega_0 t}, \qquad (1)$$

де $\omega_0 = \frac{2\pi}{T}$, T - період корельованості, $\eta_1(t)$ та $\eta_2(t)$ - стаціонарні комплексні процеси. Подавши їх через парну та непарну частини $\eta_k(t) = \frac{1}{2}\left[\eta_k^c(t) - i\eta_k^s(t)\right]$, отримаємо

$$\begin{aligned}\xi(t) = {} & \eta_1^c(t)\cos\omega_0 t + \eta_1^s(t)\sin\omega_0 t + \\ & + \eta_2^c(t)\cos 2\omega_0 t + \eta_2^s(t)\sin 2\omega_0 t\end{aligned} \qquad (2)$$

Дисперсія $\xi(t)$ має вигляд:

$$\begin{aligned}D(t) = {} & D_0 + D_1^c\cos\omega_0 t + D_1^s\sin\omega_0 t + \\ & + D_2^c\cos 2\omega_0 t + D_2^s\sin 2\omega_0 t + D_3^c\cos 3\omega_0 t + \\ & + D_3^s\sin 3\omega_0 t + D_4^c\cos 4\omega_0 t + D_4^s\sin 4\omega_0 t,\end{aligned}$$

де

$$\begin{cases} D_0 = \frac{1}{2}\left[D_{11}^{cc} + D_{11}^{ss}\right] + \frac{1}{2}\left[D_{22}^{cc} + D_{22}^{ss}\right]; \\ D_1^c = \frac{1}{2}\left[D_{21}^{cc} + D_{21}^{ss}\right] + \frac{1}{2}\left[D_{12}^{cc} + D_{12}^{ss}\right]; \\ D_1^s = \frac{1}{2}\left[-D_{21}^{cs} + D_{21}^{sc}\right] + \frac{1}{2}\left[D_{12}^{cs} - D_{12}^{sc}\right]; \\ D_2^c = \frac{1}{2}\left[D_{11}^{cc} - D_{11}^{ss}\right]; D_2^s = \frac{1}{2}\left[D_{11}^{cs} + D_{11}^{sc}\right]; \\ D_3^c = \frac{1}{2}\left[D_{21}^{cc} - D_{21}^{ss}\right] + \frac{1}{2}\left[D_{12}^{cc} - D_{12}^{ss}\right]; \\ D_3^s = \frac{1}{2}\left[D_{21}^{cs} + D_{21}^{sc}\right] + \frac{1}{2}\left[D_{12}^{cs} - D_{12}^{sc}\right]; \\ D_4^c = \frac{1}{2}\left[D_{22}^{cc} - D_{22}^{ss}\right]; D_4^s = \frac{1}{2}\left[D_{22}^{cs} + D_{22}^{sc}\right]. \end{cases} \qquad (3)$$

Значення $D_{l,k}^{c,s}$ є значеннями дисперсій взаємнокореляційних функцій формуючих ПКВП стаціонарних процесів

$$D_{l,k}(u) = \frac{1}{4}\left[D_{l,k}^{cc}(u) + D_{l,k}^{ss}(u) + i\left[D_{l,k}^{cs}(u) - D_{l,k}^{sc}(u)\right]\right]$$

$$D_{1,1}(u) = \frac{1}{4}\left[D_{1,1}^{cc}(u) + D_{1,1}^{ss}(u)\right]$$

$$D_{1,2}(u) = \frac{1}{4}\left[D_{1,2}^{cc}(u) + D_{1,2}^{ss}(u) + i\left[D_{1,2}^{cs}(u) - D_{1,2}^{sc}(u)\right]\right]$$

$$D_{2,1}(u) = \frac{1}{4}\left[D_{2,1}^{cc}(u) + D_{2,1}^{ss}(u) + i\left[D_{2,1}^{cs}(u) - D_{2,1}^{sc}(u)\right]\right]$$

$$D_{2,2}(u) = \frac{1}{4}\left[D_{2,2}^{cc}(u) + D_{2,2}^{ss}(u)\right]$$

Оскільки значення дисперсії є значеннями відповідних кореляційних функцій при нульовому зсуві і для взаємокореляційних функцій $D_{1,2}(0) = D_{2,1}(0)$, то звідси маємо:

$$D_{1,2}^{cc}(0) = D_{2,1}^{cc}(0) = D_{2,1}^{cc};\ D_{1,2}^{ss}(0) = D_{2,1}^{ss}(0) = D_{2,1}^{ss}$$

$$D_{1,2}^{cs}(0) = D_{2,1}^{cs}(0) = D_{2,1}^{cs};\ D_{1,2}^{sc}(0) = D_{2,1}^{sc}(0) = D_{2,1}^{sc}$$

ВИСНОВКИ

Досліджуючи методами ПКВП тонку структуру рейок з дефектами і порівнюючи результати з результатами для "бездефектної" моделі можна розробити методологію пошуку, локалізації і розпізнавання дефектів при швидкісній діагностиці стану рейок..

ЛІТЕРАТУРА

1. *Ісаєв І.Ю., Трохим Г.Р. Властивості характеристик сигналів електромагнітного збурення бездефектних залізничних рейок // Матеріали відкритої науково-технічної конференції молодих науковців і спеціалістів «Інженерія поверхні», 22-23 березня 2001 р., С.37-40.*
2. *Ісаєв І.Ю., Трохим Г.Р Стохастична структура імовірнісної моделі сигналу електромагнітного збурення "бездефектної" залізничної рейки // Матеріали відкритої науково-технічної конференції молодих науковців і спеціалістів «КМН-2002», Львів 2002, С.114-117.*
3. *Драган Я.П., Рожков В.А., Яворский И.Н. "Методы вероятностного анализа ритмики океанологических процессов".- Л.: Гидрометеоиздат, 1989. - 319с.*

СЕГМЕНТАЦІЯ НАПІВТОНОВИХ ЗОБРАЖЕНЬ МІКРОСТРУКТУРИ МАТЕРІАЛІВ ДЛЯ ВИЗНАЧЕННЯ ДИНАМІКИ РОСТУ КАВІТАЦІЙНИХ ТРІЩИН

Р.Я. Косаревич, Б.П. Русин

Фізико-механічний інститут ім. Г.В. Карпенка НАН України,
79601, м. Львів, вул. Наукова, 5, тел.:(0322) 65-45-30, e-mail: kosarevych@ipm.lviv.ua

Анотація У роботі розглядається проблема сегментації напівтонових зображень поверхні фрагмента матеріалу, пошкодженої кавітаційними тріщинами з метою визначення їх кількості та розмірів. Спочатку на зображенні вибираються точки, які з високою вірогідністю належать тріщинам, далі на основі запропонованого критерію до цих точок приєднуються інші точки зображення. На сегментованому зображенні проводиться фільтрація лінійних сегментів, які відповідають шліфам матеріалу.

1. ВСТУП

Багато із виникаючих проблем в матеріалознавстві сьогодні можуть бути вирішені лише з використанням автоматизованих засобів накопичення, переробки і видачі інформації про матеріали. Основним об'єктом в структурних методах дослідження матеріалів є зображення, отримані мікроскопом, на яких достатньо добре розпізнаються конструктивні елементи матеріалу. Так при визначенні міцності матеріалів важливе значення має обчислення на ділянці матеріалу кількості і розмірів кавітаційних тріщин (Рис.1). Загалом ці визначення проводяться на основі візуального аналізу фотографій. Якщо автоматизувати ці дослідження, тобто проводити обробку цифрових зображень дослідних зразків, то без сумніву вдасться зменшити вплив суб'єктивних чинників і суттєво скоротити часові затрати.

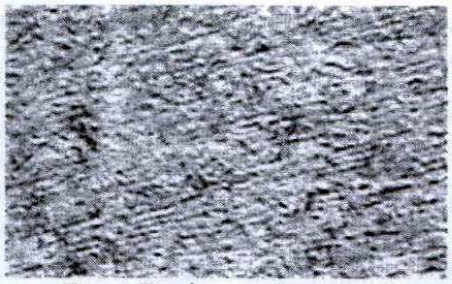


Рис.1 Вихідне зображення

2. ПРОБЛЕМАТИКА

Під час візуальних досліджень зображень мікроструктури матеріалів критичною є яскравість їх елементів. Чим більше вони темніші тим ймовірніше, що вони відображають тріщину, при цьому величина цих деталей не відіграє особливого значення, оскільки тріщина може бути дуже малою. Тріщина є кавітаційною, тобто внаслідок мікрогідроудару і може з'явитись як вздовж так і поперек текстури зображення. Сама текстура і шліфи, що утворюються внаслідок обробки зразка відіграють роль шумів оскільки спотворюють ділянки зображення, де розташована тріщина. Для аналізу зразка матеріалу необхідно провести обчислення довжин тріщин декілька раз протягом часу досліджень і встановити тим самим динаміку їх росту.

Для вирішення даної задачі необхідно сегментувати цифрове зображення, тобто розділити на фон і власне тріщини. Як відомо проблема сегментації зображення може роз'язуватись різними підходами, а саме за допомогою встановлення порогу, виділення контурів, нарощування областей та ін.

3. СЕГМЕНТАЦІЯ

Методи вибору порогу [1,2] базуються на припущені, що суміжні точки зображення, чий рівень яскравості лежить в заданих межах, належать до одного класу. За допомогою цих методів можна отримати добрі результати сегментації зображень, які включають дві компоненти. Оскільки такі методи повністю ігнорують інформацію про просторові зв'язки зображення, вони є неефективними для зображень із розмитими краями об'єктів, а також для зображень, які потрібно сегментувати на більшу кількість компонент. В нашому випадку встановлення порогу призводить до виділення великої кількості хибних об'єктів (Рис.2).

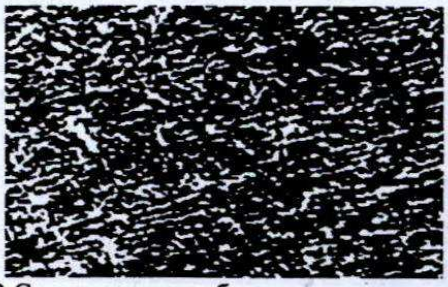


Рис.2 Сегментоване зображення за допомогою вибору порогу

Виділення контурів на зображенні грунтується на стрімкій зміні значення яскравості на межі між двома областями [3,4,5]. Детектори

країв, що використовуються при цьому або прості оператори типу Собеля, Робертса, або більш складніні, такі як Канні оператор. Але результатом детекторів краю, в більшості випадків, не є чітка замкнена лінія, а деяка сукупність точок, яка містить розриви і надлишкові точки (Рис.3). Тому після застосування таких детекторів необхідна додаткова обробка, яка включає прослідковування контурів, заповнення розривів, потоншення ліній, що вимагає додаткових затрат часу. Цих часових затрат можна уникнути якщо інтегрувати підходи на основі виділення контурів і нарощування областей.

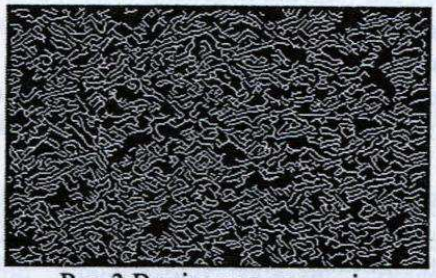


Рис.3 Виділення контурів

Методи, що полягають у нарощуванні областей [6,7,8] використовують припущення, що суміжні піксели, які належать одній області мають подібні візуальні характеристики такі як рівень напівтону, колір, текстуру. Ефективність цього підходу сильно залежить від вибраного критерію однорідності. На відміну від вибору параметрів однорідності техніка нарощування початкових областей (НПО) обмежується кількістю цих областей. Задавши їх кількість методи НПО намагаються знайти таку сегментацію зображення на області, при якій кожна звязна компонента області містить одну з початкових областей.

Вибір того чи іншого підходу до сегментації зображення диктується апріорною інформацією про об'єкти дослідження. Серед згаданих вище підходів методи НПО можуть дати найбільш задовільні результати, оскільки вихідні зображення є малоконтрасними і необхідно проводити їх локальний аналіз.

Як вже вказувалось методи НПО полягають у виборі наперед заданої кількості, початкових областей $S_1,...,S_n$, які можуть бути точками і приєднання до цих початкових областей інших точок зображення на основі деякого критерію. Кількість початкових областей може визначатися, наприклад, кількістю об'єктів, які потрібно виділити на зображенні.

3.1 Вибір початкових точок

В нашому випадку, коли кількість тріщин на зображенні наперед не відома, початкові точки будуть вибиратись, виходячи із значень їх яскравостей, за декілька етапів. Спочатку, оскільки відомо, що тріщинам відповідають темніші ділянки зображення, вибираються точки рівень яскравості яких менший за середній рівень зображення і в околі розміром 3×3 яких всі точки мають більшу або таку ж яскравість. На основі середнього рівня яскравості, відібраних на попередньому кроці точок, відкидаються ті, значення яких більше. Крім того для відбору точок використовується різниця значень середніх яскравостей двох різних за розмірами околів O_1та O_2, центри яких розташовані в одній і тій ж точці (Рис.4). Якщо різниця цих значень менша за деякий поріг, то ця точка відкидається.За поріг вибирається значення дисперсії елементів меншого околу O_2. Критерій відбору виражається як $A(i,j) \in S \Leftrightarrow C(O_1(A(i,j)), O_2(A(i,j))) > \sigma(O_2(A(i,j)))$, де $A(i,j)$ – точка відібрана на першому етапі, $S = \bigcup_{i=1}^{n} S_i$, $\sigma(\cdot)$ - середньоквадратичне відхилення елементів околу.

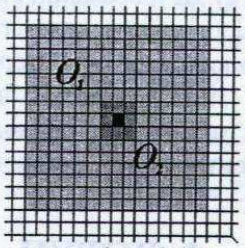


Рис.4 Околи аналізу відібраних точок

Діапазон значень величин тріщин є наперед відомим, тому розмір околів O_1 та O_2 вибирається так щоб розмір більшого був співмірним з величиною тріщини. Використання такого критерію можна пояснити наступним чином. Якщо центр обох околів O_1 та O_2 припадає на область, де розташована тріщина, то різниця середніх рівнів яскравості буде більшою за рахунок того, що в квадратному околі O_1 більша частина пікселів буде мати яскравість вищу ніж яскравість елементів тріщини. В той же час в околі O_2 всі елементи будуть належати тріщині і середнє значення буде малим. Якщо ж центр околів припадає на область, в якій немає тріщини, то середні значення яскравості в обох околах будуть приблизно однаковими і їх різниця буде меншою за поріг. Значення порогу є змінним оскільки рівень яскравості елементів не є сталим для різних тріщин.

3.2. Критерій приєднання точок

Кожній з початкових точок присвоюється деке число, мітка – її порядковий номер. На рис.5 для тестового зображення показано вибір початкових точок.

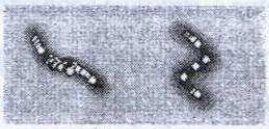


Рис. 5 Приклад виділення початкових точок

Після вибору початкових точок необхідно задати критерій приєднання до них інших точок зображення. Запропонований нами критерій також

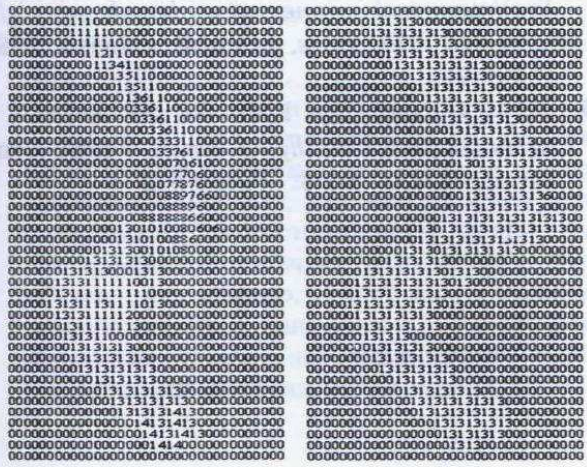
базується на використанні двох околів різних розмірів. За цим критерієм кожній точці *B(i,j)* з 3×3 околу початкової точки *A(i,j)* присвоюється її мітка, якщо різниця значень їх яскравостей менша ніж різниця середніх значень яскравостей двох околів з центром у цій точці

$$m(B(i,j)) = m(A(i,j)) \Leftrightarrow$$
$$\Leftrightarrow \left|br(B(i,j)) - br(A(i,j))\right| < C(O_3(B(i,j)), O_4(B(i,j)))$$

, де *m(B(i,j))* – мітка точки *B(i,j)*, *br(B(i,j))* – яскравість в точці *B(i,j)*. В цьому випадку розміри околів O_3 та O_4 вибираються так, щоб не перевищували розмір тріщини. Такий критерій дозволяє ефективно відбирати точки оскільки для тих, що поблизу початкової і належать тріщині різниця яскравостей невелика так як і для середніх значень в околах O_3 та O_4, а для точок які розташовані на краях тріщини різниця їх яскравостей із яскравістю початкової точки зростає але в той же час зростає і різниця середніх значень, оскільки околи містять більше точок із вищим значенням яскравості. У випадку коли точка не належить тріщині різниця яскравостей буде більшою за різницю середніх рівнів в околах, оскільки за межами тріщини ці рівні будуть приблизно однаковими для околів O_3 та O_4.

3.3. Формування зв'язних множин міток

Результатом двох вищеописаних етапів сегментації є присвоєння окремим точкам зображення міток. Можливий випадок, як видно на рис.5, коли для одного об'єкта визначається декілька початкових точок, що мають різні мітки. Тому наступним кроком є присвоєння всім точкам, що належать одному об'єкту одакової мітки. Ми виходимо з того, що тріщина - об'єкт, для елементів якого властива невелика дисперсія яскравостей (критерій НПО). Тому навіть у випадку, коли для одного об'єкта є декілька початкових точок, за вибраним критерієм приєднані точки мають формувати звязну множину. Отже вибираючи по черзі початкові точки об'єкта, і перенумеровуючи для всіх приєднаних до неї їх 4-ох або 8-ми суміжних сусідів, ми всім точкам об'єкта присвоємо однакову мітку (Рис. 6).



а) б)

Рис. 6 Присвоєння міток: (а) початкові мітки, (б) результат перенумерації за критерієм зв'язності

3.4. Фільтрація лінійних сегментів

Присвоєння однакових міток елементам на основі їх зв'язності дозволяє сформувати множини, які відповідають тріщинам на зображенні. Як вже вказувалось на зображенні присутні також текстура матеріалу і шліфи, як наслідок обробки матеріалу. Яскравість елементів зображення, що відповідають шліфам, дуже близька до яскравості тріщини, а своєю формою вони подібні до прямолінійних сегментів розташованих під одним кутом для цілого зображення. Завдяки своїм ознакам шліфи можуть спотворити загальну інформацію про кількість тріщин дослідного зразка, тому необхідне їх вилучення з вже сегментованого зображення.

Для цього можна скористатись максимальною відстанню між точками множини і прямою, що з'єднює дві найбільш віддалені точки множини як критерієм фільтрації (Рис. 7) [9]. Якщо максимальна відстань не перевищує заданий поріг, то вважається, що ця множина лінійний сегмент. Вибір правильного порогу в цьому випадку є важливим, бо як легко зауважити з рис.8 лише за таким критерієм можна відкинути значну кількість об'єктів, які хоч і мають форму близьку до лінійного сегмента, але не є шліфами.

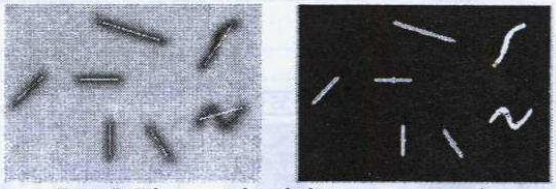
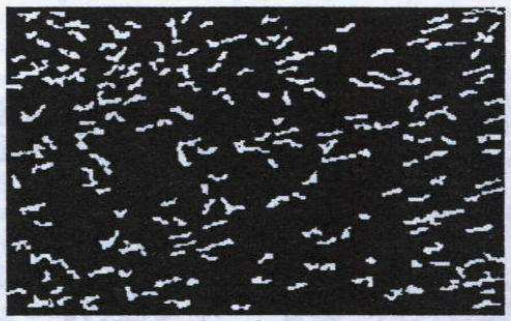
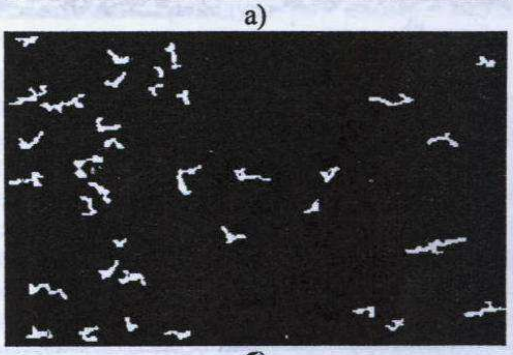


Рис.7 Фільтрація лінійних сегментів



а)



б)

Рис.8. Фільтрація лінійних сегментів з використанням різних порогів:
а) поріг рівний 1, б) поріг рівний 3.

Тому у цьому випадку слід враховувати також відомості про орієнтацію шліфів на зображенні.

4. ЕКСПЕРИМЕНТАЛЬНІ РЕЗУЛЬТАТИ

Як вже вказувалось на початку, сегментація зображення кавітаційних тріщин проводиться з метою визначення динаміки росту. На сегментованому зображенні підраховувались кількості кавітаційних тріщин (ККТ) трьох діапазонів довжин: 0-3*10^{-6}м., 3-5*10^{-6}м. та 5-10*10^{-6}м. Довжина тріщини апроксимувалась величиною, що є сумою відстані між двома найбільш віддаленими точками множини та максимальною серед відстаней точок множини до цієї прямої. На рис.9в і рис. 9е показано результати підрахунків ККТ для одного зразка з різним часом експозиції рис.9а та рис.9б. Отримані результати добре узгоджуються з реальними спостереженнями.

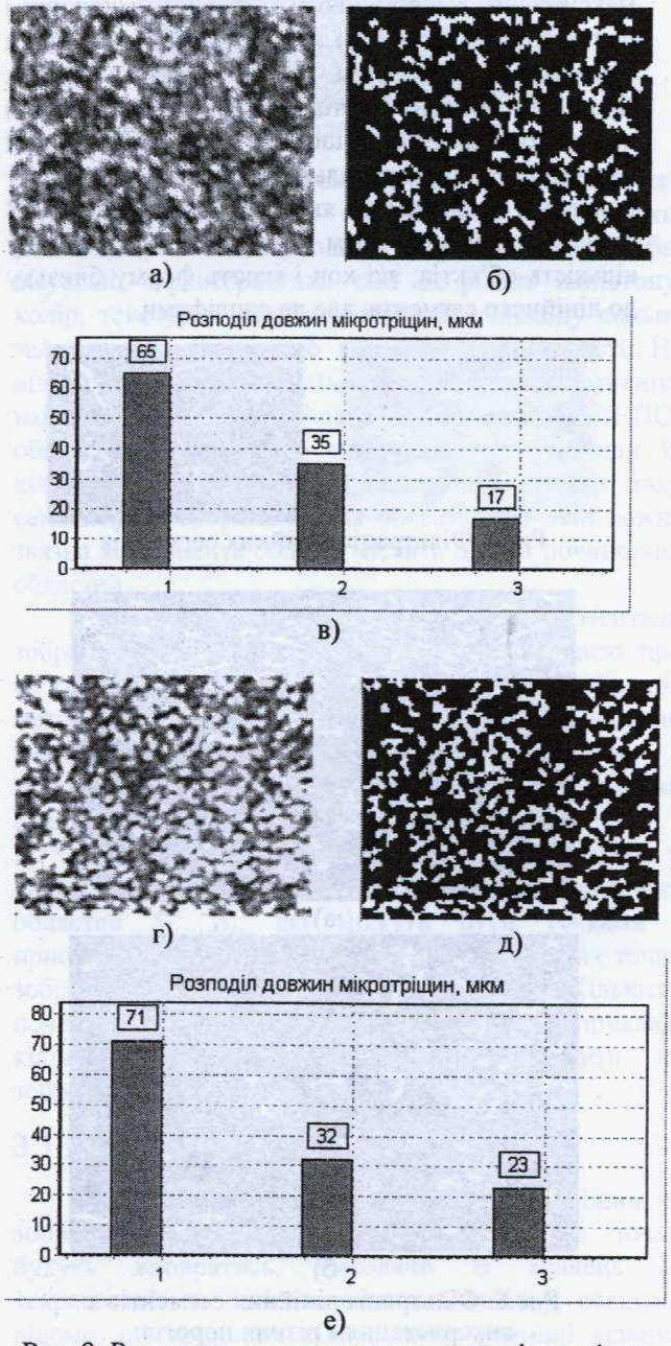


Рис. 9 Результати визначення довжин мікротріщин

ВИСНОВКИ

Адаптація відомих та розробка нових методів обробки зображень дозволяють покращити достовірність та скоротити часові затрати при аналізі просторово - часової структури матеріалів.

Запропоновані критерії виділення початкових областей та приєднання суміжних точок дозволяють ефективно проводити виділення кавітаційних тріщин на напівтонових зображеннях мікроструктури матеріалів.

ЛІТЕРАТУРА

1. H.D. Cheng, C.H. Chen, H.H. Chiu and Huijuan Xu *Fuzzy homogeneity approach to multilevel thresholding* //IEEE Trans. on Image Processing. - No 7. - 1998. - P. 1084-1088.
2. Francis H.Y. Chan, F.K. Lam, and Hui Zhu *Adaptive thresholding by variation method* ///IEEE Trans. on Image Processing. - No 7. - 1998. - P. 468-474.
3. Chao Han, Thomas S. Hatsukami, Jeng-Neng Hwang *A fast minimal path active contour model* // IEEE Trans. on Image Processing. - No.6. - 2001. - P. 865-872.
4. Andrew P. *Paplinski Directional filtering in edge detection* // IEEE Trans. on Image Processing. - No. 4. - 1998. - P. 611-615.
5. S. Ando *Image field categorization and edge/corner detection from gradient covariance* / IEEE Trans. on PAMI. - No.2. - 2000. - P. 179-190.
6. J. Fan, K.Y. Yau, Ahmed K. Elmagarmid Walid G. Aref *Automatic image segmentation by integrating color-edge extraction and seeded region growing* // IEEE Trans. on Image Processing. - No.10. - 2001. - P. 1454-1466.
7. Andre Bleau, L. J. Leon *Watershed-based segmentation and region merging* //Comp. Vision and Imag. Underst. - 2000. - P. 317-370.
8. F. Moscheni, S. Bhattacharjee, Murat Kunt *Spatiotemporal segmentation based on region merging*// IEEE Trans. on PAMI. - No.9. - 1998. - P. 897-915.
9. Teh C.H. and Chin R.T., *On the detection of dominant points on digital curves* // IEEE Trans. PAMI. - 1989. - No. 8. - P. 859-872

СТИСНЕННЯ ДАНИХ ПРИ ФОРМУВАННІ ПРОСТОРУ ДІАГНОСТІЧНІХ ОЗНАК НА ОСНОВІ НЕЛІНІЙНИХ НЕПАРАМЕТРИЧНИХ МОДЕЛЕЙ ОБ'ЄКТІВ КОНТРОЛЮ

В.Д.Павленко, О.О.Фомін
Одеський національний політехнічний університет
65044, пр. Шевченко 1, Одеса, Україна
тел. (048) 746 72 29, e-mail: pavlenko_vitalij@mail.ru

Пропонується метод модельної діагностики, заснований на непараметричній ідентифікації об'єктів контролю з використанням інтегро-ступневих рядів Вольтерра. Досліджується за допомогою методів статистичної класифікації інформативність діагностичних параметрів, що формуються на основі таких моделей: евристичних ознак, моментів багатомірний вагових функцій, а також ознак, виділених з використанням методу стиснення, заснованого на Wavelet-перетворенні.

ВСТУП

В цей час в технічній діагностиці розвивається напрям, заснований на відновленні моделі (оператора) об'єкта, що діагностується [1]. Звичайно передбачається, що несправності змінюють тільки параметри моделі об'єкта, які при діагностуванні оцінюються методами параметричної ідентифікації. Однак часто, наприклад, при виробництві виробів електронної техніки, більшість дефектів приводить до зміни не тільки параметрів моделі об'єкта, але і її структури, що обумовлює застосування методів непараметричної ідентифікації для побудови математичної моделі об'єкта контролю на основі даних експеримента "вхід - вихід".

У модельній діагностиці процедура діагностування виконується в два етапи. На першому отримують початкову (первинну) інформацію про об'єкт у вигляді сигналів відгуків на пробні впливи. На другому етапі ця інформація обробляється для виділення діагностичних ознак і рішення про належність даного об'єкта до певного класу за фізичним (технічним) станом. Для побудови діагностуючого правила використовують методи статистичної класифікації (розпізнавання образів) [2].

Існуючі методики модельної діагностики, засновані на використанні динамічних характеристик, обмежуються тільки лінійними моделями, а методики, засновані на обліку ефектів нелінійності, використовують інформацію тільки про властивості статичних характеристик. Реальні об'єкти, як правило, одночасно володіють і нелінійними і динамічними властивостями. Формально будь-який нелінійний динамічний об'єкт може бути описаний рядом Вольтерра [2, 3].

У даній роботі пропонується метод модельної діагностики, заснований на непараметричній ідентифікації об'єктів контролю з використанням рядів Вольтерра, досліджується за допомогою методів статистичної класифікації інформативність діагностичних параметрів, що формуються на основі таких моделей.

1. НЕЛІНІЙНІ НЕПАРАМЕТРИЧНІ ДИНАМІЧНІ МОДЕЛІ

В якості інформативного опису об'єктів контролю пропонується використовувати нелінійні непараметричні динамічні моделі у вигляді рядів Вольтерра

$$y(t)=\sum_{n=1}^{\infty}\int_0^t\ldots\int_0^t w_n(\tau_1,\tau_2,\ldots,\tau_n)\prod_{i=1}^{n}x(t-\tau_i)d\tau_i \quad (1)$$

де $w_1(\tau), w_2(\tau_1,\tau_2),\ldots,w_n(\tau_1,\tau_2,\ldots,\tau_n)$ - багатомірні вагові функції або ядра Вольтерра (ЯВ), *x(t)* - вхідний вплив, а *y(t)* - відгук об'єкта при нульових початкових умовах.

При цьому сукупність багатомірних ЯВ повністю характеризує і нелінійні і динамічні властивості, отже, технічний стан об'єктів контролю. Застосування моделей на основі рядів Вольтерра дозволяє більш повно і точно врахувати нелінійні і інерційні властивості об'єктів контролю, робити процедуру модельної діагностики більш універсальною, підвищує надійність діагнозу.

Діагностична процедура в цьому випадку зводиться до визначення ЯВ за даними експеримента "вхід - вихід" [3] і побудові на основі отриманих ядер діагностичної системи ознак, в просторі яких будується вирішальне (діагностичне) правило оптимальної класифікації [4].

Необхідно зазначити, що в задачах модельної діагностики адекватність моделі об'єкту контролю потрібно розуміти не в значенні точності опису відгуку об'єкта, а в значенні інформативності її з точки зору достовірного (надійного) розпізнавання технічного стану. Тому у разі застосування ЯВ при формуванні вхідного опису нелінійних динамічних

об'єктів контролю в діагностичних дослідженнях необхідно забезпечити насамперед високу точність оцінки перетинів багатомірний ЯВ малих порядків, що часто на практиці виявляється достатнім для побудови ефективної системи розпізнавання.

Застосування методів теорії розпізнавання образів для рішення задач технічної діагностики з використанням непараметричних динамічних моделей об'єктів контролю у вигляді рядів Вольтерра засноване на наступних передумовах:

1. Існує об'єктивний (але неявний) зв'язок між багатомірними ЯВ, що характеризують структуру об'єкта контролю і його технічним станом, іншими словами, існує деяка функція $F(H, S)$, зв'язуюча стан S з ЯВ $H=\{h_n(\tau_1,\ldots,\tau_n)\}^N_{n=1}$.

2. Функція $F(H, S)$, відновлена на основі ЯВ досліджених об'єктів, може бути екстрапольована на об'єкти з невідомими властивостями.

3. Структура об'єкта контролю може бути адекватно представлена за допомогою ЯВ.

Можливі різні підходи до рішення задач технічної діагностики, відмінні засобами вибору інформативних ознак і алгоритмами відновлення функції $F(H, S)$ [2,4].

Ефективність застосування методів розпізнавання образів для діагностування в основному залежить від інформативності сукупності параметрів, що використовується. Якщо вибрані параметри, що досить повно характеризують внутрішню структуру об'єкта діагностування, то основна маса об'єктів, будучи ідентичною по структурі, відобразиться в просторі цих параметрів у вигляді щільної множини точок. Об'єктам з особливостями структури (дефектним) будуть відповідати точки, що відхиляються від цієї щільної множини і розташовані значно рідше в зв'язку з різноманітністю дефектів у таких об'єктів і їх відносної нечисленності (якщо діагностуються високонадійні прилади, наприклад, інтегральні мікросхеми).

2. ФОРМУВАННЯ ПРОСТОРУ ОЗНАК

Вибір сукупності діагностичних ознак (задача формування простору інформативних ознак) впливає вирішальним чином на точність розпізнавання технічного стану об'єктів діагностування.

Математично задача вибору діагностичних ознак формулюється таким чином. Нехай заданий початковий простір ознак X, розмірністю p. Необхідно знайти перетворений простір Y, елементами якого є q - мірні вектори, притому $q<p$. Формально така задача полягає у визначенні відображення $A: X \rightarrow Y$, яке початковому простору X ставить у відповідність шуканий простір Y. Рішення цієї задачі може бути досягнуто двома шляхами.

Перший з них заснований на зважуванні тим або іншим засобом різних систем ознак з метою оцінки їх інформативності (корисність) при розпізнаванні. У даній роботі, ефективність вибраного набору ознак оцінювалась за результатами рішення задачі класифікації об'єктів екзаменаційної вибірки за допомогою побудованого одним з алгоритмів навчання вирішального правила. Набори ознак, для яких вірогідність правильного розпізнавання мала, відкидаються і як діагностична система ознак вибирається набір, для якого додання будь-якої нової ознаки не збільшує або збільшує не набагато його інформативність. Якщо параметри статистично незалежні, то діагностичну систему ознак можна сформувати шляхом оцінки інформативності кожного параметра і відкидання деякої кількості найменш інформативних з них. При такій процедурі формування ознакового простору значення самих ознак не змінюється, а тільки зменшується їх кількість.

Інший шлях вирішення задачі стиснення діагностичної інформації при розпізнаванні станів об'єктів контролю полягає в відшукуванні оптимального в деякому розумінні перетворення початкового простору векторів вимірювань в простір зображень меншої розмірності. Процедура розпізнавання станів об'єкта в редуцированому просторі ознак виявляється більш простою, оскільки обсяг діагностичної інформації, що обробляється зменшується. Зменшення розмірності ознакового простору дозволяє використати більш складні нелінійні вирішальні правила, що підвищують якість розпізнавання. Однак на відміну від описаного вище методу вибору діагностичних ознак цей метод не передбачає скорочення кількості вимірювань. У цьому випадку нові ознаки виявляються відірваними від конкретного фізичного змісту і мають тільки абстрактне інформаційне значення.

Для опису об'єкта контролю на вході системи, що розпізнає у вигляді вектора первинних ознак $\bar{x}$ можна використати або кінцевий набір значень відліків перетинів багатомірних ЯВ, або коефіцієнти розкладання ЯВ за деякою системою функцій. У ряді випадків можна сформувати деякі евристичні характеристики, які можуть входити як компоненти у вектор ознак. В якості таких характеристик можна вибрати максимальне по модулю значення і відповідний йому час, площу головного перетину ЯВ, крутість його в точці t=0, час перехідного процесу. Застосування таких евристичних ознак часто дозволяє значно скоротити розмірність вектора ознак початкового опису.

Більш універсальний підхід при формуванні вектора інформативних параметрів складається у використанні так званих моментів ЯВ об'єктів контролю, які обчислюються за формулою

$$\mu^n_{ij..k} = \int_0^\infty \ldots \int_0^\infty \tau_1^i \tau_2^j \ldots \tau_n^k w_n(\tau_1, \tau_2, \ldots, \tau_n) d\tau_1 d\tau_2 \ldots d\tau_n \quad (2)$$

тут $i,j,\ldots,k=0,1,\ldots,\infty$.

Представляє також інтерес використання методу стиснення діагностичної інформації за допомогою перетворення сплесків (Wavelet-перетворення) [5].

3. ДОСЛІДЖЕННЯ ІНФОРМАТИВНОСТІ ОЗНАК

Для дослідження засобів стиснення даних при формуванні простору діагностичних ознак був вибраний об'єкт, який описується нелінійним диференціальним рівнянням вигляду

$$\frac{dy(t)}{dt} + \alpha \cdot y(t) + \beta \cdot y^2(t) = x(t), \quad (3)$$

де α и β - постійні коефіцієнти (параметри), недоступні для вимірювань. Для такого об'єкта модель у вигляді двох членів ряду Вольтерра при нульових початкових умовах має вигляд

$$y(t) = \int_0^t w_1(\tau_1)x(t-\tau_1)d\tau_1 + \int_0^t\int_0^t w_2(\tau_1,\tau_2)x(t-\tau_1)x(t-\tau_2)d\tau_1 d\tau_2, \quad (4)$$

тут *x(t) и y(t)* – сигнали, що вимірюються відповідно на вході і виході об'єкта контролю;

$$w_1(\tau_1) = e^{-\alpha\tau_1},$$
$$w_2(\tau_1,\tau_2) = \frac{\beta}{\alpha}(e^{-\alpha\tau_1}e^{-\alpha\tau_2} - e^{-\alpha\tau_2}), \tau_1 \le \tau_2. \quad (5)$$

Звідси видно, що для діагностики станів такого об'єкта по параметрах α и β цілком досить використати ЯВ другого порядку при формуванні вектора інформативних параметрів-ознак. Як початковий опис об'єкта контролю тут використовується головний перетин ЯВ при $\tau_1 = \tau_2 = t$

$$w_2(t,t) = \frac{\beta}{\alpha}(e^{-2\alpha t} - e^{-\alpha t}). \quad (6)$$

Побудова моделі об'єкта (6) за даними експеримента "вхід - вихід" здійснюється одним з методів активної ідентифікації [1,3].

Ефективність виділених діагностичних ознак оцінюється за результатами рішення задачі класифікації об'єктів повчальної вибірки за допомогою побудованого на основі методу статистичних рішень діагностичного правила.

Ефективність виділених діагностичних ознак оцінювалася за результатами рішення задачі класифікації об'єктів екзаменаційної вибірки за допомогою побудованого на основі даних повчальної вибірки методом статистичних рішень діагностичного правила [2,4]. Для оцінки достовірності діагностування використовувався відсоток правильного розпізнавання (ВПР):

$$P = \sum_{i=1}^{m} l_i \cdot \left(\sum_{i=1}^{m} L_i\right)^{-1} \cdot 100 \quad (7)$$

де l_i – кількість правильно класифікованих елементів і-го класу в екзаменаційній вибірці; *Li* - кількість елементів і-го класу в екзаменаційній вибірці; *m* - кількість класів.

Реалізації об'єктів з різними значеннями параметрів α и β генерувалися коштами імітаційного моделювання в середовищі MATLAB. Повчальна і екзаменаційна вибірки формувалися на основі дискретних значень функцій (6) для об'єктів чотирьох класів: 1- й клас складають об'єкти умовно прецездатні (параметри α и β змінюються в межах ±5% від номінальних значень $\alpha_н$ и $\beta_н$); 2-й клас - об'єкти умовно непрацездатні по параметру α (параметр β змінюється в межах ±5%$\beta_н$, а параметр $\alpha \in (-1.1\alpha_н; -1.05\alpha_н) \cup (1.05\alpha_н; 1.1\alpha_н)$); 3-й клас - об'єкти умовно непрацездатні по параметру β (параметр α змінюється в межах ±5%$\alpha_н$, а параметр $\beta \in (-1.1\beta_н; -1.05\beta_н) \cup (1.05\beta_н; 1.1\beta_н)$); 4-й клас - непрацездатні об'єкти по α та β одночасно.

Таким чином для розділення цих класів знаходилися три вирішальних функції *$d_1(x)$, $d_2(x)$, $d_3(x)$*. Вирішальна функція *$d_1(x)$* відділяє об'єкти 1-го класу від 2-го, 3-го і 4-го; *$d_2(x)$* - відділяє *об'єкти* 2-го класу від 3-го і 4-го; *$d_3(x)$* -розділяє *об'єкти* 3-го і 4-го класів.

Досліджувалася інформативність наступних евристичних ознак, виділених на основі функцій (6): 1- экстремум функції, 2-час экстремума; 3-похідна функції в точці *t*=0, 4-інтеграл від функції, 5-час перехідного процесу. Найкращі результати досліджень інформативності окремих ознак і їх сукупностей у вигляді отриманих значень ВПР для трьох вирішальних правил приведені в таблиці 1.

Таблиця 1.

Ознаки	$d_1(x)$	$d_2(x)$	$d_3(x)$
4	73.03	98.91	48.69
1,4	96.73	99.78	99.67
1,2,4	96.24	99.12	99.67

Найбільш високі ВПР досягаються для поєднання двох ознак {1,4}: экстремума і інтеграла від функції (6). Для сполучень ознак з трьох і вище ВПР істотно нижче.

При стисненні за допомогою моментів функцій (6) досліджувалася інформативність чотирьох моментів: 0-го, 1-го, 2-го і 3-го порядків. Відповідні номери цих ознак при дослідженнях: 1,2,3 і 4. Найкращі з отриманих результатів ВПР приведені в таблиці 2.

Таблиця 2.

Ознаки	$d_1(x)$	$d_2(x)$	$d_3(x)$
4	85,83	88,66	100.00
2,3	97,16	100.00	100.00
1,2,3	95,83	100.00	100.00
1,3,4	95,16	100.00	100.00

При використанні перетворення сплесків (Wavelet-перетворення) від функцій (6) виділялися в якості ознак наступні характерні точки: 1 - максимум функції (по модулю); 2 - час, при якому досягається максимум функції; 3 - значення функції в момент часу t=0; *4* - різниця між значенням функції в момент часу t=0 і значенням максимума функції; 5 - інтеграл від функції. Найкращі з отриманих результатів досліджень у вигляді ВПР приведені в таблиці 3.

Таблиця 3.

Ознаки	$d_1(x)$	$d_2(x)$	$d_3(x)$
3	91,11	90,06	94,71
1,4	98,80	100	98,08
1,3,4	98,80	100	98,08
1,3,4,5	98,56	100	98,08

Крім того, досліджувалися на інформативність ознаки, що являють собою значення Wavelet-образів, взяті через рівні проміжки часу на інтервалі, де сконцентрована вся енергія перетворених функцій. При дослідженнях таких ознак вибрано п'ять: 1,2,...,5. Отримані результати ВПР приведені в таблиці 4.

Таблиця 4.

Ознаки	$d_1(x)$	$D_2(x)$	$d_3(x)$
3	82,83	91,77	100
2,4	97,66	100	100
2,3,4	92,83	100	100
1,2,3,4,5	94,83	100	100

З таблиць 3 і 4 видно, що інформативність ознак, виділених на Wavelet-образах ЯВ вище, ніж інформативність ознак виділених в характерних точках ЯВ (таблиця 1).

ВИСНОВКИ

Запропонований в роботі метод діагностичного контролю, заснований на використанні для опису нелінійних об'єктів контролю багатомірний вагових функцій (ЯВ) і стисненні діагностичної інформації, досліджений експериментально за допомогою імітаційного моделювання. Отримані результати показують ефективність методу при розпізнаванні чотирьох класів станів для розглянутого тестового нелінійного об'єкта контролю.

Аналіз діагностичної цінності сукупностей ознак, що формуються, показав, що найбільш висока інформативність ознак досягається при стисненні діагностичної інформації за допомогою Wavelet перетворення.

ЛІТЕРАТУРА

1. Данилов Л.В., Матханов П.Н., Филиппов Е.С. Теория нелинейных электрических цепей. - Л.: Энергоатомиздат. Ленигр. отд-ние, 1990. - 256 с.
2. Ту Дж., Гонсалес Р. Принципы распознавания образов. / Пер. с англ. Под ред. Ю.И.Журавлева. - М.: Мир, 1978. - 411 с.
3. Апарцин А.С., Солодуша С.В. О математическом моделировании нелинейных динамических систем рядами Вольтерры. - Электронное моделирование, 1999, №2, с.3-12.
4. Павленко В.Д., Фомин А.А. Комбинированный метод построения решающего правила статистической классификации. – Электронное моделирование, 2001, №4, с.34-39.
5. Лотоцький Р.В. Методи стискання даних за допомогою перетворення сплесків.-У зб.: Оброблення сигналів і зображень та розпізнавання образів: Праці 5-ої Всеукраїнської міжнародної конференції УкрОБРАЗ'2000, Київ, 27 листопада -1 грудня 2000 р., с.63-66.

ФОРМУВАННЯ ФУНКЦІЇ ПРОСТОРОВОГО РОЗРІЗНЕННЯ РОЗПОДІЛЕНОЇ МЕРЕЖІ ДАТЧИКІВ У СИСТЕМІ НЕРУЙНУЮЧОГО КОНТРОЛЮ

В.Ф. Миргород, І.М. Гвоздєва, С.М. Ковешніков

Одеський національний політехнічний університет, Україна, 65044, м. Одеса, пр. Шевченка, 1;
тел.: +380 482 250-104, факс: +380 482 343-848; e-mail: acrobat@eurocom.od.ua
http://www.eurocom.od.ua/~acrobat/ тел.: (0482) 288-847, 288-227, 344-233
Одеський державний екологічний університет, Україна, 65016, м. Одеса, вул. Львівська, 15;
http://www.ogmi.farlep.odessa.ua

На основі апроксимації бажаної функції просторового розрізнення усіченим рядом Фур'є-Бесселя запропонована методика структурного і параметричного синтезу сигнал-процесора розподіленої вісесиметричної мережі датчиків для систем ультразвукового неруйнуючого контролю.

ВСТУП

Для неруйнуючого контролю різноманітного технологічного устаткування в енергетиці і двигунобудуванні широко використовується метод високочастотного акустичного зондування. Виявлення і класифікація мікродефектів при такому зондуванні вимагає одержання високої розрізнювальної здатності як у подовжньому (по глибині зондування), так і в поперечному (по куту) напрямках. Наприклад, при контролі корпусних стінок використовуються робочі частоти порядку 2,25 МГц, що дозволяє одержати найкраще розрізнення по дистанції (3...5) мм, однак досяжне поперечне розрізнення у кілька разів гірше за розрізнення по глибині [1]. Встановлені в даний час стандарти вимагають проведення зондування при декількох різних кутах падіння пучка, тому функціонально необхідним є використання розподіленої мережі датчиків, організованої у фазовані антенні грати (ФАГ) при прийомі ультразвуку. Класичні аддитивні алгоритми обробки сигналів у таких ґратах дозволяють одержати кутову розрізнювальну здатність, обмежену, як відомо [1], критерієм Релея. Подальше підвищення кутової розрізнювальної здатності у класі лінійних алгоритмів можливо лише шляхом збільшення розмірів ФАГ, що на практиці не завжди можливо.

ОСНОВНІ РЕЗУЛЬТАТИ

Висока просторова когерентність вторинного акустичного поля може бути ефективно використана для підвищення розрізнення в поперечному напрямку шляхом нелінійної обробки вихідних сигналів, наприклад, мультиплікативної, однак дотепер не було запропоновано досить загального підходу до синтезу таких алгоритмів. Для

багатомірних поліномів другого порядку синтез алгоритмів просторової обробки в одномірних ФАГ може бути здійснений на основі апроксимації бажаної функції просторового розрізнення поліномами Чебишева, а в конформних нееквідістантних ФАГ — шляхом розкладання в ряд Фур'є по просторових гармоніках [2, 3]. У даній роботі отримані раніше результати узагальнені на мережу датчиків, що має циліндричну симетрію.

Вихідний сигнал мультиплікативних кільцевих ФАГ з N датчиків за аналогією з [2] може бути представлений у вигляді:

$$U = \int_0^T \vec{x}^T(t) W \vec{x} dt, \qquad (1)$$

де W — деяка позитивно визначена симетрична матриця;

$\vec{x}(t)$ — вектор сигналів датчиків, що складається з компонентів вигляду

$$x_m(t) = X\cos\left[\omega t + ka\sin\gamma\cos\left(\frac{2\pi m}{N} - \varphi\right)\right];$$

$k = 2\pi/\lambda$ — хвильове число;

λ — довжина хвилі вузькосмугового прийнятого сигналу;

γ, φ — кути спостереження у вертикальній і горизонтальній площині відповідно;

T — інтервал усереднення, $T >> 2\pi/\lambda$;

a — радіус ФАГ;

X — амплітуда сигналу.

Матриця W визначає структуру сигнал-процесора мультиплікативних ФАГ, що містить у загальному випадку $(N^2 - N)/2$ каналів, оскільки квадратичні компоненти в (1) є неінформативними. Виділимо для ілюстрації в (1) N каналів і після інтегрування одержимо:

$$U(\gamma,\varphi) = X^2 \cdot \sum_{m=0}^{N-1} b_m \cos\left[2ka\sin\gamma\cos\left(\frac{2\pi m}{N} - \varphi\right)\right], \qquad (2)$$

де b_m — коефіцієнти передачі по каналах ФАГ.

Розкладаючи (2) у ряд по функціях Бесселя, діаграмоформуючу формулу для мультиплікативних ФАГ можна представити у вигляді:

$$U(\gamma,\varphi) = NX^2 J_0(2ka\sin\gamma) + 2X^2\sum_{\nu=1}^{\infty}(-1)^{\nu} J_{2\nu}(2ka\sin\gamma)\cdot \sum_{m=0}^{N-1} b_m \cos\left(\frac{2\pi m}{N} - \varphi\right). \qquad (3)$$

Співвідношення (3) є вихідним для синтезу різноманітних конфігурацій ФАГ з просторово розподіленої мережі датчиків з циліндричною симетрією. Важливим частковим випадком є кільцеві грати з простим перемножуванням сигналів діаметрально розташованих датчиків з наступним підсумовуванням і усередненням. Вихідний сигнал таких ФАГ має вигляд:

$$U(\gamma,\varphi) = \frac{N}{2}X^2 J_0(2ka\sin\gamma) + NX^2\sum_{i=1}^{\infty}(-1)^{iN/2} J_{iN}(2ka\sin\gamma)\cos iN\varphi. \qquad (4)$$

З огляду на властивості функцій Бесселя з великим індексом, при виконанні умови

$$2ka\sin\gamma < N \qquad (5)$$

у розкладанні (4) досить враховувати тільки перший член ряду. У порівнянні з аддитивною обробкою розрізнювальна здатність, що досягається, може бути поліпшена не менш, ніж у 2 рази, тому що ширина основного пелюстка функції просторового

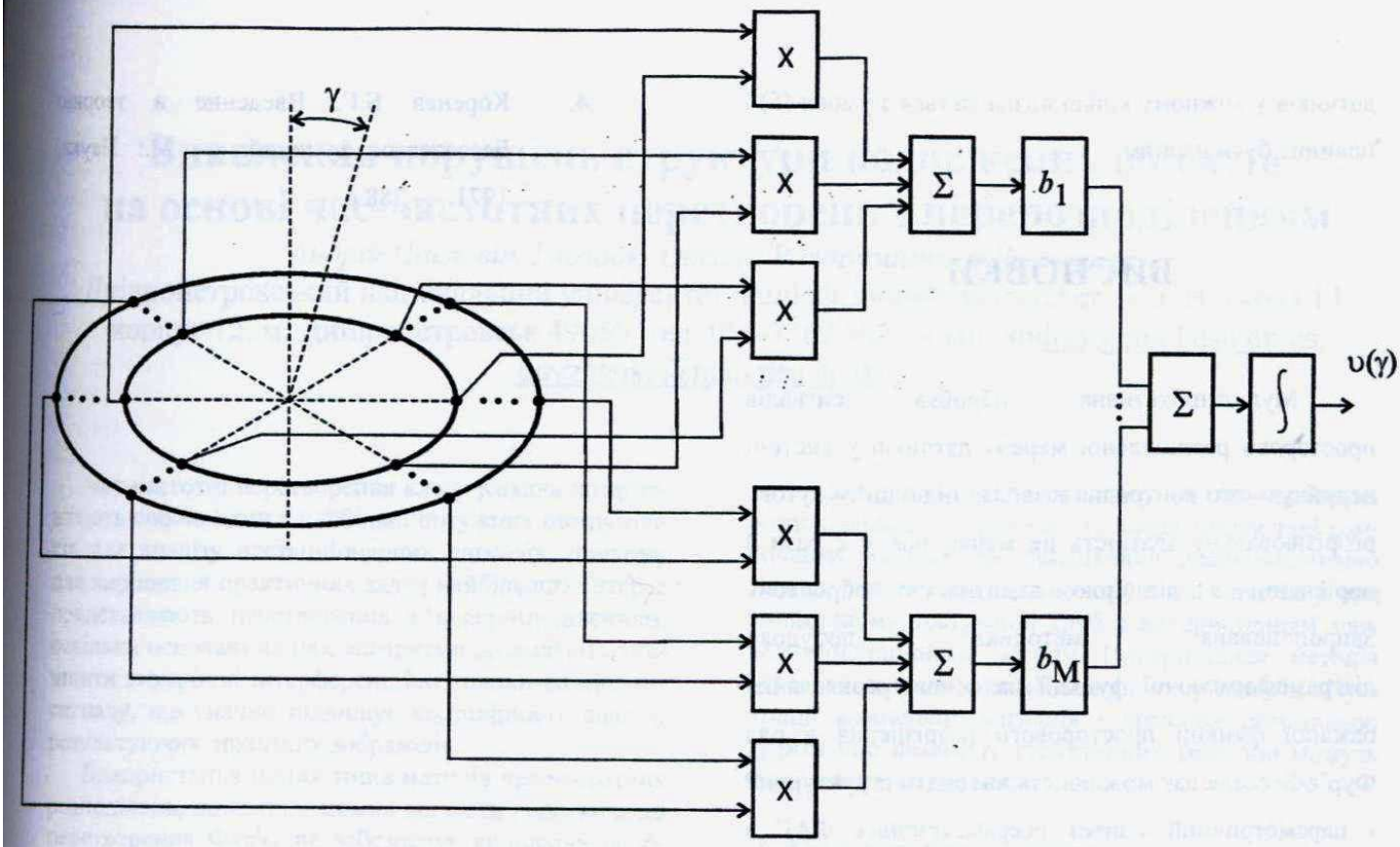


Рис 1. Мультиплікативні ФАГ з кільцевою структурою

розрізнення визначається співвідношенням $\gamma_0 = 2\arcsin(2{,}405/2ka)$.

Подальше підвищення розрізнювальної здатності досягається об'єднанням мережі датчиків у концентричні кільця з аддитивним зважуванням (рис. 1), вихідний сигнал якої має вигляд:

$$U(\gamma) = \rho_0 \sum_{k=1}^{M} b_k J_0(\rho_k x), \qquad (6)$$

де $x = (2ka\sin\gamma)/\mu_1$;

$\rho_k = a_k / a_1$;

a_k — радіус k-го кільця,

M — кількість кілець ФАГ,

ρ_0 — коефіцієнт нормування,

b_k — коефіцієнти зважування.

Якщо $f(x)$ — бажана функція просторового розрізнення, то при досить загальних припущеннях її можна представити у вигляді розкладання в ряд Фур'є-Бесселя [4]

$$f(x) = \sum_{k=1}^{\infty} z_k J_0(\mu_k x), \qquad (7)$$

де μ_k — позитивні нулі функції J_0, розташовані в порядку зростання,

z_k — коефіцієнти розкладання,

$$z_k = \frac{2}{[J_1(\mu_k)]^2} \int_0^1 s f(s) J_0(\mu_k s) ds$$

Зіставляючи (6) і (7), методика синтезу мультиплікативних ФАГ зводиться до апроксимації бажаної функції просторового розрізнення усіченим рядом вигляду (7) з необхідною точністю. Число членів ряду, необхідне для досягнення необхідної точності апроксимації, визначає число кілець ФАГ, а коефіцієнти отриманого розкладання — коефіцієнти передачі по каналах підсумовування. При цьому радіуси кілець повинні співвідноситися між собою, як співвідносяться позитивні нулі функції J_0. Число

датчиків у кожному кільці визначається з умови (5) і повинне бути парним.

ВИСНОВКИ

Мультиплікативна обробка сигналів просторово розподіленої мережі датчиків у системі неруйнуючого контролю дозволяє підвищити кутову розрізнювальну здатність не менш, ніж у 2 рази в порівнянні з лінійною аддитивною обробкою. Запропонована методика побудови діаграмоформуючої функції на основі розкладання бажаної функції просторового розрізнення в ряд Фур'є-Бесселя дає можливість виконати структурний і параметричний синтез осесиметричних ФАГ з наперед заданою точністю.

ЛІТЕРАТУРА

1. Кайно Г. Акустические волны: Устройства, визуализация и аналоговая обработка сигналов / Пер. с англ. — М.: Мир, 1990. — 656 с.
2. Миргород В.Ф., Гвоздева И.М. Формирование акустических изображений при неразрушающем контроле технологического холодильного оборудования // Холодильная техника и технология. — 1999, вып. 64 — с. 126—129.
3. Гвоздева И.М., Миргород В.Ф. Алгоритм параметрического синтеза процессоров пространственно-временной обработки сигналов конформных антенных решеток. // Электронное моделирование. — 2002. — т. 24. — № 3. — с. 61—66.
4. Коренев Б.Г. Введение в теорию Бесселевых функций. — М.: Наука, 1971. — 288 с.

Виявлення порушень структури подовжених об'єктів на основі час-частотних перетворень з перерозподіленням

Андрій Павлович Гвоздак, Оксана Володимирівна Довженко

Дніпропетровський національний університет, Радіофізичний факультет, вул. Наукова 13, корпус 12, м. Дніпропетровськ 49050. тел. 056-7769092. e-mail: andrey@mail.dsu.dp.ua, dovzhenko@mail.dsu.dp.ua .

Час-частотні перетворення класу Кохана представляють собою один з найбільш потужних інструментів для аналізу нестанціонарних сигналів, причому для вирішення практичних задач найбільший інтерес представляють перетворення з перерозподіленням, оскільки основані на них алгоритми дозволяють зменшити внутрішні інтерференційні впливи компонент сигналу, що значно підвищує достовірність аналізу результуючих яркістних зображень.

Використання інших типів методів час-частотних розподілень, до котрих можна віднести короткочасні перетворення Фур'є, не забезпечує виділення необхідних компонент сигналу, що пов'язано з наявністю обмежень, обумовлених принципом Гейзенберга, і вимагає додаткової обробки та складної інтерпретації результатів.

В рамках даної роботи метод, оснований на час-частотному перетворенні з перерозподіленням, було використано для виявлення дефектів при магнітно-індукційному тестуванні гаряче катаних труб. Експериментальні дослідження, проведені для реальних сигналів, показують, що метод перерозподілення дозволяє значно підвищити чутливість аналізу форми і положення випадкових дефектів у зрівнянні з стандартними підходами, а також забезпечує гнучке регулювання порогу чутливості аналізу за рахунок зміни параметрів типу і розміру вікон згладження.

1. Вступ

Метод магнітно-індукційної дефектоскопії (МІД) [1], оснований на використанні короткочасного перетворення Фур'є (КЧПФ), які відносять до лінійних час-частотних перетворень, що належать до перетворень першого класу, не забезпечує необхідної чутливості аналізу нестаціонарних сигналів.

Використання час-частотного аналізу дозволяє дослідити, як функція розподілення представляє компоненти вихідного сигналу розподіленням енергій як за часом, так і за частотою. В даній роботі розглядається практична можливість застосування перетворення класу Кохана в області квадратичного час-частотного розподілення (TFD (time-frequency distribution) – ЧЧП), причому до класу Кохана відносяться всі квадратичні розподілення, що мають властивість коваріантності до зсуву як за часом, так і за частотою.

Традиційні методи, основані на час-частотних розподіленнях, не дозволяють локалізувати такі компоненти сигналу, як випадковий високочастотний сигнал, викликаний зміною структури матеріалу при динамічному тестуванні труб з використанням магнітно-індукційного методу. Використання методів перерозподілення сильно підвищує чутливість візуалізації компонент сигналів і дозволяє однозначно судити про наявність структурних змін, що можуть відноситися до дефектів.

2. Час-частотні перетворення класу Кохана

На відміну від лінійних час-частотних перетворень, що розкладають сигнал на елементарні компоненти, метою енергетичних перетворень є представлення енергії сигналу двома змінними – часом і частотою.

Початковим у цьому контексті поняттям є енергія сигналу $x(t)$, що може бути отримана як квадрат модуля, як сигналу, так і його Фур'є перетворення,

$$E_x = \int_{-\infty}^{+\infty} |x(t)|^2 \, dt = \int_{-\infty}^{+\infty} |X(f)|^2 \, df, \qquad (1)$$

з чого ми можемо інтерпретувати $|x(t)|^2$ і $|X(f)|^2$ як енергетичні щільності відповідно за часом і за частотою. Поєднуючи щільності енергій, визначимо функцію $\rho_x(t, f)$, для котрої справедливий вираз:

$$E_x = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \rho_x(t, f) \, dt \, df \, , \qquad (2)$$

котрий отримуємо з (1). Таким чином, енергія є квадратичною функцією сигналу, відповідно до чого і енергія час-частотного перетворення спектра також в загальному випадку має квадратичне представленням.

Енергетична щільність розподілу повинна задовольняти двом наступним маргінальним властивостям:

$$\int_{-\infty}^{+\infty} \rho_x(t, f) \, dt = |X(f)|^2 \qquad (3)$$

$$\int_{-\infty}^{+\infty} \rho_x(t, f) \, df = |x(t)|^2 , \qquad (4)$$

які позначають, що якщо інтегрувати функцію енергетичної щільності час-частотного перетворення за одною змінною, то вийде енергетична щільність перетворення за іншою змінною.

Якщо для перетворень справедливі вирази (2), (3) і (4), можна накласти додаткові обмеження на $\rho_x(t, f)$ таким чином, що ці перетворення будуть задовольняти іншим необхідним властивостям, серед котрих фундаментальну важливість має коваріантність до зміщення за часом і частотою. Енергетичні перетворення, що задовільняють всім вищепереліченим вимогам, відносяться до класу Кохана [4].

Базовим перетворенням для енергетичних час-частотних розподілів класу Кохана, є розподіл Вігнера–Віля (WVD–Wigner-Ville distribution-ВВП), що представляється наступним чином:

$$W_x(t,f)=\int_{-\infty}^{+\infty} x(t+\tau/2)x^*(t-\tau/2)\, e^{-j2\pi f\tau}d\tau, \quad (5)$$

Коваріантність за часом і за частотою представляє собою найбільш важливу характеристику час-частотного перетворення, оскільки саме вона гарантує, що якщо сигнал має затримку в часі і є модульованим, його час-частотний розподіл є зміщеним на деяку величину в частотній області. Клас енергетичних час-частотних розподілів вводить коваріантність процесів за часом і за частотою в наступному виразі:

$$C_x(t,f;q)=\iint\int_{-\infty}^{+\infty} e^{i2\pi\xi(s-t)}q(\xi,\tau)x(s+\tau/2)\times$$
$$x^*(s-\tau/2)e^{-i2\pi f\tau}d\xi\, ds\, d\tau\ ,$$

де $q(\xi,\tau)$ - двовимірна функція, котра має назву параметризаційної (тобто такої, що описує фізичні процеси за допомогою параметрів моделі). Цей клас перетворень відомий як клас перетворень Кохана, і можна записати:

$$C_x(t,f;\Pi)=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\Pi(s-t,\xi-f)\times$$
$$W_x(s,\xi)ds\, d\xi, \quad (6)$$

де

$$\Pi(t,f)=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} q(\xi,\tau)e^{-i2\pi(f\tau+\xi t)}dt\, df$$

є двохвимірним перетворенням Фур'є параметризаційної функції $q(\xi,\tau)$.

Велику кількість існуючих енергетичних час-частотних перетворень можна віднести до цього класу, серед котрих і перетворення ВВП, у якому функція $\Pi(t,f)$ є подвоєною функцією Дірака:

$$\Pi(t,f)=\delta(t)\delta(f),$$

тобто $q(\xi,\tau)=1$.

Функція згладження $\Pi(t, f)$ має наступні характеристики:

- специфікація параметрізаційної функції $q(\xi,\tau)$ дозволяє отримувати найбільш відомі енергетичні розподілення;
- забезпечується можливість перетворення необхідних обмежень перетворень в умову припустимості параметризаційної функції;
- забезпечується можливість використовувати такі припустимі аргументи, що клас рішень формується у відповідності до заданих умов.

Клас Кохана містить у собі всі квадратичні час-частотні розподіли, що забезпечують широкий набір засобів для аналізу нестаціонарних сигналів.

Базова ідея класу перетворень полягає у введенні об'єднуючої за часом і частотою функції, котра описує енергетичну щільність або інтенсивність сигналу одночасно в обох областях. Найбільш важливим елементом цього класу розподілів можна вважати розподіл Вігнера-Віля, що задовольняє цілому ряду необхідних властивостей. Оскільки ці розподіли є квадратичними, вони містять взаємовпливаючі компоненти в час-частотній площині, котрі можуть порушувати можливість локалізації компонент досліджуваного сигналу. Одним із способів зниження впливу інтерферентності є згладжування розподілу за часом і частотою, однак застосування згладження призводить до втрат роздільної здатності як за часом, так і за частотою.

Крім класу Кохана енергетичних розподілень існує цілий ряд розподілень, що не задовільняють властивості коваріантності, наприклад клас афінних перетворень.

3. Метод перерозподілення

Білінійні час-частотні перетворення забезпечують широкий діапазон методів для аналізу нестаціонарних сигналів. Однак проблема усунення інтерференційних компонент на етапі візуалізації результатів залишається актуальною. Розглянемо один з варіантів її вирішення, котрий має назву методу перерозподілення.

Початкова ціль методу перерозподілу [2] - підвищення чутливості візуалізації спектрограм, оскільки саме спектрограми відбивають відношення інтерференційних і інформативних компонент сигнала.

Принцип перерозподілення ґрунтується на припущенні, що вагове усереднення значень не розподіляється симетрично навколо точок час-частотного простору (t, f), котрі представляють собою геометричні центри області [3]. З цієї причини усереднення повинне наближатися до центру тяжіння області, котра представляється більш репрезентативною в розподіленні локальної енергії сигналу. В механічній аналогії локальні енергії можуть бути розглянуті як розподілення мас, а центр тяжіння – як центр абсолютної маси.

Таким чином, сам метод представляється як реалізація наближення кожного значення спектрограми, обчисленого в (t, f), до іншої точки $(\hat{t},\hat{f})$, що є центром тяжіння розподілення енергії сигналу навколо (t,f):

$$S_x^{(r)}(t',f';h)=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} S_x(t,f;h)\delta\left(t'-\hat{t}(x;t,f)\right)\times$$
$$\delta\left(f'-\hat{f}(x;t,f)\right)dt\, df. \quad (7)$$

Іншими словами, відбувається перерозподіл спектрограм в такий спосіб, що значення в будь-якій точці результуючої спектрограми (t', f') представляє собою суму всіх значень перерозподілених до поточного центру точок.

Для практичної реалізації методу, (7) замінюється наступними виразами:

$$\hat{t}(x;t,f)=t-\Re\left\{\frac{F_x(t,f;t\cdot h(t))F_x^*(t,f;h)}{|F_x(t,f;h)|^2}\right\},$$

$$\hat{f}(x;t,f)=f-\Im\left\{\frac{F_x\left(t,f;\frac{dh(t)}{dt}\right)F_x^*(t,f;h)}{|F_x(t,f;h)|^2}\right\}.$$

Така реалізація методу дозволяє уникнути високих обсягів обчислень,а відповідно, витрат ресурсів обчислювальної системи.

В якості підсумку можна зазначити, що перерозподілення спектрограм не представляє собою білінійну функцію,однак задовольняє властивостям коваріаційного зсуву, збереження енергії (припускається, що $h(t)$ представляє собою одиницю енергії) і невід'ємності. Це також можна довести, як перетворення Вігнера-Віля являє собою ідеальне перетворення для локалізації лінійно-частотно модульованих та імпульсних сигналів, люба перерозподілена спектрограма також відповідає цим властивостям:

$$x(t)=Ae^{i(f_0t+\alpha t^2/2)}\Rightarrow\hat{f}=f_0+\alpha\hat{t},$$

$$x(t)=A\delta(t-t_0)\Rightarrow\hat{t}=t_0.$$

4. Використання методу перерозподілення для контролю змін структури подовжених об'єктів

Представлений принцип перерозподілення може бути застосований для цілого ряду перетворень. Якщо розглянути загальний вираз розподілення класу Кохана як двовимірну згортку з ВВП,

$$C_x(t,f;\Pi)=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\Pi(t-s,f-\xi)\times W_x(s,\xi)\,ds\,d\xi,$$

замінюючи частину ядра згладження $W_h(s,\xi)$ на довільне ядро $\Pi(s,\xi)$, визначимо перерозподілення для кожного члена класу Кохана:

$$\hat{t}(x;t,f)=\frac{\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}s\Pi(t-s,f-\xi)W_x(s,\xi)\,ds\,d\xi}{\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\Pi(t-s,f-\xi)W_x(s,\xi)\,ds\,d\xi},$$

$$\hat{f}(x;t,f)=\frac{\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\xi\,\Pi(t-s,f-\xi)W_x(s,\xi)\,ds\,d\xi}{\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\Pi(t-s,f-\xi)W_x(s,\xi)\,ds\,d\xi},$$

$$C_x^{(r)}(t',f';\Pi)=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}C_x(t,f;\Pi)\times\delta(t'-\hat{t}(x;t,f))\delta(f'-\hat{f}(x;t,f))\,dt\,df.$$

Останнє перерозподілене перетворення вигідно поєднує як властивість зменшення впливу компонент інтерференції, так і посилення концентрації інформативних компонент сигналу. Крім того, це перетворення має властивість коваріантності до зсуву за часом і частотою і проявляє високу чутливість до локалізації лінійної ЧМ і імпульсів [5].

5. Експериментальна частина

Експериментальні дослідження проводилися на основі сигналів, знятих з магнітно-індукційного дефектоскопа.

В рамках експериментів було досліджено різні типи час-частотної репрезентації сигналу на основі таких перетворень класу Кохана як спектрограми Габора, Margenau Hill розподілення, розподілення Вігнера-Вілля і згладжене розподілення Вігнера-Вілля. Аналіз результатів показує, що найбільш вигідним з точки зору представлення локалізацій компо нентів дефекту, а також витрат обчислювальних ресурсів, є використання Margenau Hill розподілення, рівняння котрого може бути представлене у наступній формі

$$MH_x(t,f)=\int_{-\infty}^{+\infty}\frac{1}{2}(x(t+\tau)x^*(t)+x(t)x^*(t-\tau))e^{-j2\pi f\tau}d\tau$$

де $x(t)$ - вхідний сигнал.

Інтерференційна структура перетворення Margenau Hill відрізняється від структури розподілення Вігнера-Віля, яка є базовою для розподілень класу Кохана: інтерференційні компоненти які локалізовані в двох точках (t_1,f_1) і (t_2,f_2), будуть розташовані в точках з координатами (t_1,f_2) і (t_2,f_1).

Таким чином, використання розподілення Margenau Hill для сигналів які мають багато компонент локалізованих в одному місті в часі або частоті може дати помилкові результати, тому що компоненти інтерференції будуть накладатися на компоненти сигналу. Тому метод перезозподілення дуже рідко використовується для перетворень, які не мають в ядрі функцію сгладжування інтерференційних компонент сигналу. Метод перезподілення на практиці використовується до псевдо або перетворень з поменшенною інтерференцією (Reduce Interference Distribution).

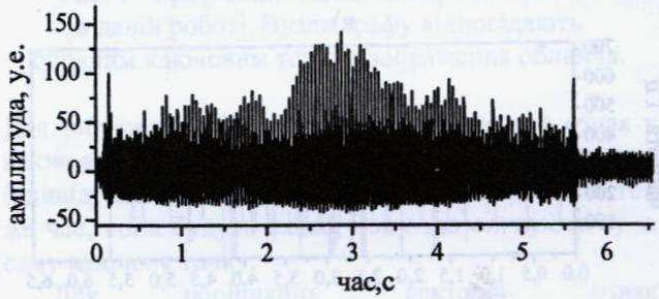


Рисунок 1. Сигнал, записаний з магнітно–індукційного дефектоскопа.

Розглянемо застосування методу перерозподілення для Margenau Hill розподілів і порівняємо результати з короткочасними перетвореннями Фур'є.

На рис.1 представлений сигнал, знятий з магнітно-індукційного дефектоскопа з частотою дискретизації 5000 Гц.

На рис. 2 а-г, відповідно, представлені розподіли КЧПФ, розподіл Margenau Hill, псевдо Margenau Hill і перерозподілений розподіл Margenau Hill. Для КЧПФ критерієм наявності дефекту є розщеплення основного компонента сигналу в час-частотній площині, для перерозподіленого Margenau Hill розподілення - локалізація компонентів сигналу дефектів, що є відображенням структури матеріалу труби на час-частотній площині при використанні вузькополосного зглажуючого вікні Hanning'а.

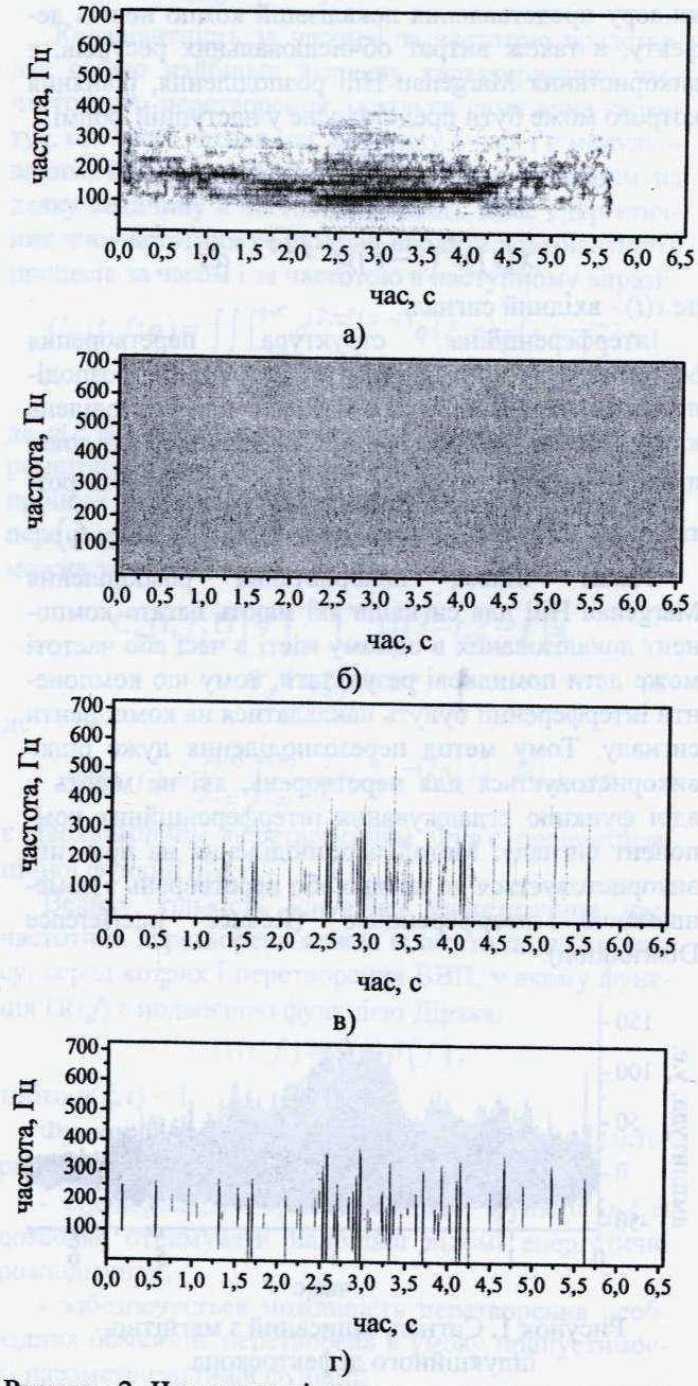


Рисунок 2. Час-частотні перетворення сигналу рис.1 з параметрами: розмір вікна 512 вибірок, крок вікна 4 вибірки, тип згладжування вікна - Hanning, розмір вікна згладжування 63 для перетворень класу Кохана: а - КЧПФ, б - розподіл Margenau Hill, в- псевдо перетворення Margenau Hill, г- перерозподілений розподіл Margenau Hill.

Візуальний аналіз отриманих результатів дозволяє зробити однозначний висновок про переваги методу перерозподілення на основі перетворень класу Кохана для виявлення компонент дефектів в час-частотній площині у зрівнянні з традиційними підходами. Метод перезподілення (рис. 2г) дозволяє визначити положення змін структури об'єкта що тестується. а за довжиною спектральних рисок визначити належить чи така зміна до дефекту.

6. Висновки

Використання методу перерозподілення на основі час-частотних перетворень класу Кохана дозволило:

- поліпшити візуалізацію компонентів сигналів у час-частотній області, у порівнянні з методом короткочасних перетворень Фур'є;
- зменшити помилки взаємовпливу компонент у порівнянні з іншими типами перетворень класу Кохана;
- метод збільшує концентрацію компонентів сигналу в час-частотній області, що дозволяє збільшити співвідношення сигнал/шум;
- реалізація методу перерозподілення практично не вимагає додаткових обчислювальних можливостей, що дозволяє реалізувати метод у реальному масштабі часу;

До недоліків можна віднести наявність компонент взаємовпливу в час-частотній області, що при близьких частотах може призводити до появи інтерференційних ефектів.

Література

1. A.P. Gvozdak, V. M. Dolgov, *Time-frequency NDT in real time mode based on DSP*, Signal/Image Processing and Pattern Recognition, Proceeding, The Fourth All-Ukrainian International Conference, 19-23 October, Kiev, pp. 113-116, 1998.
2. F. Auger and P. Flandrin, *The why and how of time-frequency reassignment*, IEEE International Symposium on Time-Frequency and Time-Scale Analysis, pp. 197-200, USA Philadelphia 1994.
3. F. Auger and P. Flandrin, *Improving the Readability of Time-Frequency and Time-Scale Representations by the Reassignment Method*, IEEE Transactions on Signal Processing, vol. 43(5): 1068-1089, 1995.
4. L. Cohen, *Time-Frequency Distribution – A Review*, Proceedings of the IEEE, 77(7): 941-980, 1989.
5. P. Flandrin, *On Detection-Estimation Procedures in the Time-Frequency Plane*, In International conference on acoustics, speech and signal processing, pp. 43.5.1-4, 1986.

ВИЯВЛЕННЯ КЛЮЧОВИХ ТОЧОК НА ЗОБРАЖЕННІ ОБЛИЧЧЯ ЛЮДИНИ

Муригін К.В.
Донецький державний інститут штучного інтелекту
83050, м. Донецьк, проспект Богдана Хмельницького, 84
Тел.: (0622) 926082, факс: (0622) 926082
E-mail: kir@iai.donetsk.ua

У даній роботі описуються кілька схожих підходів до задачі локалізації ключових точок обличчя на зображенні. Усі методи засновані на кореляційному пошуку в просторі векторів ознак, отриманих на основі згорток зображення обличчя людини з набором функцій Габора різних масштабів і кутів повороту. Обговорюються переваги і недоліки описаних методів, приводяться результати їх порівняння.

ВСТУП

Для рішення задачі розпізнавання людини по зображенню обличчя в даний час використовується велика кількість різноманітних методів. Однак більшість з них зіштовхуються з однією з основних труднощів – проблемою відповідності порівнюваних властивостей. Наприклад, невеликі розбіжності в масштабах та кутах зйомки можуть істотно вплинути на результат роботи алгоритмів розпізнавання, що працюють із зображенням обличчя в цілому. Одним з напрямків рішення цієї проблеми є застосування методів, що працюють з окремими деталями обличчя. У рамках таких методів проблема відповідності порівнюваних властивостей трансформується в проблему правильного виявлення і локалізації відповідних деталей зображення обличчя. Подібно тому, як у фізиці не можна порівнювати кілограми з метрами, так і при розпізнаванні не можна порівнювати, наприклад, брова й око людини.

Таким чином, задача визначення відповідності або локалізації точок обличчя, що використовуються при розпізнаванні (ключових точок), представляється дуже важливою, і виступає необхідним етапом на шляху до побудови системи ідентифікації людини по фотографії.

1. ОПИС МЕТОДУ

Серед методів заснованих на локальних властивостях останнім часом велике поширення одержав метод, що використовує як ознаки згортки вихідного зображення з набором функцій Габора [1,2]. Популярність цього підходу багато в чому порозумівається даними біології про те, що схожу на габоровську форму має рецепторне поле клітини частини головного мозку [3,4], що відповідає за зір.

Як ознаки, що визначають властивості зображення обличчя, метод використовує вектора, що складаються з комплексних елементів, отриманих за допомогою згортки вхідного зображення обличчя з функціями Габора різних масштабів і кутів повороту [1,2]. У даній роботі використовувалися сорок функцій Габора п'яти різних масштабів і восьми орієнтацій. Таким чином, кожна точка зображення перетворювалася в описаний вектор ознак, що у деякій формі містить інформацію про розподіл яскравості навколо даної точки.

Ключові точки розташовувалися на зображенні обличчя у виді сітки (графа обличчя) прямокутної структури (4×7 точок) (див. рис. 1).

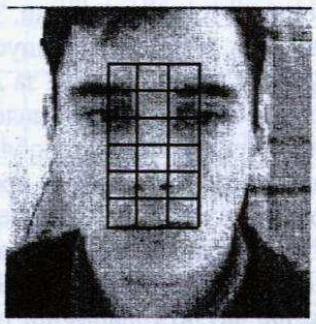


Рис. 1 - Граф обличчя, що використовувався в даній роботі. Вузли графу відповідають обраним ключовим точкам зображення обличчя.

Для зображень облич різних людей вектора ознак у ключових точках будуть відрізнятися, що пов'язано з індивідуальними особливостями обличчя, але, в той же час, вони будуть схожі, тому що описують ту ж саму ключову точку.

Для порівняння векторів ознак використовувалася міра подібності у виді косинуса кута між ними:

$$S(\vec{V}_i, \vec{V}_j) = \frac{\sum_n V_{ni} V_{nj}}{|\vec{V}_i| \cdot |\vec{V}_j|}.$$

Така міра дозволяє зробити порівняння інваріантним стосовно лінійних перетворень вхідних зображень,

тому що елементи векторів ознак не залежать від постійного зсуву зображень по шкалі яскравості, а сама міра подібності інваріантна стосовно постійних множників.

2. ПОСТАНОВКА ЗАДАЧІ І РОЗПОДІЛ НА ЕТАПИ

Метою проведених досліджень була побудова алгоритму виявлення ключових точок на зображенні обличчя, незалежно від індивідуальних особливостей.

Як вихідні дані для досліджень використовувалася база даних векторів ознак для кожної ключової точки обличчя, розрахована на основі бази даних, що містить зображення облич 200 різних людей. Усі зображення з бази даних отримані при нормальних умовах освітлення і попередньо приведені до одного масштабу (за масштабний коефіцієнт була обрана відстань між зіницями, еталонна відстань дорівнювала 34 пікселя). На кожному зображенні обличчя були вручну відзначені ключові точки, що відповідають особливим деталям обличчя (зіниці, брови, кінчик носу, краї губ), по яких автоматично будувався граф обличчя, у вузлах якого витягалися вектора ознак. У такий спосіб були враховані розбіжності в пропорціях зображень облич різних людей.

Як відзначалося в роботах [1,2], амплітуда комплексних елементів векторів ознак є повільно мінливою функцією, що дозволяє розділити задачу локалізації ключових точок на два етапи. З огляду на зазначену властивість, можна припустити, що точність розташування графа обличчя за допомогою векторів амплітуд буде слабко залежати від невеликих відмінностей у масштабі обличчя, його пропорціях, невеликого збільшення просторового кроку при кореляційному пошуку. Усе це дозволить швидко знаходити приблизне розташування графа обличчя, побудованого з урахуванням середніх пропорцій облич, що звузить діапазони пошуку для більш точного алгоритму локалізації. Як показують дані, отримані на основі використовуваної бази даних зображень облич, середній граф обличчя має наступні параметри:

$x_{ij} \approx 11{,}3 \times i$ и $y_{ij} \approx 9 \times j$, де $i = \{0,..,3\}$, а $j = \{0,..,6\}$.

Таким чином, другий етап локалізації ключових точок на зображенні обличчя зводиться до уточнення отриманих приблизних координат.

3. УТОЧНЕННЯ ПОЛОЖЕННЯ КЛЮЧОВИХ ТОЧОК

Для більш точного визначення координат ключових точок, амплітуд комплексних векторів ознак недостатньо. Виникає необхідність використання фазової інформації, що має характеристичну частоту, яка відповідає частоті габоровського ядра.

Як перший і найпростіший метод був досліджений метод, що використовує середні по базі даних вектора ознак. На рисунку 2 приведена функція щільності розподілу результатів порівняння (ФЩРРП) середнього вектора ознак і векторів з бази даних для точки ока.

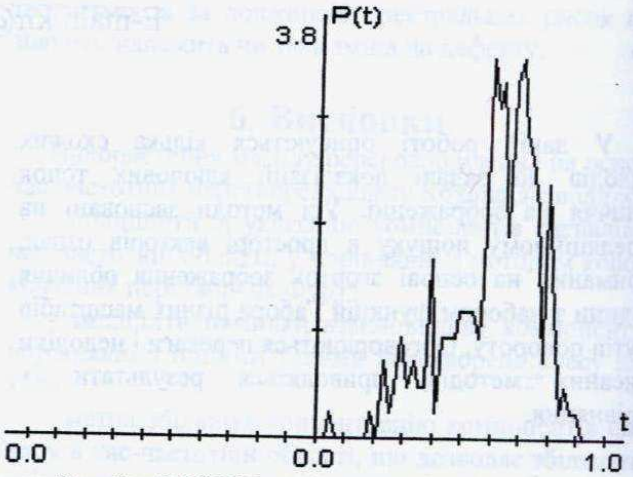


Рис. 2 - ФЩРРП середнього вектора ознак з векторами з БД.

Недоліки цього методу проявилися при першому ж випробуванні. Виявилося, що метод сильно орієнтований на середні пропорції облич. Наприклад, для точки ока пропорції обличчя визначаються взаємним розташуванням центра ока і брови, що проявилося в ефективному визначенні ключової точки для окремого класу облич і майже цілком помилкові результати для інших класів.

Для зменшення впливу на локалізацію локальних пропорцій обличчя були введені вагові коефіцієнти для елементів векторів ознак. Коефіцієнти були знайдені з умови мінімізації середньоквадратичної відстані між зображеннями одного класу, при накладенні обмеження, що виключає тривіальне не змістовне рішення. У монографії [5] відзначається, що така метрика (див. систему (1)) виражає властивість підмножини зображень, у якій вони найбільш подібні.

$$\begin{cases} d(a,b) = \sqrt{\sum_i w_i (a_i - b_i)^2}, \\ \sum_i w_i = 1. \end{cases} \quad (1)$$

Мінімізація дисперсії дає наступні значення для вагових коефіцієнтів:

$$w_i = \frac{1}{\sigma_i^2 \sum_n \frac{1}{\sigma_n^2}}.$$

На основі отриманої метрики були зроблені розрахунки ФЩРРП середнього вектора ознак і

векторів з бази даних. Як приклад на рисунку 3 приведена ФЩРРП для точки ока. Отриманий розподіл має менше середнє відхилення від середнього по базі даних вектора ознак, ніж для метрики, що не використовує вагові коефіцієнти. Однак, експерименти також показали сильну залежність точності виявлення від локальних пропорцій обличчя. Крім того, уведення вагових коефіцієнтів знижує кількість аналізованої при кореляційному порівнянні інформації, що підвищує імовірність помилкового виявлення шуму, а також знижує точність локалізації ключових точок.

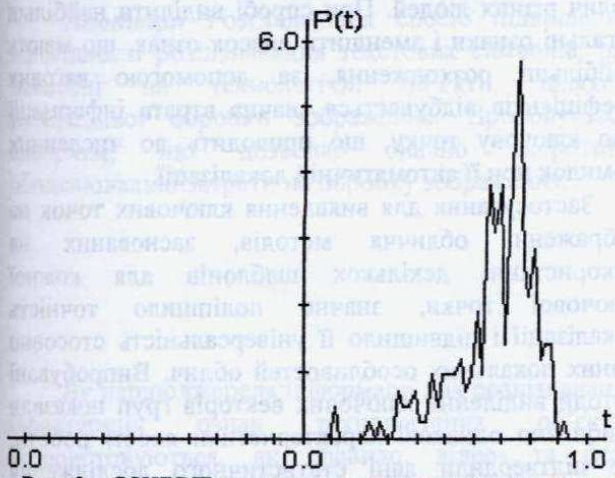


Рис. 3 - ФЩРРП середнього вектора з векторами з БД з урахуванням вагових коефіцієнтів.

Аналіз отриманих результатів показав необхідність уведення декількох шаблонів для кожної ключової точки обличчя. Це привело до задачі розділення всього класу локальних характеристик обличчя людини на групи, що мають найбільш схожі риси. Такий підхід дозволить більш точно знаходити ключові точки на зображенні обличчя, а також за рахунок класифікації локальних рис приведе до спрощення представлення обличчя в базі даних і, у кінцевому рахунку, до скорочення її обсягу. Останнє припускає подальше дослідження можливостей використання розділення на групи для вирішення задачі розпізнавання.

Для розділення векторів ознак на групи використовувалася умова обмеження відмінності між усіма векторами з групи деяким заданим числом ε. Після розподілу векторів ознак по групах, усі групи упорядковувалися по кількості членів і вибиралися N груп з найбільшою кількістю членів. Параметром оцінки ефективності розбивки W була кількість векторів властивостей, що потрапили в обрані N найбільш численних груп. У наших експериментах ми задалися наступними значеннями параметрів: N=10, W=150 (з 200). Потім для N=10 була обчислена таблиця 1, у якій відбита залежність числа облич, що попадають у N груп, від числа ε. Значення ε наведені в першому рядку, індекси ключових точок, відповідно до графа рис. 1,– у першому стовпці. Напрямок нумерації вузлів графа: ліворуч - праворуч і зверху - униз. Після цього для W=150 по таблиці 1 були обчислені значення ε для кожної ключової точки.

Таблиця 1 Залежність W від ε для кожної ключової точки обличчя.

№	0.1	0.2	0.3	0.4	0.5	0.6
0	200	191	178	148	120	93
1	199	191	167	144	113	79
2	200	191	168	138	111	77
3	200	195	185	160	131	92
4	199	190	169	126	98	55
5	200	191	163	114	83	48
6	200	191	158	129	96	56
7	200	193	182	144	102	61
8	190	165	130	96	61	43
9	200	200	192	158	120	75
10	200	199	191	165	129	78
11	199	171	132	97	68	43
12	190	157	125	98	62	38
13	200	200	197	178	129	96
14	200	200	200	187	152	108
15	199	174	138	107	70	52
16	189	177	134	105	69	48
17	200	200	197	174	140	87
18	200	200	197	179	134	100
19	199	184	152	124	84	50
20	187	159	125	106	69	42
21	200	200	193	177	149	117
22	200	200	198	185	158	120
23	197	170	137	96	72	49
24	148	114	83	61	47	33
25	190	171	149	120	80	52
26	190	177	149	121	82	63
27	139	121	92	62	50	35

Після поділу векторів ознак на групи були проведені два дослідження з визначення ключового вектора групи. У першому випадку ключовий вектор обчислювався як середній по групі. В другому – як середній між максимальним і мінімальним значенням кожної координати векторів групи. Результати обчислення ФЩРРП векторів ознак з бази даних і максимально схожих на них з 10 ключових векторів груп для обох випадків приведені відповідно на рисунках 4 і 5.

Обидва методи виділення ключових векторів груп показали добрі результати при тестуванні алгоритму виявлення ключових точок обличчя на зображеннях різних людей.

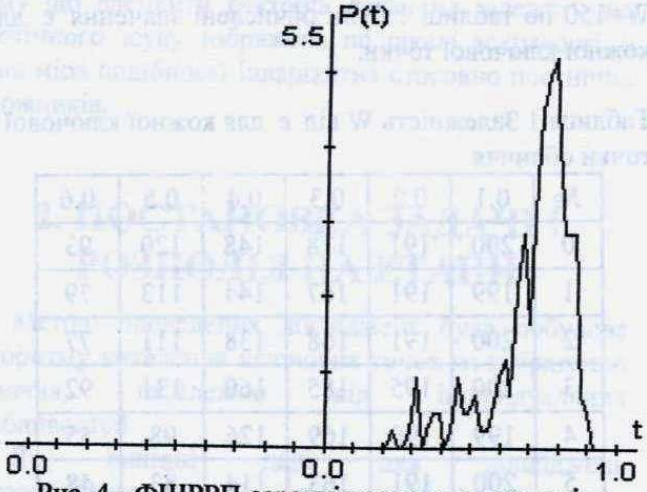


Рис. 4 - ФЩРРП середніх по групах векторів з векторами з БД.

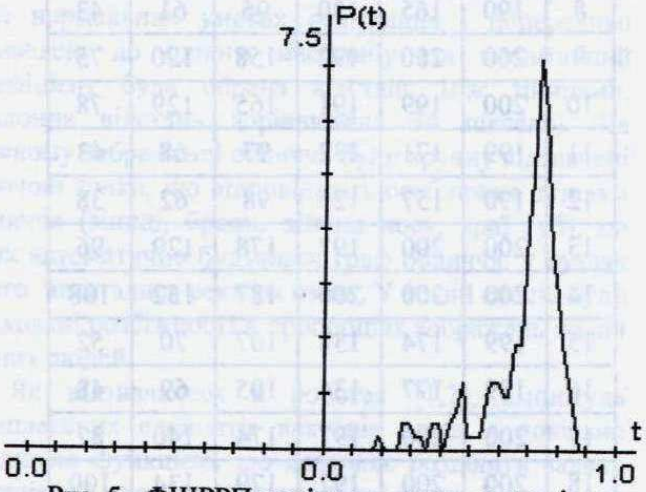


Рис. 5 - ФЩРРП медіан по групах векторів з векторами з БД.

Для порівняння усіх випробуваних методів були обчислені наступні статистичні дані: DM- середня відмінність векторів ознак з бази даних і шаблонних; P(s>por1)- імовірність того, що величина порівняння вектора з бази і шаблона буде більше por1; P(s>por2)- імовірність того, що величина порівняння вектора з бази і шаблона буде більше por2. Порівняльна характеристика для точки ока приведена в таблиці 2.

Таблиця 2 Порівняльна характеристика алгоритмів локалізації ключових точок на зображенні обличчя. Наведені дані відповідають точці ока.

Характеристика	DM	P(s>0.7)	P(s>0.8)
Метод середнього	0.59	0.35	0.05
Середній з вагами	0.66	0.57	0.15
Метод медіан по групах	0.67	0.69	0.08
Метод середнього по групах	0.70	0.70	0.21

ВИСНОВКИ

Проведені експерименти по тестуванню алгоритмів локалізації ключових точок на зображенні обличчя показали недостатність застосування одного шаблона для кожної ключової точки. Це пов'язано з індивідуальними особливостями зображень обличчя, які необхідно враховувати вже на етапі виявлення деталей обличчя. Такий результат порозуміється значним розходженням між векторами ознак, що відповідають одній ключовій точці, у зображень облич різних людей. При спробі виділити найбільш загальні ознаки і зменшити внесок ознак, що мають найбільші розходження, за допомогою вагових коефіцієнтів відбувається значна втрата інформації про ключову точку, що приводить до численних помилок при її автоматичній локалізації.

Застосування для виявлення ключових точок на зображенні обличчя методів, заснованих на використанні декількох шаблонів для кожної ключової точки, значно поліпшило точність локалізації і підвищило її універсальність стосовно різних локальних особливостей облич. Випробувані методи виділення ключових векторів груп показали приблизно однакові характеристики якості роботи, що підтвердили дані статистичного дослідження (див. таб. 2, рядки 4 і 5).

Ще одна перевага підходу, заснованого на використанні декількох шаблонів, це можливість використання отриманої класифікації ключових точок обличчя для задачі розпізнавання.

ЛІТЕРАТУРА

1. Wiskott L., Fellous J.M., Krueger N. and von der Malsburg C. Face Recognition and Gender Determination. // Proc. of the Int. Workshop on Automatic Face- and Gesture-Recognition, Zuerich, 1995. -p.92-97.

2. M. Lades et al. Distortion invariant object recognition in the dynamic link architecture. IEEE Trans. Comput., 42(3): 300-311,1993.

3. J. Jones, L. Palmer An evaluation of the two-dimensional Gabor-filter model of simple receptive fields in cat striate cortex. // J. Neurophysiol., 1987.- p.1233-1258.

4. D. Burr, M. Morrone, D. Spinelli Evidence for edge and bar detectors in human vision. // Vision Res., 1989, p.419-431.

5. Г.С. Себестиан Процессы принятия решений при распознавании образов. // перевод под редакцией В.И. Иваненко, «Техника», Киев, 1965.

Попередня обробка зображень

Копитчук М.Б., Олещук О.В.

Одеський Національний політехнічний університет

65044, пр-т Шевченко 1.

тел. (0482)28-85-37, e-mail: logos104@rambler.ru

Анотація Розглядається спосіб підвищення вірогідності розпізнавання текстових символів, що нанесені на технологічні об'єкти, шляхом попередньої обробки зображення. Пропонується алгоритм, що дозволяє значно скоротити обчислювальні затрати на обробку зображення.

Вступ

Як першоджерела інформації для розпізнавання характерних ознак технологічних об'єктів, використовуються, як правило, відео- та фото зображення [1]. При цьому на вказаних зображеннях можуть бути присутні об'єкти довільного вигляду. тому перед безпосереднім аналізом зображених предметів необхідно виконати ряд операцій, що дозволяють отримати зображення тільки об'єктів визначеного типу. Існуючі методи обробки зображень [2]-[3] недостатньо враховують специфіку об'єктів спостереження та небажаних об'єктів, що попадають у поле зору.

В рамках даної статті обмежимось розгляданням зображень текстових символів, хоча розглянуті принципи обробки можуть бути застосовані і до деяких інших типів зображень.

Критерії формування характерних ознак

Для вирішення поставленого завдання потрібно визначити, які критерії класифікації можуть бути використані. Через те, що текст звичайно наноситься чорною чи білою фарбою, то можна очікувати, що найбільш темні чи світлі ділянки зображення є символами. На практиці за рахунок поганого освітлення чи наявності різних елементів конструкції та інших об'єктів, що попадають у поле зору, на зображенні часто присутні більш яскраві чи більш темні ділянки, ніж зображення тексту. З тієї ж причини деякі ділянки з текстом не є найбільш контрастними. З іншого боку, в областях, що безпосередньо містять текст, найбільш контрастними ділянками є переходи від зображення символу до фону зображення.

Також необхідно враховувати, що поверхня, на якій зображений текст, не є абсолютно однорідною, як за рахунок нерівномірного фарбування поверхні, так і за рахунок нерівномірного висвітлення.

Виходячи з вищесказаного, виходить, що одним з перших етапів повинне бути виділення фону зображення і вирівнювання яскравості і контрастності по всій області зображення. Для виділення фону необхідно знати колір символів, що складають напис (чорний чи білий). Через те, що на цьому етапі колір символів невідомий, необхідно зробити потрібні операції для обох кольорів, а вже на більш пізніх етапах виконати визначення кольору символів.

Виділення фону варто робити на невеликих ділянках (чим менше, тим краще, але так, щоб туди не потрапив символ цілком). Спочатку коротко опишемо основні перетворення над матрицями, що використовуються для виділення фону. Умовимося вважати, що код чорного кольору має більше значення, ніж код білого кольору.

1. Поширення максимуму на сусідні точки:

$$D = DMax(S, k_x, k_y)$$
$$D(x,y) = \max(S(u,v)) \qquad (1)$$

де

$$u = \overline{x - k_x/2 + 1, x + k_x/2 - 1},$$
$$v = \overline{y - k_y/2 + 1, y + k_y/2 - 1}.$$

Зміст перетворення *DMax* полягає в тому, що значення кожного елемента (x,y) матриці D встановлюється в максимальне значення матриці S, знайдене в околиці точки (x,y). Величина околиці для пошуку максимуму задається параметрами k_x, k_y.

2. Поширення мінімуму на сусідні точки:

$$D = DMin(S, k_x, k_y)$$

$$D(x,y) = \min(S(u,v)), \qquad (2)$$

де

$$u = \overline{x - k_x/2 + 1, x + k_x/2 - 1},$$

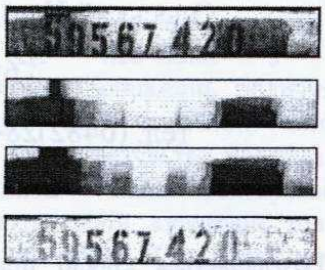
$$v = \overline{y - k_y/2 + 1, y + k_y/2 - 1}.$$

Дане перетворення аналогічне перетворенню *DMax*, за винятком того, що виділяє мінімальне значення в заданій області.

Якщо як розмір області задати значення, що перевищують очікувану товщину ліній символів і застосувати операцію *DMax* над матрицею M_0 із зображенням цифр білого кольору, то на отриманій матриці M_1 області, що раніше містили символи, будуть містити значення, що відповідають кольору фону. Однак при цьому спостерігається деяке збільшення більш темних ділянок зображення за рахунок зменшення світлих ділянок. Це найбільше помітно при яскраво вираженій неоднорідності фону зображення. Щоб позбутися цього небажаного ефекту, потрібно до матриці M_1 застосувати операцію *DMin* з тими ж параметрами. На отриманій матриці M_2 темні ділянки зображення повернуться до свого вихідного розміру, але на відміну від M_0 в матриці M_2 будуть відсутні світлі ділянки, розміром менше області пошуку мінімумів і максимумів. Якщо тепер з матриці M_2 відняти матрицю M_0, то буде отримане зображення темних символів на білому фоні.

Аналогічно, якщо до зображення з темними символами на світлому фоні послідовно застосувати операції *DMin* і *DMax*, а потім відняти отриману матрицю з вихідного зображення, то буде отримане зображення темних цифр на білому фоні (мал.1).

Зображення після видалення фону (мал.1г) містить темні ділянки розміром меншим, ніж $k_x * k_y$, де k_x і k_y – параметри в операціях *DMin* та *DMax*. Вибір параметрів здійснюється виходячи з очікуваної товщини ліній символів на зображенні. Якщо вибрати параметри k_x і k_y занадто великими, то після видалення фону зображення буде містити темні ділянки з фонової області. Чим більше значення параметрів k_x і k_y, тим більше помилкових ділянок фону залишиться на зображенні. При виборі занадто малих значень параметрів k_x і k_y околиці для пошуку мінімальних і максимальних значень виявляться настільки малими, що лінії символів будуть сприйняті як ділянки фону і будуть вилучені з зображення (мал.2).



Мал. 1. Видалення фону зображення: а – вихідне зображення, б – результат виконання операції DMin, в – результат виконання операції DMax над матрицею б, г – різниця зображень а і б.

Мал. 2. Виділення фону з різною величиною околиці пошуку: а – вихідне зображення, б – околиця 64x64, в – околиця 16x16, г – околиця 8x8.

Як правило, вказівка трохи більшого розміру області пошуку не так значно впливає на якість зображення, як область пошуку, менша, ніж товщина лінії символів. Це необхідно враховувати, якщо точно невідома товщина символів чи символи мають різну товщину ліній. У цьому випадку рекомендується вибирати параметри k_x і k_y виходячи з максимально можливої товщини лінії символів.

Після видалення фону зображення символи, як правило, мають різну яскравість. Це пов'язано як із різною яскравістю символів, так і з різною яскравістю фону на вихідному зображенні (світлі символи на темному фоні будуть яскравіше, ніж світлі символи на світлому фоні). У зв'язку з цим виникає завдання вирівнювання яскравості по всьому зображенню.

Для цього введемо операцію нормалізації області

$$D = NArea(S, k_x, k_y),$$

де

$$D(x,y) = (S(x,y) - M_{Min}(x,y)) / (M_{Max}(x,y) - M_{Min}(x,y)),$$
$$M_{Min} = DMin(S, k_x, k_y), \qquad (3)$$
$$M_{Max} = DMax(S, k_x, k_y).$$

Віднімаючи з матриці S матрицю M_{Min}, тобто мінімальні значення в області, отримаємо, що значення в ній будуть змінюватися від нуля і вище. У знаменнику знаходиться величина, що визначає різницю між максимальним і мінімальним значенням в області обраного розміру. Розділивши на цю величину, одержимо матрицю з діапазоном значень від нуля до одиниці. Необхідно передбачити дії при рівності M_{Min} і M_{Max} у деякій точці (x,y), щоб запобігти ділення на нуль. Можливі дії в цьому випадку:

Додати деяку постійну величину γ у знаменник формули (3).

$$D(x,y) = (S(x,y) - M_{Min}(x,y)) / (M_{Max}(x,y) - M_{Min}(x,y) + \gamma)$$

Недоліком цього способу є залежність результату від співвідношення величини γ і значень матриці S.

Інше рішення - у випадку рівності M_{Min} і M_{Max} у деякій точці (x,y) як значення матриці $D(x,y)$ вибрати деяку постійну величину μ:

$$D(x,y) = \begin{cases} Если \quad M_{Min}(x,y) \neq M_{Max}(x,y), & то \quad \dfrac{S(x,y) - M_{Min}(x,y)}{M_{Max}(x,y) - M_{Min}(x,y)} \\ Если \quad M_{Min}(x,y) = M_{Max}(x,y), & то \quad \mu \end{cases}$$

Вибір значення μ варто робити виходячи з особливостей матриці S. Якщо вона є зображенням загального вигляду, то рекомендується встановити μ, рівним 0.5, тобто середині діапазону значень матриці D. Якщо в матриці S попередньо був вилучений фон зображення, то в ній з усіх великих однотонних областей могли залишитися тільки області білого кольору, тобто області із значенням, рівним нулю, тому значення μ варто встановлювати рівним нулю.

Мал. 3. Нормалізація області: а – вихідне зображення, б – нормалізація областей розміром 4x4, в – розміром 8x8, г – розміром 16x16, д – розміром 64x16.

В результаті нормалізації області значення кожної точки (x,y) у матриці D буде залежати тільки від значень точок в області розміром $k_x * k_y$ з центром у точці (x,y). Параметри k_x і k_y вибираються виходячи з очікуваних розмірів символів на зображенні, відстані між ними, розташування символів стосовно інших об'єктів на зображенні (елементів конструкції, різного роду написів та ін.) і виду цих об'єктів. Вибір малих значень k_x і k_y приводить до того, що між символами можуть з'явитися помилкові темні ділянки і з'явиться велика кількість темних ліній у безпосередній близькості до символів, що зробить важчим подальше виділення символів із зображення. При виборі занадто великих значень k_x і k_y на якість зображення деяких символів можуть впливати яскраво виражені об'єкти, що знаходяться поблизу від символів, у тому числі сусідні символи.

При необхідності після виконання нормалізації області можна перейти до перетворення зображення в монохромне. Це дозволить спростити подальшу обробку і зменшити витрати пам'яті для збереження зображення. Для перетворення зображення в двоколірне необхідно вибрати граничне значення r, при перевищенні якого точка зображення вважається чорною і їй привласнюється код, рівний одиниці, інакше точка вважається білою з присвоєнням їй коду, рівного нулю.

Виконана послідовність перетворень дозволяє одержати зображення, що містить яскраво виражені об'єкти з заданими геометричними характеристиками. Результат виконання вищеописаних операцій дозволяє легко робити подальшу обробку на основі зв'язності точок, особливо якщо зображення, що спостерігається, містить послідовність недотичних об'єктів. Більш докладно про аналіз монохромних зображень і розпізнавання зображених на них об'єктів див. у [2], [4].

До недоліків розглянутих операцій можна віднести їх відносно велику обчислювальну вартість. Для усунення цього недоліку необхідно оптимізувати алгоритм їхньої реалізації.

Оптимізація операцій

Час виконання операції $DMax(S,k_x,k_y)$ при її виконанні для кожної точки матриці S окремо пропорційний $k_x k_y n_x n_y$,

де n_x, n_y – кількість точок матриці S по осях x та y.

Обчислювальні витрати при виконанні операції поширення максимуму в такому виді істотно залежать від параметрів k_x і k_y і є дуже значними при великих значеннях k_x і k_y.

Очевидно, що при обчисленні результату операції для двох близько розташованих точок доводиться порівнювати між собою велику кількість тих самих значень, тому доцільно виконувати цю операцію, використовуючи загальні проміжні результати.

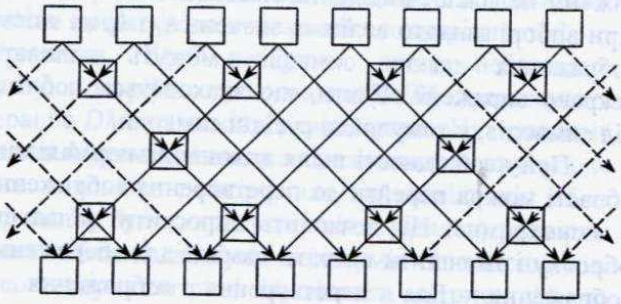
Для простоти розглянемо одномірний вектор V з числом елементів, рівним n,

де $n = 2^m$, $m = 1,2,3,...$

і реалізуємо для нього операцію поширення максимуму $DMax(V,k)$.

де $k = 2^p$, $p = 1,2,3,...$

Запропонована схема реалізації алгоритму показана на мал.4.



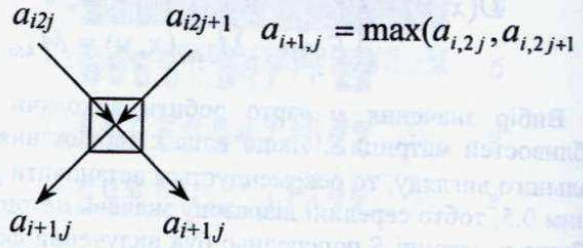
Мал.4. Схема реалізації операції поширення максимуму.

Алгоритм реалізації операції $DMax(V,k)$ умовно можна розділити на два етапи. На першому етапі здійснюється пошук максимумів серед сусідніх елементів векторів.

На першому кроці попарно порівнюються сусідні елементи вектора V, і максимальне значення з кожної пари записується у вектор V_1, що має число елементів у два рази менше, чим вектор V. Потім ті ж дії виконуються над вектором V_1 з одержанням вектора V_2 і т.д. доти, поки не буде отриманий вектор V_q,

де $q = \log_2(k) - 1$.

На другому етапі виконується порівняння елементів векторів різного рівня. Спочатку кожен елемент вектора V_q порівнюється з двома елементами вектора V_{q-1}, що є сусідніми з елементами, з яких отримане значення поточного елемента вектора V_q. Далі отриманий вектор V_{q+1} аналогічно поелементно порівнюється з вектором V_{q-2} і т.д. доти, поки не буде отриманий вектор V_{2q}, що і є шуканим вектором.



Мал.5. Умовне позначення одиничної операції поширення максимуму

Якщо необхідно виконати операцію поширення максимуму, циклічно замикаючи краї матриці, то використовуються зв'язки, зазначені пунктиром. У противному випадку пунктирні зв'язки відкидаються.

Швидкість виконання операції поширення максимуму при використанні оптимізованого алгоритму пропорційна $\log(k_x)\log(k_y)n_x n_y$.

Висновки

Запропонований спосіб попередньої обробки зображень символів, нанесених на технологічний об'єкт, дозволяє поліпшити якість зображення, за рахунок чого вірогідність розпізнавання символів зростає до 85%.

Алгоритм оптимізації запропонованих операцій при обробці областей зображень розміром 16x16 дозволяє скоротити час обробки зображень у 16 разів, а при розмірі областей 32x32 – у 40 разів.

Література

1. Копитчук М.Б., Олещук О.В. "Использование метода развертки изображения для распознавания символов" – Праці УНДІРТ, 2001, №3(27). – С. 63 – 67.
2. Дуда Р., Харт П. "Распознавание образов и анализ сцен" – М.: Мир, 1976.
3. Василенко Г.И. Тараторин А.М. "Восстановление изображений" – М.: Радио и связь, 1986.
4. Вапник В.Н., Червонекис А.Я. "Теория распознавания образов" – М.: Наука, 1974.

ГІСТОГРАМНИЙ МЕТОД РОЗПІЗНАВАННЯ РУКОПИСНОГО ТЕКСТУ І СИМВОЛІВ

Олександр Ільченко
Українська Академія Друкарства
79039 м. Львів вул. Підголоско 19 тел.59-94-20 e.mail:alex_i2002@ukr.net

Абстракт. У роботі описаний розроблений гістограмний метод і сформульовані основні принципи побудови системи розпізнавання рукописного тексту і символів.

1. ВСТУП

Розпізнавання образів і, зокрема, розпізнавання рукописного тексту і символів на даний час дуже актуальна науково-технічна проблема. Розв'язання її значно полегшить взаємодію людини із сучасною обчислювальною технікою, комп'ютерами, телекомунікаційним обладнанням, тощо. Відомо, що, як правило, введення інформації у комп'ютерну техніку здебільшого виконується вручну. Це займає багато часу і потребує великої кількості ресурсів.

2. РОЗПІЗНАВАННЯ РУКОПИСНОГО ТЕКСТУ І СИМВОЛІВ

Рукописним текстом вважаємо сукупність слів, окремих літер, чисел, цифр, розділових знаків написаних від руки, кириличним або латинським алфавітом. До рукописного тексту відноситься не тільки рукописний текст, як ми його уявляємо, а й текст написаний від руки друкованими літерами, а також текст, написаний від руки стилізованими літерами або цифрами, наприклад, на поштових конвертах.

Рукописні символи це більш широке поняття, у яке входять усе вищеперелічене, а також спеціальні символи, наприклад, математичні, тощо.

3.ОГЛЯД МЕТОДІВ РОЗПІЗНАВАННЯ РУКОПИСНОГО ТЕКСТУ І СИМВОЛІВ

Відомо декілька методів розпізнавання рукописного тексту і символів. Докладніше зупинимося на трьох із них, а власне на методі вертикального сканування, хвильовому методі знаходження скелета растрового зображення, і на методі, так званого "фонтанного перетворення" (від англійського **font** – шрифт).

Принцип дії цих методів розпізнавання рукописного тексту і символів полягає у зведенні рукописного тексту і рукописних символів до спрощеного, стандартизованого вигляду.

Хвильовий метод знаходження скелета растрового зображення базується на принципі відслідковування за проходженням сферичної хвилі по контуру растрового зображення. Докладно цей метод описаний у [1].

Метод вертикального сканування викладений у [2] та полягає у побудові стандартизованого скелету растрового зображення окремих рукописних літер, шляхом пошуку екстремальних значень контура зображення.

Метод "фонтанного перетворення" заснований на принципі структурно-плямових еталонів, і використовується у широко відомій системі розпізнавання рукописного тексту "Fine Reader Рукопис", компанії ABBYY, яка широко використовується для обробки анкет Пенсійного фонду та Податкової інспекції [3,4].

Основним недоліком вищеперелічених методів є те, що вони мало придатні дла розпізнавання рукописного тексту, складні в практичній реалізації, потребують значних системних ресурсів.

Розроблений автором гістограмний метод розпізнавання рукописного тексту і символів, який описаний нижче, не має вищеперелічених недоліків.

4. ГІСТОГРАМНИЙ МЕТОД РОЗПІЗНАВАННЯ РУКОПИСНОГО ТЕКСТУ І СИМВОЛІВ

Гістограмний метод розпізнавання рукописного тексту і символів заснований на побудові графіка (гістограми) залежності кількості інформативних пікселів матриці бінарного растрового зображення, відкладеної по осі ординат, від координати рядка або стовпця, відкладеної по осі абсцис (рис.1) та порівнянні її з набором еталонних гістограм навчальної вибірки.

Під бінарним растровим зображенням розуміється двовимірна матриця з чорних та білих точок, в якій об'єкт задається чорними пікселями растру, а фон – білими. Під інформативними пікселями розуміються чорні пікселі растру.

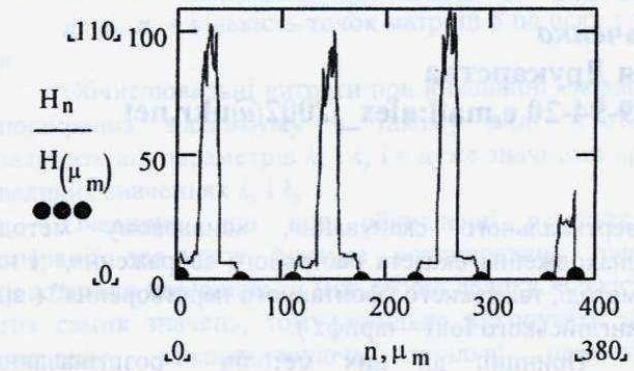


Рис.1. Гістограма рукописного тексту, наведеного на рис.2.

На рис.1 наведена гістограма тексту "В останні роки розпізнавання образів знаходить усе більше застосування в повсякденному житті", зісканованого із роздільчою здатністю 300 dpi (рис.2).

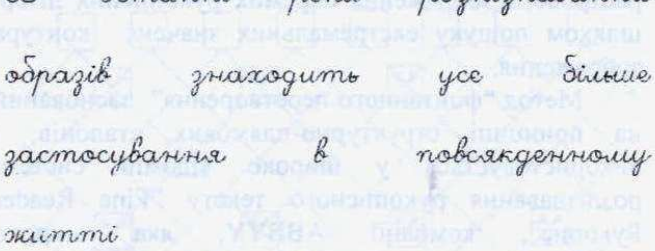


Рис.2. Рукописний текст, гістограма якого наведена на рис.1.

Гістограмний метод розпізнавання рукописного тексту і символів дозволяє із 100%-ю вірогідністю розділяти рядки у суцільному тексті і окремі слова, літери, цифри і символи у рядку, шляхом знаходження локальних мінімумів на гістограмі (рис.3,4).

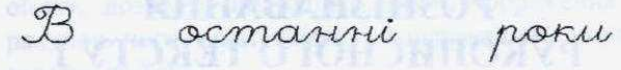


Рис.3. Зразок рукописного тексту для ілюстрації розділення слів і окремих літер у рядку.

На рис.4 наведена ілюстрація розділення слів у рядку рукописного тексту, знаходженням локальних мінімумів на гістограмі. Локальні мінімуми на гістограмі позначені крапками.

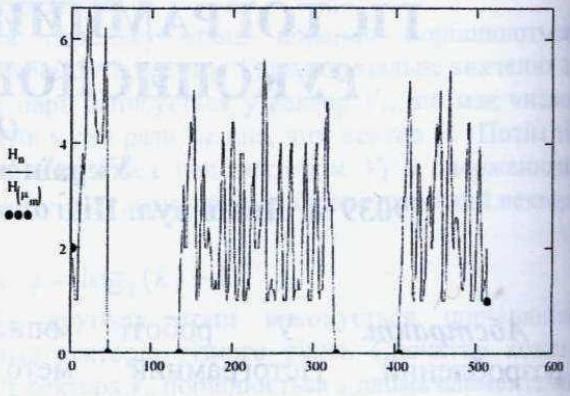
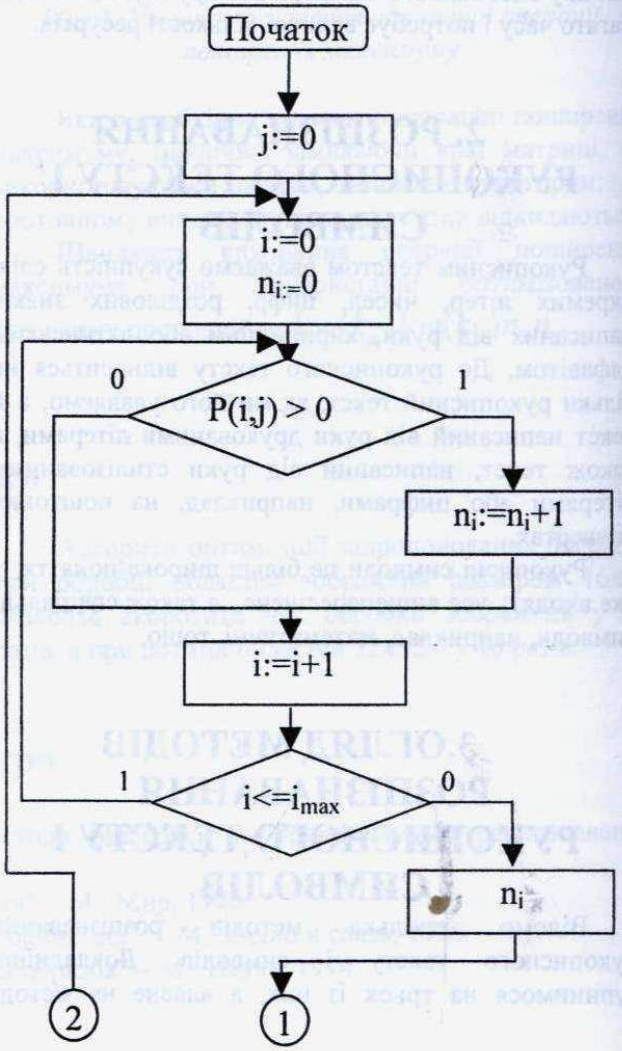


Рис.4. Гістограма окремих слів тексту, наведеного на рис.3.

5. АЛГОРИТМ ПОБУДОВИ ГІСТОГРАМИ БІНАРНОГО РАСТРОВОГО ЗОБРАЖЕННЯ

На рис.5 наведена блок-схема алгоритму побудови гістограми бінарного растрового зображення.



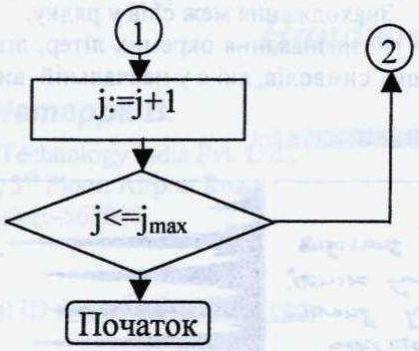


Рис. 5. Блок-схема алгоритму побудови гістограми бінарного растрового зображення.

Де

n_i - кількість інформативних пікселів у рядку матриці бінарного растрового зображення;

i - номер (координата) рядка матриці;

j - номер (координата) стовпця матриці;

P(i,j) – координата кожного пікселя матриці;

У блоках 1 та 2 блок-схеми алгоритму побудови гістограми бінарного растрового зображення відбувається початкова ініціалізація змінних.

У блоці 3 перевіряється чи є кожен піксель матриці інформативним, тобто чорним.

Якщо піксель є інформативним, тоді до кількості інформативних пікселів у i – му рядку матриці - n_i додається одиниця (блок 4).

У блоці 5 відбувається перехід на наступний рядок матриці.

У 6-му блоці перевіряється чи даний рядок не є останнім у матриці.

Виводиться значення кількості інформативних пікселів у i – му рядку матриці бінарного растрового зображення (блок 7).

У блоці 8 відбувається перехід на наступний стовпець матриці.

У 9-му блоці перевіряється, чи даний стовпець не є останнім у матриці. Гістограмний метод розпізнавання рукописного тексту і символів заснований на побудові графіка (гістограми) залежності кількості інформативних пікселів матриці бінарного растрового зображення, відкладеної по осі ординат, від координати рядка або стовпця, відкладеної по осі абсцис (рис.1) та порівнянні її з набором еталонних гістограм навчальної вибірки.

6. СИСТЕМА РОЗПІЗНАВАННЯ РУКОПИСНОГО ТЕКСТУ І СИМВОЛІВ ІЗ ВИКОРИСТАННЯМ ГІСТОГРАМНОГО МЕТОДУ

Система розпізнавання рукописного тексту і символів (рис.7) складається із сукупності програмних і апаратних засобів. Апаратна частина системи стандартна, і складається із комп'ютера, сканера і принтера і в межах даної статті розглядатися не буде.

Загальний алгоритм роботи системи розпізнавання рукописного тексту і символів приведений на рис.6.

Рис.6 Блок-схема алгоритму роботи системи.

1. Введення навчальної вибірки шляхом сканування необхідного набору літер, цифр,

розділових знаків, символів тощо, введення їх значення з клавіатури.

2. Побудова гістограми кожного символу навчальної вибірки і присвоєння їй введеного з клавіатури значення.

3. Введення зі сканера тексту, який буде розпізнаватися.

4. Побудова гістограми тексту, який буде розпізнаватися.

5. Знаходження меж рядків у тексті.

6. Знаходження меж слів у рядку.

7. Розпізнавання окремих літер, літер у словах та інших символів, які є у навчальній вибірці.

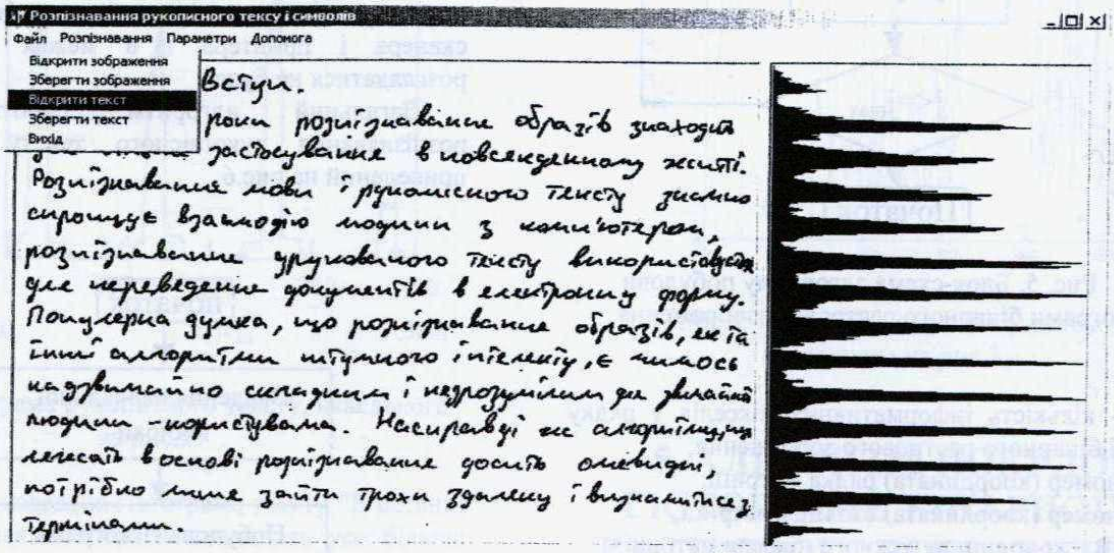


Рис. 7. Інтерфейс системи розпізнавання рукописного тексту і символів

8.Перевірка правильності розпізнавання слів з допомогою вбудованого словника.

Запис розпізнаного тексту у файл, або вивід на друк.

Розроблена математична модель системи у середовищі MathCad 2001Pro.

Фрагмент програми на внутрішній мові програмування середовища MathCad 2001Pro.

Знаходження меж рядків у тексті і меж слів у рядку відбувається шляхом знаходження локальних мінімумів на гістограмах. Ці локальні мінімуми і є межами рядків або слів (рис.7).

Розпізнавання літер у словах відбувається так. Будується гістограма слова. Від цієї гістограми відтинається відрізок, ширина якого дорівнює мінімальній ширині літери. Гістограма цього відрізку порівнюється зі всіма гістограмами символів навчальної вибірки. При збігу гістограм, гістограмі відрізку, яка відповідає першій літері слова присвоюється значення символу гістограми навчальної вибірки із якою вона збіглася. Якщо відрізок мінімальної ширини не збігся ні з одною гістограмою навчальної вибірки, тоді до відрізку мінімальної ширини додається певний крок, будується нова гістограма відрізку і вся процедура порівняння нової гістограми із набором гістограм символів навчальної вибірки повторюється. Так відбувається доти, поки не відбувається збігання гістограми відрізку із набором гістограм символів навчальної вибірки. Після цього береться наступний відрізок мінімальної ширини і всі дії повторюються знову. Так відбувається доти, поки не буде розпізнано останню літеру слова.

Після цього відбувається перевірка правильності розпізнавання слів шляхом порівняння набору літер розпізнаного слова із вбудованим словником ситеми розпізнавання.

7. ВИСНОВКИ

1.Розроблений метод розпізнавання рукописного тексту і символів, який грунтується на на побудові гістограми тексту, щорозпізнається, і порівнянні її з набором еталонних гістограм навчальної вибірки.

2.Сформульовані основні принципи побудови системи розпізнавання рукописного тексту і символів із використанням розробленого методу.

7. ЛІТЕРАТУРА

1.Применение волнового алгоритма для нахождения скелета изображения. *http://www.ocrai.narod.ru/vectory.html*

2.Макуцький Я., Овсяк В. Алгоритми пошуку екстремальних точок і побудови стандартизованого скелету символів рукописного тексту.//Комп'ютерні технології друкарства. Збірник наукових праць №3. Львів, 1999р. с. 90 – 97.

3.Абраменко А. Компьютер читает. *http://www.ocrai.narod.ru/fr.html.*

Novel Segmentation and Feature extraction Techniques for Handwritten Text Recognition.

Hemappa B.
Intel Technology India Pvt. Ltd.,
#136, 3rd Floor, Airport Road,
Bangalore-560017
India.

Email ID :hems_bs@hotmail.com

Dr.N.V.Subba Reddy
Professor and Head,
Department of Computer Science
Manipal Institute of Technology
MAHE (Deemed University)
Manipal-576 119, India.
Email Id: dr_nvsreddy@rediffmail.com

Abstract: *In this paper, we proposed novel slant detection and correction techniques, segmentation and feature extraction methods for handwritten text recognition. The segmentation and feature extraction module works on data, which is to be preprocessed. This is necessary because of the output of the scanned document is noisy and usually contains too many unwanted impression happened due to humidity and uneven ink distribution on paper. The novel techniques are used to detection and correction of any rotation that may occur during the writing or scanning process. The slant correction algorithm then generates a non_slant image by rotating the blocks, rather then the individual pixels. The proposed segmentation algorithm will count the block pixels in horizontal lines of the binaraized image and estimate the width and base line of each text line in the document. This ensures easy to segment the document into text of sentences. Each sentence again segmented into words. Each word processed and extract the prominent feature for further classification. The proposed system based on neural network feature extraction gives promising results in terms of classification and recognition.*

Keywords: *pattern recognition, Preprocessing, Trimming, Segmentation, Feature extraction, Slant detections, Slant correction, Neural network.*

1. Introduction

The area of handwriting recognition can be divided into online and offline [2,3]. In online recognition the writer is connected via an electronic pen or a mouse to the computer and the handwriting is recorded as a function of time. By contrast, in offline recognition the handwriting is captured by means of a scanner and becomes available for processing and recognition in the form of an image. Because of the availability of temporal information, online recognition is often considered the easier problem [1]. Automatic processing of handwritten forms requires form analysis [4, 5], field extraction [6,7], handwritten recognition, and so on. Texts/Lines separation has two non-trivial problems: one is to detect and remove the lines, and the other is to restore some character strokes distorted by the deleted lines [8]. A number of methods have been reported for sentences/lines separation [9, 10], but most of them can handle lines with in ±5° because of computational burden of their algorithm and fails in some exceptional situations where a straight line is curved a little or has a non-uniform thickness.

2. Developed system

The structure of the developed text recognition system shown in fig 1. Notice that both the segmentation and feature extraction modules works on data, which is preprocessed. This is necessary because of the output of the scanned document is noisy and usually contains too many unwanted impression happened due to humidity and uneven ink distribution on paper. The novel techniques are used to detection and correction of any rotation that may occurred during the writing or scanning process. The volume of the data to processed can be reduced and prominent features of patterns (Texts) can be extracted by incorporating trimming and removal of underlines, by which efficiency of preprocessor improves and it will have influence on the over all performance of the pattern recognition system(11). The proposed segmentation algorithm will count the block pixels in horizontal lines of the binaraized image and estimate the width and base line of each text line in the document. So that is easy to segment the document into text of sentences. And each sentence again segmented into words. Each words processed and extract the prominent feature for further classification and recognition system and to improve the accuracy of the classification, we will propose neural network based feature extraction system. The proposed

methods work well for multi lingual language documents and pattern recognition system.

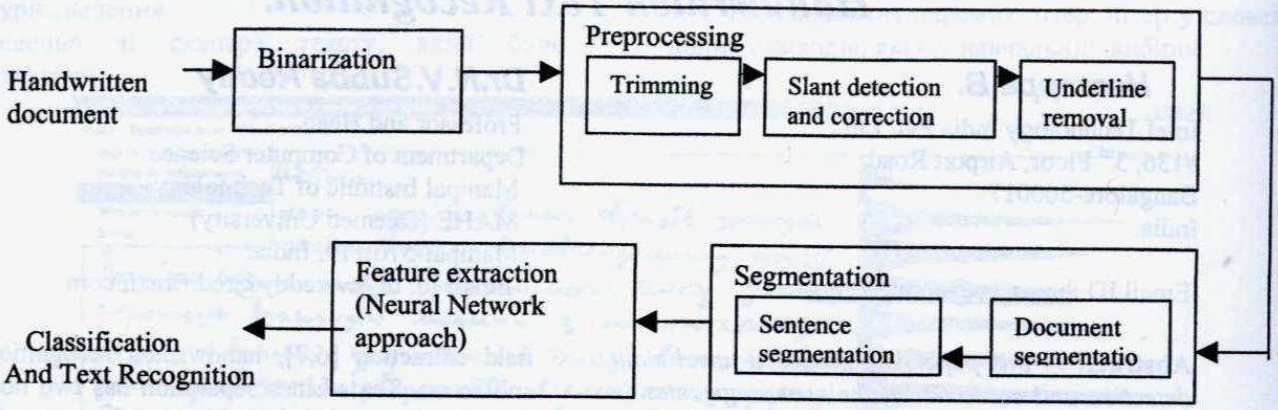


Fig 1. Overview of proposed model for printed and handwritten text recognition system.

3. Preprocessing and Normalization

The preprocessing model takes as input data and gives as output pattern containing the word to be recognized without any other element irrelevant to the recognition process. The tasks performed at this level depend on the data: when the document from which the word is extracted presents a background pattern, the latter should be removed. In some other cases, the words are extracted from forms showing boxes and lines that should be eliminated. For our data, binarization, trimming[11] the pattern to remove the unwanted impression happened due to humidity and uneven ink distribution on paper.

Once the image is preprocessed, it must be normalized. The normalization is deals with removing slant and slope of the document, line of texts and words. The slant detection and correction of the document images is particularly crucial among the document processing operations. Following algorithms explains the slant detection and corrections technique of document , sentence and words.

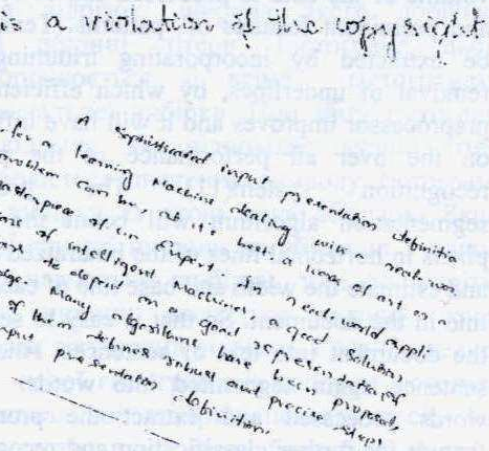


Fig 2. Sample inputs

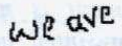


Fig 3. Slanted input

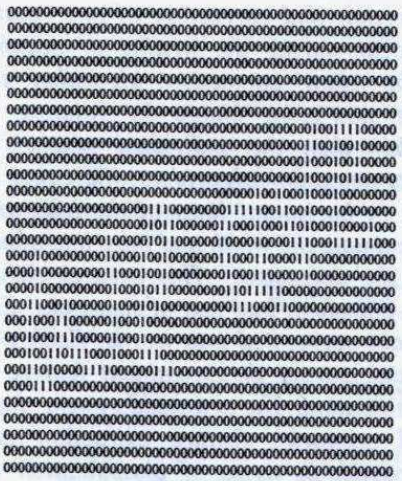


Fig.4 After Binarization

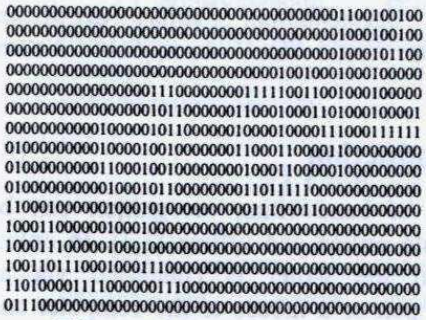


Fig 5 After Trimming

3. 1. Slant Detection Algorithm

The Slant detection algorithm takes preprocessed data as input and gives the degree of the slant towards right or left of the image.

Algorithm:

1. Initialize the TopLeft, TopRight, BottomLeft, BottomRight flags to zero.
2. Scan the 1/8th of binary pattern row wise from top to bottom.
3. if rows contains continues 0's
   A[I][j] = 1;
   Else
   A[I][j] = -1;
   Where I=0 … int(M x 1/8)
   J=0…N
   M= height of the pattern
   N = width of the pattern.
4. if each row of array A contains 1,s in left side , Set TopLeft =1 else TopRight =1;
5. Scan the 1/8th of binary pattern row wise from bottom to top.
6. Initialize the Array A to 0, Repeat the Step 3.
7. if each row of array A contains 1,s in left side , Set BottomLeft =1 else BottomRight =1;
8. if TopLeft = 1 and BottomRight =1 then Return LeftDownward Slant.
   Else if TopRight = 1 and BottomLeft =1 then return RightDownward Slant
   Else if((TopLeft=1 & bottomRight =0)or(TopRight=1 & BottomLeft =0) then return Partial slant.
   Else return no slant.

3.2 Slant correction

Algorithm :

1. Assign the pattern into M x N matrix
2. if (LeftDownward slant)

```
for(I=0;I<M*0.5;I++){
  for(j=0;j<N;j++){
    if(patten[I][j]! = 1)
      pattern[I][j]=pattern[I+1][j];
    else
    break;
  }
}
```

3. else if(RightDownward slant)

```
for(I=0;I<M*0.5;I++){
  for(j=N;j>0;j--){
    if(patten[I][j]! = 1)
      pattern[I][j]=pattern[I+1][j];
    else
    break;
  }
}
```

```
000000000000000001110000000011111001000110011110
000000000001000001110000000011111001000110010010
010000000001000001110000000011111001000100010010
010000000001000001110000000011011001000100010110
110000000001000001110000000010001001000100010000
110000000001000001011000000010001000011100011111
110001000000100010100000000001110001100000000000
100011000000100010000000000000000000000000000000
100001110000010001000000000000000000000000000000
100110111000100011100000000000000000000000000000
110100001111000000111000000000000000000000000000
011100000000000000000000000000000000000000000000
```

Fig 6. After Slant correction.

4. Segmentation

Text regions posses a unique texture because they typically follow a specific arrangement rule: each regions consists of text line of the same orientation with approximately the same spacing between them and each text line consists of characters of approximately the same size. The above observations suggest that the primary component of the page layout – Text lines can be discriminated using the following novel segmentation algorithm.

4.1 Page segmentation.

Algorithm:

1. Scan the binarized data row wise.
2. Count the number of on cells in each row.
   D[i]= OC;
   Where I = 0…R
   OC =Total number of on Cells in a row.
   R =Total number of rows in page.
3. Mark the minimum or no on cells rows.
   If (D[I] = 0 || D[I]< Minimum)
   M[j]=I;
   J= 0 …R/4;
4. Find the width between the lines and separate the line of text.

4.2 Segmenting the text lines into words

To find the word boundaries within a text, We first set the distance between each words in a text line. Scan the binarized text line pattern vertically to identify the space between words and divide the text line into individual words to use for feature extraction. Preprocessed text line gives as input to the text line segmentation, which trimmed and underline removed pattern.

1. Set the space between the words
2. Scan the text line pattern column wise
3. if the space between words is equal to assumed space. Then divide the word.

```
00000000000000000001110      01111100100011001111100
00000000000100000001110      01111100100011001001000
01000000000100000001110      01111100100010001001000
01000000000100000001110      01101100100010001011000
11000000000100000001110      01000100100010001000000
11000000000100000001011      10000100001110001111111
11000100000010001010000      00111000110000000000000
10001100000010001000000
10001110000010001000000
10011011100010001110000
11010000111100000011100
01110000000000000000000
```

Fig 7. segmented and trimmed words

5. Feature extraction

In most systems, the word image is segmented into small parts supposed to be the basic information units. In our feature extraction, preprocessed whole word is segmented into 16 numbers of regions, extraction of the feature from each region and generates the feature

vectors. This feature vectors are especially useful in classification of words using neural network system.

Preprocessed pattern divided into 16 numbers of regions.

Feature value = on cells /total number of on and off cels in a regions.

Feature vector = { G1, G2,G3,G4,G5,G6, D1, D2,D3,D4,D5,D6, M1,M2,M3}

1. Divide the preprocessed pattern into number of regions.

2.find the total number of On cells in each regions

3.Find the total number of ON and OFF cells in a region,

4.Calculate the feature value

Feature value = Total number of ON cell in a region/Total number of ON and OFF cells.

Repeat the step 4 for all the regions and generate the feature vector.

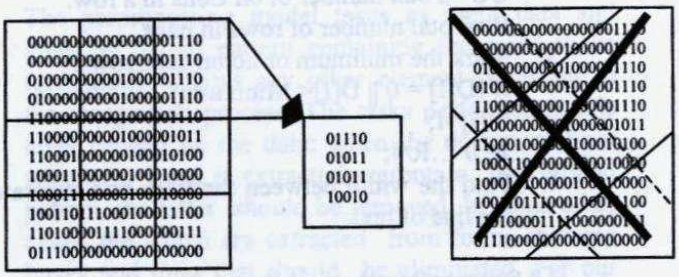


Fig 8. Segment the word into Grids(G1,G2..G6).

Doted lines segments the word as D1,D2,D3 regions, Solid lines divide as D3,D4,D5 regions

Solid Thick line segments M1,M2,M3,M4. regions

Conclusion:

The segmentation and feature extraction mechanisms developed are very useful in the recognition of text. The segmentation and feature extraction are two main components of the preprocessor. The overall performance of the text recognition system will have influence on better use of these two components. In this work proper care has been taken to improve the overall development of the text recognition system.

Reference:

1. U.V.Marti and H. Bunke, "*Using A Statistical Language Model to Improve the Performance of an HMM-Based Cursive Handwriting Recognition System*," Inter Journal of pattern recognition and Artificial Intelligence, Vol.15, No. 1(2001) pp.65-90.

2. I Guyon, M. Schenkel and J.Denker, "*Overview and synthesis of on-line cursive handwriting recognition techniques*," Handbook of character Recognition and document Image Analysis, eds H.Bunke and P.S.P Wang, World Scientific, 1997, Chap 7, pp. 227-258.

3. J.C.Simon, "*Off-line cursive word Recognition*", Proc. IEEE80, 7(1992) 1150-1161.

4. A. Dengel and G.Barth, High level document analysis guided by geometric aspects, Int'l J. Pattern Recognition and Artificial Intelligence, Vol. 2, No. 4, pp 641-655, 1988.

5. T. Watanable, Q. Luo, and N.Sugie, "*Layout recognition of multi-kinds of table-form documents*", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol 17, no. , pp. 432-445, 1995.

6. F. Cesarini, M. Gori, S. Marinai and G.Soda, Data extraction from form images, Proc. Conference DEXA 95, pp. 438-448, London(UK), September 1995.

7. L.Y. Tseng and R.C. Chen, "*Recognition and data extraction of form documents based on three types of line segments, Pattern Recognition*", Vol.31,no. 10, pp. 1525-1540, 1998.

8. S.H Kim, S.H. JEONG, H.K. Kwag , "*Line Removal and Character Restoration using Bag Representation of Form Images*" ,Proceedings of seventh international Workshop on frontiers in handwritten recognition, Sept-2000, Amsterdam, ISBN 90-76942-01-3, Nijmegen: International Unipen Foundation, pp 43-52.

9. B. Yu and A.K. Jain, *A Generic system for form dropout*, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 18,no 11, pp. 1127-1131, 1996.

10. M.D. Garris, "*Method and evaluation of character stroke preservation on handprint recognition*", Technical Report NISTIR 5687, July 1995.

11. Hemappa B, Dr.N.V.Subba Reddy, "*Preprocessing Technique for Handwritten Pattern Recognition*," National Conference on Intelligent and Efficient Electrical system, January 19-20,2001. No. II-5.

ІНВАРІАНТНА МОДЕЛЬ ОБРОБЛЕННЯ ТА РОЗПІЗНАВАННЯ ІЗОПЛАНАТИЧНИХ ЗОБРАЖЕНЬ

В. В. Гнатушенко

Дніпропетровський національний університет

49050, Дніпропетровськ, пров. Науковий , корпус 12

E-mail: vvgnat@ukr.net

Тел. (0562) 43-36-30. Факс (0562) 46-55-23

АНОТАЦІЯ

Робота присвячена побудові інформаційно-геометричної моделі форм ізопланатичних зображень.

1. ВСТУП

Як відомо [1], дифракційні явища відіграють важливу роль при роботі оптичних приладів, призначених для одержання зображень об'єктів. Зображення, отримані за допомогою фокусувальних пристроїв (лінз), ніколи не відтворюють об'єкт з ідеальною точністю. Вони бувають спотворені унаслідок усякого роду недосконалостей оптичних систем (аберацій). Але навіть ідеальна лінза, вільна від аберацій, не може дати ідеального зображення через хвильову природу світла. Унаслідок дифракційних явищ порушуються закони геометричної оптики та відображення точок об'єкта в точки зображення перестає бути проективним. В термінах теорії хвильових процесів це означає, що зображення перестає бути стигматичним. Таким чином, виникає важлива задача відновлення й ідентифікації нестигматичних зображень, які не охоплюються розвиненою в [2, 3] інформаційно-геометричною моделлю.

2. ІНВАРІАНТНА ГЕОМЕТРИЧНА МОДЕЛЬ ПОДАННЯ ІЗОПЛАНАТИЧНИХ ЗОБРАЖЕНЬ

2.1. ОСНОВНІ СПІВВІДНОШЕННЯ ФОРМУВАННЯ ІЗОПЛАНАТИЧНИХ ЗОБРАЖЕНЬ

Розглядається некогерентне джерело з інтенсивністю випромінювання $I_0(\mathbf{x})$, яке розташоване у предметній площині (некогерентність означає відсутність статистичного зв'язку між випроміненнями різних точок джерела). Випромінення від елемента поверхні об'єкта, розташованого у точці $\mathbf{x}$, поширюється через середовище і падає на лінзу діаметра D, центр якої розташовано у точці $\mathbf{q}$ фотограмметричної системи координат XYZ. Зображення з інтенсивністю $I(\mathbf{r})$ спостерігається в площині зображення, паралельній площині лінзи на відстані d за нею (рис. 1), причому $\mathbf{r}$ відлічується від головної точки зображення A — основи перпендикуляра з точки $\mathbf{q}$ на площину зображення. Через $\boldsymbol{\rho}$ позначено радіус-вектор точки в площині лінзи.

Як відомо з теорії хвильових процесів, у загальному випадку інтегральний зв'язок між інтенсивністю зображення $I(\mathbf{r})$ та інтенсивністю випромінення точок об'єкта $I_0(\mathbf{x})$ має вигляд [3]:

$$I(\mathbf{r}) = \int_{(\Omega)} I_0(\mathbf{x}) P(\mathbf{r},\mathbf{x}) d\mathbf{x}, \qquad (1)$$

де $P(\mathbf{r},\mathbf{x})$ — функція розмиття точки (ФРТ). Інтегрування ведеться по області розташування джерела (поверхні об'єкта).

У загальному випадку $P(\mathbf{r},\mathbf{x})$ є функцією двох змінних $(\mathbf{r},\mathbf{x})$, але, якщо довжина хвилі випромінення λ суттєво менша за характерні розміри іконічної системи та відстані, з якої формується зображення, ФРТ з високою точністю може бути подана у вигляді $P(\mathbf{r},\mathbf{x}) = P(\mathbf{r}-\mathbf{x}')$ (ізопланатичне наближення [4]), де $\mathbf{x}' = -\frac{d}{q}\mathbf{x}$. З урахуванням цього функція яскравості ізопланатичного зображення має вигляд:

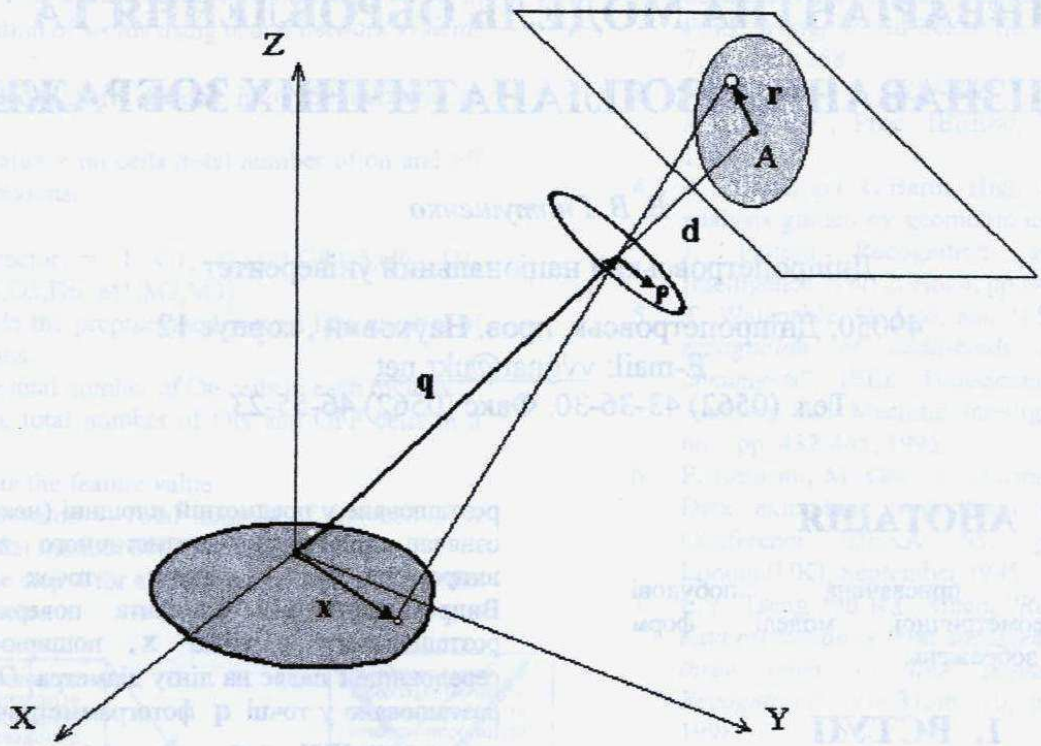


Рис. 1 Формування зображення в дальній фотограмметрії

$$I(\mathbf{r}) = \frac{1}{\lambda^4 q d^3} \int\limits_{(\Omega)} I_0\left(-\frac{q}{d}\mathbf{x}'\right) P(\mathbf{r}-\mathbf{x}')d\mathbf{x}', \quad (2)$$

де ФРТ для круглої лінзи з апертурою діаметра D

$$P(\mathbf{r}-\mathbf{x}') = \frac{\pi D^2}{4} \int \exp\left[\frac{jk\boldsymbol{\rho}_d}{d}(\mathbf{r}-\mathbf{x}')\right] \times \times \exp\left[\frac{-jk\mathbf{q}\boldsymbol{\rho}_d}{q}\right] K\left(\frac{\rho_d}{D}\right) d\boldsymbol{\rho}_d \quad (3)$$

Через $K\left(\frac{\rho_d}{D}\right)$ у виразі (3) позначено функцію

$$K\left(\frac{\rho_d}{D}\right) = \begin{cases} \frac{2}{\pi}\left[\arccos\left(\frac{\rho_d}{D}\right) - \left(\frac{\rho_d}{D}\right) \times \sqrt{1-\left(\frac{\rho_d}{D}\right)^2}\right], & \rho_d \le D \\ 0, & \rho_d > D, \end{cases} \quad (4)$$

яка дорівнює площі перетину двох кругів діаметра D з відстанню між центрами ρ_d.

Зазначимо, що у випадку $D \to 0$, $K\left(\frac{\rho_d}{D}\right) \to \delta(\rho_d)$ і відповідність точок об'єкта й зображення описується виразом $\mathbf{r} = -\frac{d}{H}\mathbf{x}$, ($H$ — висота центра проеціювання над предметною площиною), що збігається із загальним виразом для стигматичних зображень [2, 4].

2.2. ІНВАРІАНТНІ БАГАТОВИДИ ПОДАННЯ ГЕОМЕТРИЧНИХ ФОРМ ІЗОПЛАНАТИЧНИХ ЗОБРАЖЕНЬ

Інформаційно-геометрична модель подання зображень проекційної природи, розвинена в [5], будується у просторі інформаційних характеристик — безрозмірних семіінваріантів, які визначаються на основі інтегрального перетворення Фур'є відповідних функцій яскравості. З метою узагальнення моделі на випадок ізопланатичних зображень, виконаємо перетворення Фур'є від виразу (2), який пов'язує функцію яскравості зображення та власну яскравість точок об'єкта. Врахуємо при цьому, що цей вираз має вигляд згортки власної яскравості точок об'єкта $I_0(\mathbf{x})$ й функції розмиття точки $P(\mathbf{r})$. Зі спектрального аналізу відома теорема, у відповідності з якою перетворення Фур'є від згортки двох функцій

дорівнює добутку перетворень Фур'є від них [6]. З урахуванням цього одержуємо:

$$\mathrm{F}(\mathbf{u})=\frac{\pi D^2}{q^2 d^2}\mathrm{F}_0\left(-\frac{d}{q}\mathbf{u}\right)\mathcal{P}(\mathbf{u}), \qquad (5)$$

де $\mathrm{F}(\mathbf{u}), \mathrm{F}_0(\mathbf{u})$ — відповідно перетворення Фур'є від функції яскравості зображення та власної яскравості об'єкта. Через $\mathcal{P}(\mathbf{u})$ позначено перетворення Фур'є від ФРТ (3):

$$\mathcal{P}(\mathbf{u})=\frac{\pi D^2}{4}\left(\frac{\lambda d}{2\pi}\right)^2 K\left[\frac{\lambda d}{2\pi D}\sqrt{u_1^2+u_2^2}\right]. \qquad (6)$$

На підставі цього зв'язок між семіінваріантами зображення S_{nm}, скалярної яскравості об'єкта $S_{nm}^{(0)}$ та функції розмиття точки $\mathbf{s}_{nm}$ має вигляд:

$$S_{nm}=(-1)^{n+m}\left(\frac{f}{q-f}\right)^{n+m} S_{nm}^{(0)}+\mathbf{s}_{nm}, \qquad (7)$$

де f — фокусна відстань лінзи.

У відповідності з принципом формування інформаційних характеристик зображень, які описуються семіінваріантами парного порядку $N=n+m$ [2, 5], маємо:

$$Z_i=\frac{S_{i-1,N+1-i}}{F^{N/2}(\mathbf{0})}, i=\overline{1,N+1}. \qquad (8)$$

З урахуванням співвідношення (7) зв'язок між векторами подання двох зображень $\mathbf{Z}^{(1)}$, $\mathbf{Z}^{(2)}$, отриманих при різних позиційних умовах їхньої фіксації, набуває вигляду:

$$Z_i^{(2)}=\left(\frac{q_1-f}{q_2-f}\right)^{n+m}\left(\frac{M_{00}^{(1)}}{M_{00}^{(2)}}\right)^{\frac{n+m}{2}} Z_i^{(1)}+ \frac{\mathbf{s}_{i-1,N+1-i}^{(2)}-\left(\frac{q_1-f}{q_2-f}\right)^{n+m}\mathbf{s}_{i-1,N+1-i}^{(1)}}{\left(M_{00}^{(2)}\right)^{\frac{n+m}{2}}}, \qquad (9)$$

де $M_{00}^{(1)}, M_{00}^{(2)}$ — декартові моменти нульового порядку від функцій яскравості відповідних зображень:

$$M_{00}^{(1)}=\frac{I_0(\mathbf{0})}{\left(\lambda^2 q_1 d_1\right)^2}\mathcal{P}_1(\mathbf{0})=\frac{D_1^2 I_0(\mathbf{0})}{16\pi\lambda^2 q_1^2},$$

$$M_{00}^{(2)}=\frac{I_0(\mathbf{0})}{\left(\lambda^2 q_2 d_2\right)^2}\mathcal{P}_1(\mathbf{0})=\frac{D_2^2 I_0(\mathbf{0})}{16\pi\lambda^2 q_2^2}. \qquad (10)$$

Зазначимо, що вектори $\mathbf{Z}^{(1)}$, $\mathbf{Z}^{(2)}$ залежать від позиційних умов фіксації відповідних зображень і геометричних параметрів іконічної системи: в розглядуваному випадку — $\mathbf{q}, D$. Далі розглядаємо випадок, коли $D_1=D_2=D$.

Приймаючи $\mathbf{Z}^{(1)}$ за подання еталонного зображення (далі $\mathbf{Z}$), для якого зазначені параметри формоутворення вважаються відомими, покладемо співвідношення (9) в основу узагальнення інформаційно-геометричної моделі на випадок ізопланатичних зображень. Далі рискою позначаємо величини, що пов'язані з аналізованим зображенням.

В загальному випадку довільної вимірності простору моделі зв'язок між векторами $\mathbf{Z}^{(1)}$ та $\mathbf{Z}^{(2)}$ є досить громіздким і тут не наводиться. В окремому випадку тривимірного простору моделі $(N=2)$, маємо:

$$\left.\begin{aligned}
\bar{Z}_1&=\left(\frac{\bar{q}(q-f)}{q(\bar{q}-f)}\right)^2 Z_1+\frac{16\pi\lambda^2\bar{q}^2}{D^2 I_0(\mathbf{0})}\times\\
&\times\left[\bar{\mathbf{s}}_{02}-\left(\frac{q-f}{\bar{q}-f}\right)^2\mathbf{s}_{02}\right],\\
\bar{Z}_2&=\left(\frac{\bar{q}(q-f)}{q(\bar{q}-f)}\right)^2 Z_2,\\
\bar{Z}_3&=\left(\frac{\bar{q}(q-f)}{q(\bar{q}-f)}\right)^2 Z_3+\frac{16\pi\lambda^2\bar{q}^2}{D^2 I_0(\mathbf{0})}\times\\
&\times\left[\bar{\mathbf{s}}_{20}-\left(\frac{q-f}{\bar{q}-f}\right)^2\mathbf{s}_{20}\right],
\end{aligned}\right\}. \qquad (11)$$

Виходячи з вигляду функції розмиття точки (3) можна показати, що її семіінваріанти симетричні за першим та другим індексами, $\mathbf{s}_{02}=\mathbf{s}_{20}$. Здійснимо лінійну заміну координат у просторі моделі $Y_1=\frac{1}{2}(Z_1+Z_3)$; $Y_2=Z_2$; $Y_3=\frac{1}{2}(Z_1-Z_3)$. Врахуємо також поворот на кут γ картинної системи координат, в якій задане

зображення, що аналізується, стосовно координатної системи еталонного зображення. Виходячи з означення семіінваріантів [7] та виразу для перетворень координат картинної площини при повороті на кут γ:

$$\bar{x} = x\cos\gamma - y\sin\gamma, \quad \bar{y} = x\sin\gamma + y\cos\gamma,$$

можна показати, що зв'язок між інформаційними характеристиками зображення, підданого такому перетворенню, має вигляд:

$$\left.\begin{aligned} \overline{y_1} &= y_1; \\ \overline{y_2} &= y_2\cos(2\gamma) + y_3\sin(2\gamma); \\ \overline{y_3} &= -y_2\sin(2\gamma) + y_3\cos(2\gamma) \end{aligned}\right\}. \quad (12)$$

На підставі виразів (11), (12), одержимо:

$$\left.\begin{aligned} \overline{Y}_1 &= kY_1 + V(\bar{q}, q), \\ \overline{Y}_2 &= kY_2\cos(2\gamma) + kY_3\sin(2\gamma), \\ \overline{Y}_3 &= -kY_2\sin(2\gamma) + kY_3\cos(2\gamma), \end{aligned}\right\}, \quad (13)$$

де $k = \left(\dfrac{\bar{q}(q-f)}{q(\bar{q}-f)}\right)^2$,

$$V(\bar{q}, q) = \frac{32\pi\lambda^2\bar{q}^{-2}}{D^2 I_0(\mathbf{0})}\left[\bar{\mathbf{s}}_{02} - \left(\frac{q-f}{\bar{q}-f}\right)^2 \mathbf{s}_{02}\right].$$

Застосовуючи методику побудови афінно-інваріантних багатовидів другого степеня, розвинену в роботі [5], одержимо рівняння квадрики, інваріантної до перетворень (13) компонент вектора подання ізопланатичних зображень

$$F_2^2 = \frac{1}{k^2}\left\{\left[Y_1 - V(\bar{q}, q)\right]^2 - Y_2^2 - Y_3^2\right\}. \quad (14)$$

Порівнюючи вираз (14) з рівнянням афінно-інваріантної квадрики для стигматичних зображень (точкової відповідності об'єкта та зображення) [5, 8], легко бачити, що урахування ізопланатичності веде до зміщення центра квадрики вздовж осі OY_1 на величину $V(\bar{q}, q)$ та перетворення інваріантної інформаційної ознаки F_2^2.

Зазначимо, що, як випливає з (13), (14), при збільшенні відстані, з якої одержано зображення, квадрика (14) стягується в точку $\left[V(\bar{q}, q); 0; 0\right]$.

3. ВИСНОВКИ

Встановлені закономірності геометричної моделі подання ізопланатичних зображень дозволяють здійснювати їх морфологічний (позиційно-інваріантний) аналіз на основі належності точки, яка подає розподіл яскравості зображення, інваріантному багатовиду другого степеня (14), що визначається еталонним зображенням, та розташування цього багатовиду у просторі інформаційних характеристик. Взаємне розташування точок подання еталонного зображення та зображення, що аналізується, на інваріантному багатовиді дає підстави для обчислення відмінностей цих зображень у випадку їх належності одному класу еквівалентності геометричних форм.

ЛІТЕРАТУРА

1. Борн М., Вольф Э. "Основы оптики": Пер. с англ. — М.: Наука, 1973. — 720 с.
2. Михайленко В.Є., Корчинський В.М. "Багатовимірні геометричні моделі в інваріантному розпізнаванні проекційних зображень" // Proc. of the Third All-Ukrainian International Conf. "Signal/Image Processing and Pattern Recognition"(UkrOBRAZ'96). — Kiev (Ukraine). — 1996. —P.94-96.
3. Korchins'kyj V. "Informative features of multiplied plane images for their invariant recognition" // Proc. of the Third All-Ukrainian International Conf. "Signal/Image Processing and Pattern Recognition"(UkrOBRAZ'96). — Kiev (Ukraine). — 1996. — P.87-88.
4. Исимару А. "Распространение и рассеяние волн в случайно-неоднородной среде". Том 2: Пер с англ. — М.: Мир, 1981. — 318 с.
5. Михайленко В. Є., Корчинський В. М. "Концепція геометричного об'єкту в морфологічному аналізі проеційних зображень" // Прикладна геометрія та інженерна графіка: К.: КДТУБА, 1997. — Вип.61. — С. 59 – 63.
6. Васильев В. Н., Гуров И. П. "Компьютерная обработка сигналов в приложении к интерферометрическим сигналам." — СПб.: БХВ — Санкт-Петербург, 1998. — 240 с.
7. Гнатушенко В. В. "Багатовимірна геометрична модель напівтонових зображень у вигляді центральних проекцій". // Збірка праць міжнар. наук.-практ. конф. "Сучасні проблеми геометричного моделювання". — Харків, 2001. — С. 202-204.
8. Гнатушенко В. В. "Графічне подання перетворень проекційних зображень у багатовимірній інформаційно-геометричній моделі." // Прикладна геометрія та інженерна графіка. К.: КНУБА — 2001. — Вып. 69. — С. 217 – 219.

ВИДІЛЕННЯ ІНФОРМАТИВНИХ ОЗНАК ТЕКСТУРНИХ ЗОБРАЖЕНЬ НА ОСНОВІ МЕТОДА РЕКУРЕНТНОГО КІЛЬКІСНОГО АНАЛІЗУ

О.М.Ахметшин, М.Г.Комарова

Дніпропетровський національний університет
49050 м.Дніпропетровськ, пер. Науковий 13
Email: akhm@mail.dsu.dp.ua

АНОТАЦІЯ

У роботі розглянуто новий метод кількісного аналізу текстурних зображень, який грунтується на ідеї аналізу нелінійних динамічних процесів у багатовимірному фазовому просторі, що дозволяє підвисити чутливіст аналізу. Подані результати досліджень на прикладах текстурних зображень та даних рентгентовської комп'ютеної томографії для виявлення ранніх стадій розвитку остеопорозу.

1.ВСТУП

Проблема аналізу текстурних зображень традиційно викликає велике зацікавлення та знаходиться на межі методів комп'ютерного бачення та штучного інтелекту. Особливо актуальною ця проблема виявляється при застосуванні до задач ранньої медичної диагностики та контролювання динаміки процесу лікування онкологічних захворювань. У роботі [1] було розглянуто підхід, заснований на ідеї відображення зрізів текстурних зображень у двовимірному просторі ядер Вольтерра другого порядку, а в роботі [2] для відображення одновимірніх зрізів використана двовимірна площина кумулянтів третього порядку. Обидва розглянутих підхода є достатньо чутливими, але за своєю суттю подають аппарат якісного аналізу, оскільки базуються на тому, що синтезовані зображення візуально відрізняються від еталонних.

Метою поданої роботи є подальший розвиток цього спрямування, але з напиранням на підвищення загальної чутливості аналізу на основі кількісних характеристик, що обчислюються за допомогою метода рекурентного кількісного аналізу (РКА) .

2.МЕТОД

РКА, запропонований Дж.П.Збілутом та Ч.Л.Веббером [3, 4], являє собою технологію для аналізу нелінійних динамічних сигналів. Головною перевагою метода є можливість його застосування до часових та просторових рядів незалежно від від їх статистичної природи, наявності шуму від будь-якого джерела (самої системи або спостереження) та довжини ряду. Метод РКА дає локальний огляд поведінки сигналу, надає можливість досліджувати короткочасову поведінку системи.

При дослідженні нелінійної динаміки процесів вважається, що змінні, які описують систему, змінюються за часом залежно одна від одної. Якщо систему описують за допомогою d змінних x_1, x_2… x_d, то така взаємозалежність може бути подана у формі вектора (x_1…x_d). Але на практиці для багатьох систем можливо замірювати лише поведінку за однією змінною. Для подальшого дослідження динаміка такої системи повинна бути змодельована у відбудованому фазовому просторі. Процедура побудови такої моделі спирається на концепцію, що скалярний сигнал, що спостережується, є одновимірною проекцією динамічної поведінки системи, що подається *d* вимірами. Динаміка може бути реконструйована як багатовимірна за допомогою даних лише одного скалярного часового сигналу, який було спостережено. Для застосування процедури відбудови записані дані повинні являти собою квантований сигнал *s(n)*. Вектор *Y(n)* формується з затриманих версій сигналу *s(n)* [5]:

$$Y(i) = (s(i), s(i+l)...s(i+(d-1)L)), \qquad (1)$$

де *d* – обрана кількість вимірів, *L* - затримка. Теорема Такенса [6] встановлює математичне співвідношення між розмірністю побудованого вектора та реальною розмірністю атрактора відповідної динамічної системи (D): $d=2D+1$, де *d* є мінімальною розмірністю, що забезпечує можливість реконструктування атрактора. Під атрактором розуміємо множину точок в фазовому просторі, на яку виходить система при нескінченному збільшенні часу спостерігання. Наведене вище співвідношення є справедливим для детермінованих систем при відсутності шуму. При наявності шуму від будь-якого джерела (самої системи чи вимірювання) необхідний більш високий порядок.

Технологія рекурентного кількісного аналізу не передбачає побудування атрактора у реконструйованому фазовому просторі, але на основі даних, достатніх для цього, дозволяє визначити деякі

показники хаотичності системи (ентропію процесу та значення максимальної експоненти Ляпунова).

Рекурентний кількісний аналіз грунтується на обчисленні матриці Євклідових відстанів D_{ij} між всіма можливими комбінаціями пар рядків відбудованої матриці *Y(n)*.

$$D(i,j)=\sqrt{\sum_{k=1}^{n}(i_k-j_k)^2}\,, \qquad (2)$$

де i, j – вектори, що відповідають рядкам матриці *Y(n)*, n – розмірність цих векторів. Отримана матриця використовує функцію Хевісайда для значень відстаней між скалярними рядами, що є меншими за деякий напередвизначений поріг. За значеннями, що повертає така функція, будується графік, що носить назву рекурентного, а затемнені точки на ньому, які відповідають відстаням, меншим за значення порогу, - рекурентних точок. Графічне подання являє собою квадратну матрицю з затемненою головною диагоналлю ($D_{ij}=D_{ji}$) та може бути розглянуте як глобальна автокореляційна структура системи. Найбільш важливими для розглядання є рекурентні точки, що утворюють диагональні структури, які є паралельними відносно головної диагоналі. Наявність ділянок с такими структурами на рекурентному графіку вказує на наявність детермінізму у динамічному процесі, що розглядається.

На основі побудованого рекурентного графіка можуть бути обчислені дескриптори динамічного процесу.

Оскільки рекурентний графік є симетричним відносно головної диагоналі, до розглядання залучається тільки верхня трикутна область за виключенням нульових евклідових відстанів, що утворюють головну диагональ.

1. Рекурентність (R) являє собою відношення кількості затемнених (рекурентних) точок до загальної кількості точок у верхній трикутній області графіка. Рекурентні точки відповідають "рівнинам" фазового простору, в таких точках значення величини, що спостерігається, визначається початковими умовами (детерміновано процедурою введення), відповідно, величина, що дорівнює 100-R, визначає пропорцію точок, значення в яких зумовлено стохастичною поведінкою системи.

$$R=\frac{N_r}{N}, \qquad (3)$$

де N_r - кількість рекурентних точок, N - загальна кількість точок.

2. Детермінизм (D) визначає відсоток рекурентних точок, які утворюють на графіку лінії, паралельні головній диагоналі. Наявність диагональних ліній є сигнатурою детермінованої структури системи. Значення цього параметру визначається у значній мірі напередвизначеним параметром LINES, що вказує, яку кількість підряд розташованих точок слід вважати лінією.

$$D=\frac{N_d}{N_r}, \qquad (4)$$

де N_d - кількість рекурентних точок, що фомують диагональні структури, N_r - кількість рекурентних точок.

3. Ентропія (E) розподілення довжини диагоналей, паралельних головній, є мірою невизначеності (хаотичності) ймовірносної системи. По мірі наближення системи до стану рівноваги значення ентропії зростає. Визначається за формулою Шеннона для інформаційної ентропії.

$$E=\sum_{i=1}^{n}P_i\log_2\frac{1}{P_i} \qquad (5)$$

4. Максимальна довжина диагональної лінії (M) зворотньо пропорційна максимальній експоненті Ляпунова (L) для процесу, що розглядається,

$$L=\frac{1}{M} \qquad (6)$$

являє собою міру швидкості розходження сусідніх траєкторій у фазовому просторі, тобто швидкість "забування" початкових умов системи (ступень "самостійності" поведінки системи). Аналіз цього показника дозволяє оцінити можливість передбачення подальших значень ряду.

5. Тренд локальної рекурсії (T) є мірою стаціонарності процесу, визначає міру зміни густини рекурентних точок в напрямку від головної диагоналі графіка. Стаціонарним сигналам відповідає близьке до нуля значення цього показника.

3.Застосування запропонованого методу

Метод РКА був застосований для оцінки стохастичних процесів, поданих текстурними зображеннями та томографічними знимками кісток людини, хворої на остеопороз.

Текстурні зображення, обрані для порівняння, подані на Рис.1.

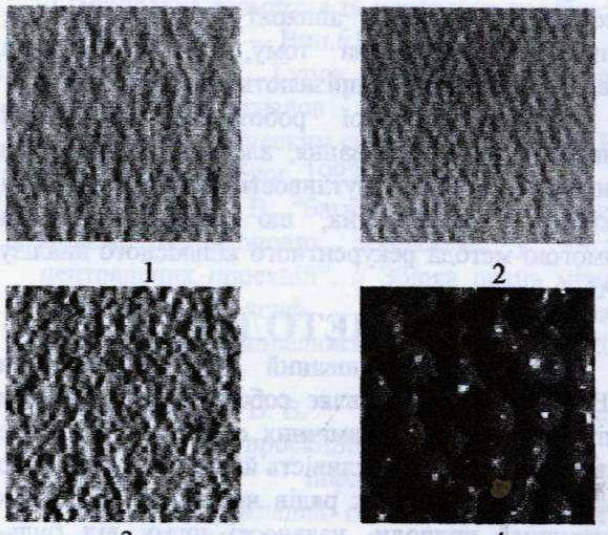


Рис.1.Текстурні зображення

Оскільки метод РКА призначений для оцінювання рядів даних, для визначення параметрів текстурних зображень потрібно працювати з їх зрізами. Зріз зберігає інформацію про спектральний склад текстури, але втрачаються дані про об'єкти, що присутні на зображенні.

Поданим текстурним зображенням відповідають рекурентні графіки, приведені на Рис.2.

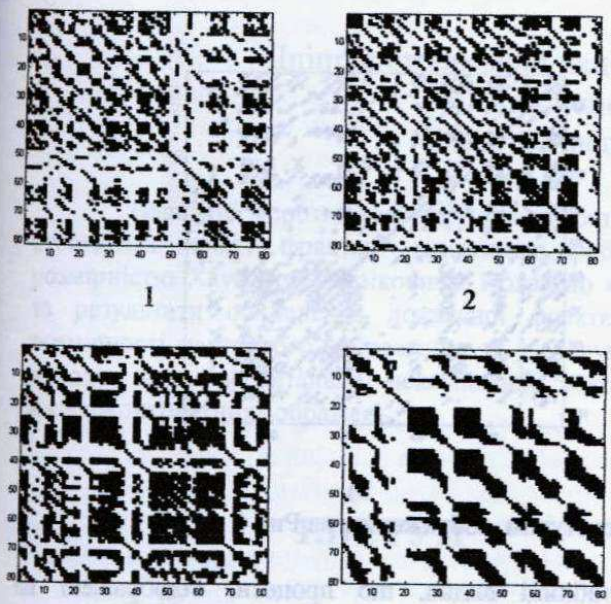


Рис.2.рекурентні графіки, побудовані для зрізів текстур, зображених на Рис.1.

Графіки, що мають стуктуру майже однорідної густини, відповідають рядам з невеликою дисперсією значень. Пробіли у рекурентному графіку свідчать про наявність "скачків" даних у розглянутій послідовності. для значення відбудованої розмірності більше одиниці така залежність переноситься на поведінку траєкторій у фазовому просторі. Для побудування рекурентних графіків (Рис.3) була обрана розмірність d=2, оскільки при виборі одиничної розмірності кількісні показники зображень 1 та 2 відрізняються на незначні величини. При обиранні занадто малої розмірності відбудованого фазового простору деяка інформація про структуру даних може бути загублена. Значення порогу визначення рекурентних точок дорівнює 20 відсотків від максимальної з отриманих євклідових відстанів. У Табл.1 наведені п'ять основних дескрипторів РКА та величина, зворотна довжині максимальної диагональної лінії на графіку, яка відповідає максимальній експоненті Ляпунова для процесу, що розглядається.

На основі кількісних характеристик текстурних зображень можна зробити висновки про характер процесів, які вони відображають. По-перше, як за візуальною оцінкою, так і за кількісним аналізом рекурентних графіків схожі початкові зображення виявляються досить різними. Найбільше значення параметру R відповідає текстурі 3, найменше – текстурі 1, тобто вона виявляє найбільш стохастичну поведінку. За коефіцієнтом детермінізму D більш детермінований процес серед усіх розглянутих являє текстура 4, текстура 3 має найменшу швидкість "забування" початкових умов. Ентропія текстури 4 найбільша, що свідчить про те, що процес, який вона відображає, знаходиться ближче до стану рівноваги, ніж решта розглянутих. З точки зору стаціонарності найближче до нуля значення тренду локальної рекурсії відповідає текстурі 1.

Таблиця 1

Кількісні характеристики аналізованих текстур

	R	D	M	L	E	T
текстура 1	27.500	21.473	16	0.0625	1.890	9.066
текстура 2	29.475	26.492	12	0.0833	2.392	78.586
текстура 3	42.469	42.078	19	0.0526	2.527	141.510
текстура 4	29.557	64.240	16	0.0625	2.970	34.566

Метод РКА був застосований також для аналізу томографічних знимків двох кісток людини (Рис.3), які знаходяться на різних ступенях захворювання на остеопороз. Білі лінії на зображеннях вказують розташування на знимках зрізів, які були взяті для аналізу.

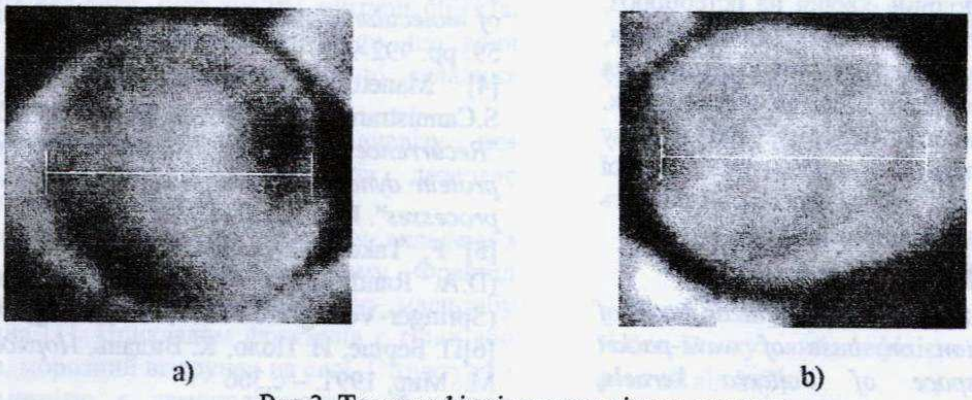
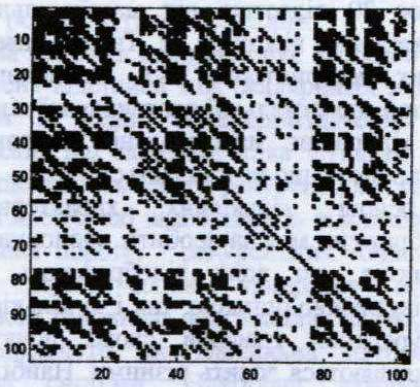
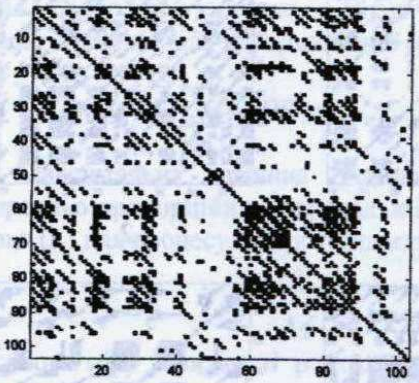


Рис.3. Томографічні знимки кісток людини
a) – нормальна кістка, b) – кістка, хвора на остеопороз.

Предметом розглядання є знаходження та зіставлення кількісних характеристик цих знимків з метою виділення характерних ознак, які б відрізняли їх один від одного. Як і при аналізі текстурних зображень, кількісні характеристики обчислені для зрізів. Перетворення двовимірного зображення на одновимірний сигнал хоча і відповідає умовам застосування методу, але при такому поданні повністю загублюється достовірність інформації – обробці підлягає вже деякий періодичний ряд. Оскільки кістка має пористу структуру, деякі ознаки повинні зберігатися на обох зображеннях, і відхилення значень кількісних показників буде незначним. Тому для дослідження були взяті градієнтні перетворення зрізів зображень (тільки високочастотні компоненти).



a)



b)

Рис.4. Рекурентні графіки, що відповідають текстурам, зображеним на Рис.4

Для побудування рекурентних графіків, поданих на Рис.4, були обрані такі вхідні параметри: відбудована розмірність дорівнює трьом, поріг рекурентних відстаней – 30 відсотків від максимальної євклідової відстані для кожного зображення, лінією вважаються розташовані поряд 10 рекурентних точок. Кількісні характеристики, обчислені для графіків (Рис.4), наведені у Таблиці2.

З таблиці видно, що процеси, відображені на знимках, відрізняються за всіма характеристиками. Кістка на зображенні b, яка знаходиться на більш пізньому ступені захворювання, являє собою процес, який є менш стаціонарним та детермінованим, ніж процес, що відображений на знимку b, але водночас більш залежить від "початкових умов".

Таблиця 2

Кількісні характеристики аналізованих зображень

	R	D	M	L	E	T
Зображення a	55.797	38.315	23	0.0435	3.264	108.598
Зображення b	34.875	12.828	17	0.0588	2.247	123.867

4.ВИСНОВКИ

В ході роботи проілюстровано застосування методу РКА для підвищення чутливості та отримання кількісних характеристик зображень. Метод був успішно застосований для аналізу текстурних зображень. Результати аналізу томографічних знимків двох кісток людини, хворої на остеопороз, які знаходяться на різних ступенях захворювання, підтвердили можливість застосування методу для медичної диагностики навіть на ранніх стадіях. Зручність підходу полягає в тому, що аналізу підлягають дані, не описані за допомогою будь-якої математичної моделі. Це робить область застосування методу дуже широкою.

ЛІТЕРАТУРА

[1]. Akhmetshin A.M., Shutin D.V., *Nonlinear model of qualitayive classification on base of wave-packet decomposition in space of Volterra kernels*, UkrOBRAZ'2000, Kyjiv, pp. 153-156, 2000.

[2]. Akhmetshin A.M., Mikhalev D.A., *Classification of texture images with the usage of artificial neural nets and higher-oder stsistics*, UkrOBRAZ'2000, Kyjiv, pp. 173-176, 2000.

[3] Manetti, C., M.-A. Ceruso, A. Giuliani, C.L. Webber, Jr., and J.P. Zbilut (1999). "*Recurrence quantification analysis as a tool for the characterization of molecular dynamics sumilations*", Physical Rev. E. 59: pp. 992-998.

[4] Manetti, C., A. Giuliani, M.-A, Ceruso, S.Cannistraro, C. L. Webber, Jr., and J.P. Zbilut (2001). "*Recurrence analysis of hydration effects on nonlinear protein dynamics: multiplicative scaling and additive processes*". Phys. Lett. A 281: 317-323

[5] F. Takens, *Dynamical Systems and Turbulence* (D.A. Rand and L.-S. Young, eds.) pp. 366-381. (Springer-Verlag, NewYork, Heidelberg, Berlin, 1980).

[6]П. Берше, И. Поло, К. Видаль, *Порядок в хаосе*, - М.: Мир, 1991. – с.366

СЕГМЕНТАЦІЯ НИЗЬКОКОНТРАСТНИХ ЗОБРАЖЕНЬ НА ОСНОВІ МЕТОДУ МУЛЬТИФРАКТАЛЬНОЇ ДЕКОМПОЗИЦІЇ

Ахметшин О.М. Киргизов І.О.

Дніпропетровський національний університет, Україна, 49050,
м. Дніпропетровськ, пр. Науковий 13.
Електронна пошта: akhm@mail.dsu.dp.ua

Наведені короткі теоритичні відомості, що стосуються поняття фрактал – множини з дробовою розмірністю Хаусдорфа-Безіковича. Показано метод та результати обчислення локальної фрактальної розмірності растрового зображення та застосування методу мультифрактальної декомпозиції текстур і низькоконтрастних зображень.

1.ВСТУП

Однією з головних задач розпізнавання образів є виділення об'єктів. Особливо гостро і актуально ця проблема існує при сегментації низькоконтрастних зображень.

Кожний об'єкт на зображенні, який потрібно виділити, має властиві йому і тільки йому деякі характеристики. Точки зображення, що мають спільні характеристики відносяться до одного об'єкту, а точки з іншими характеристиками – до інших об'єктів.

Одним із простих методів сегментації зображення є його квантування. Тоді точки з однаковими рівнями інтенсивності об'єднують в один клас або сегмент даних. Зрозуміло, що цей метод занадто тривіальний, оскільки, багато різних об'єктів можуть мати одну і ту ж саму інтенсивність і навпаки, один об'єкт може мати різні рівні інтенсивності. Але об'єкти можуть відрізнятися не тільки рівнями інтенсивності, а й своєю формою.

Деякі методи використовують знаходження краю (грані) для того, щоб знайти контури об'єкта. Одним із недоліків таких методів є те, що вони дають небажані грані при дослідженні складних об'єктів або текстур.

Застосування фрактального аналізу для сегментації та класифікації об'єктів дозволяє уникнути вищезгаданих проблем.

Фракталом називають структуру, складену з частин, які в деякому сенсі подібні цілому . Фрактал має однаковий вигляд, незалежно від масштабу спостерігання [1]. Прикладом фракталів є гілки або кора дерева, морозний візерунок на склі. Структури з такою геометрією є самоподібними і їх можна класифікувати таким критерієм, як фрактальна розмірність D.

Відомо, що просторово ізотропні поверхні дають зображення, інтенсивність яких є фрактальною функцією тієї ж дробної розмірності, що і початкова поверхня. Це дозволяє, вимірюючи характеристики зображення, проводити аналіз самої поверхні.

2.ФРАКТАЛЬНА РОЗМІРНІСТЬ

Фрактали можна розглядати як множину точок, вкладених у простір. Множина точок, що утворюють площину в евклідовому просторі має топологічну розмірність D_T=2. Евклідова розмірність простору E=3. Для визначення розмірності Хаусдорфа-Безіковича (фрактальної розмірності D) потрібно визначити поняття відстань між точками в просторі. Потрібно визначити величину множини точок $\Im$ у просторі. Один із засобів виміряти площу поверхні полягає в тому, щоб розділити простір на невеликі куби із стороною δ . Якщо помістити центр куба в деякій точці множини, то всі точки в кубі будуть покриті цим кубом. Підраховуючи число кубів, необхідних для покриття множини точок, ми отримуємо міру величини множини. Площу кривої можна визначити як число кубів $N(\delta)$. Продовжуючи підрахунок і, знайшовши $N(\delta)$ при менших значеннях δ . Для фрактальних поверхонь буде виконуватись співвідношення:

$$N(\delta) \equiv \frac{1}{\delta^D} \qquad (1)$$

Звідси, фрактальна розмірність D визначається:

$$D = \lim_{\delta \to 0} \frac{\ln(N(\delta))}{\ln(\frac{1}{\delta})} \qquad (2)$$

Тобто, D – кутовий коефіцієнт графіка $\ln N(\delta)$ як функції від $\ln(\delta)$.

Розмірність D, яка визначається по формулі (2)

через підрахунок кількості кубів, необхідних для покриття множини $\Im$ в залежності від розміру куба, називають клітковою розмірністю.

Кожен об'єкт (підмножина $\Im_i$) зображення $\Im$ характеризується своєю фрактальною розмірністю D_i і множина точок $\Im$ є мультифрактал та являє собою об'єднання підмножин $\Im_i$ [1]:

$$\Im = \bigcup_i \Im_i \qquad (3)$$

Суть розробленого методу мультифрактальної декомпозиції полягає в обчисленні D_i та в об'єднанні (групуванні) точок в підмножини $\Im_i$.

2. РЕАЛІЗАЦІЯ МЕТОДУ

2.1. ОПИС ПОЧАТКОВИХ ДАНИХ

Спочатку дамо опис даних, над якими буде проводитись дослідження.

1) Зображення – це двовимірна матриця значень інтенсивності;
2) Зображення представляється в градаціях сірого: інтенсивність має значення цілого числа в інтервалі [0, 255], де 0 – чорний, а 255 – білий.
3) Координати пікселя (ряд, стовбець). Початок координат точка (0, 0) – верхній лівий кут;
4) Точка матриці I з координатами (i, j) – I[i, j].

2.2. АЛГОРИТМ

Структура методу полягає в наступному. В зображенні для кожного пікселя (i, j) обчислюється локальна фрактальна розмірність $D_{i,j}$ для деякого невеликого вікна P. Псевдокод алгоритму методу має вид:

1. I – це матриця розміром [m, n] іненсивностей.
2. ID- матриця фрактальних розмірностей для кожного пікселя матриці I.
3. Для кожного пікселя I[i,j]:
   - 3.1. P – невелике вікно розміром [k, k], що проходить по зображенню I[i, j].
   - 3.2. Обчислюється фрактальна розмірність D для вікна P.
   - 3.3. Назначаємо D → ID[i, j].
4. Отримуємо розподілення розмірностей ID і кожній групі назначаємо свій колір. Проводимо сегментацію зображення I[i, j], оскільки, кожній точці зображення I[i, j] відповідає точка ID[i, j].

2.3. АЛГОРИТМ ОБЧИСЛЕННЯ ФРАКТАЛЬНОЇ РОЗМІРНОСТІ

Особливу увагу потрібно звернути на крок 3.2. Оскільки, фрактальна розмірність є тільки апроксимацією на наших дискретних даних.

Розмір квадратного вікна P вибираємо [k, k], де k-число степені 2 або $k=2^v$, для деяких цілих чисел v. Слід звернути увагу на те, що розмір вікна є парним, отже немає центрального пікселя якому можна поставити у відповідність обчислене значення фрактальної розмірності. Оскільки, вікно пробігає всі піксели по черзі, то обчислене значення розмірності присвоюється лівому верхньому пікселю вікна P.

Ідея обчислення фрактальної розмірності для вікна P базується на покритті множини точок зображення цим вікном та визначенні діапазону інтенсивності, що займають піксели у вікні. Далі, розбиваючи вікно на дрібніші підвікна, проводимо ті ж обчислення. Потім будуємо залежність $\ln(N(\delta))$-кількість зайнятого діапазону від $\ln(1/\delta)$ -масштабу та, використовуючи метод найменшого квадрата (МНК), визначаємо кут нахилу даної прямої, яка являє собою фрактальну розмірність вікна.

Вибір розміру вікна рівний числу степені 2 пояснюється зручністю розбиття вікна P на менші підвікна однакового розміру. Наприклад, вікно розміром P[16,16] можна розбити на 4 підвікна розміром [8, 8], на 16 підвікон розміром [4,4] і т.д.

Алгоритм:

1. Для різних значень дробового числа $h \in (0, 1]$
   - 1.1. Ділимо вікно P на $(1/h)^2$ квадратів.
   - 1.2. Розділяємо рівні інтенсивності зображення [0, 255] на 1/h рівнів. Нумеруємо ці рівні від 1 до 1/h.
   - 1.3. Для кожного підвікна $(a,b) \in P$ виконуємо:
     - 1.3.1. Позначаємо u та w мінімальний та максимальний рівні інтенсивності (1..1/h) відповідно.в даному підвікні.
     - 1.3.2. Записуємо $w-u+1 \rightarrow n_{a,b}(h)$.
2. Будуємо залежність $\ln[N(h)]$ та $\ln(h)$ і використовуючи (МНК) знаходимо фрактальну розмірність пікселя для покриваючого множину вікна P.

Вікно P[kxk] розбивається на підвікна наступним чином [(k*h)x(k*h)]. Для розбиття вікна P h приймає значення $h=1,½,¼,\ldots(1/h)^v$. Де параметр v потрібен для визначення розміру вікна в пікселях 2^v.

Для отримання фрактальної розмірності виконуємо підбір параметрів лінійної функції y=ax+b методом найменших квадратів [5], де x=ln[1/h], y=ln[N(h)], а параметр a=-D. Отже,

$$a = \frac{K_{xy}}{D_x}; \qquad (4)$$

де K_{xy} – кореляційний момент, D_x –дисперсія

$$K_{xy} = \frac{\sum_{i=1}^{n}(x_i - m_x)(y_i - m_y)}{n}; \qquad (5)$$

$$D_x = \frac{\sum_{i=1}^{n}(x_i - m_x)^2}{n}. \qquad (6)$$

m_x та m_y –математичне очікування

$$m_x = \frac{\sum_{i=1}^{n} x_i}{n} \qquad (7)$$

$$m_y = \frac{\sum_{i=1}^{n} y_i}{n}. \qquad (8)$$

2.4. РЕЗУЛЬТАТИ

Описані вище алгоритми дозволяють отримати оцінку фрактальної розмірності зображення з різними її значеннями та на цій основі проводити сегментацію і класифікацію зображень.

Але не визначеним питанням є те, яким слід вибирати розмір вікна для покриття множини, по якому буде знаходитись фрактальна розмірність. Оскільки до розміру ставляться зовсім протилежні вимоги:

- для більш точного визначення границь між різними текстурами необхідно зменшувати розмір вікна;
- для підвищення точності обчислення фрактальної розмірності необхідно його збільшувати.

Вибір розміру вікна робиться емпірично при задовільних результатах виділення контурів та сегментації текстур.

Для сегментації були вибрані чотири різних за своєю структурою текстури (рис.1).

На рис.2. представлена карта фрактальних розмірностей – фрактальні розмірності кожної точки текстури.

Заключною стадією є сегментація. З рис.3. видно, що після цієї операції кожен сегмент має декілька точок іншого сегмента. Цей результат можна пояснити тим, що в текстурах зустрічаються подібні по своїй структурі частини зображення.

Однією з переваг такого метода є те, що сегментація текстур не залежить від інтенсивності (кольору) текстури, а залежить тільки від її форми.

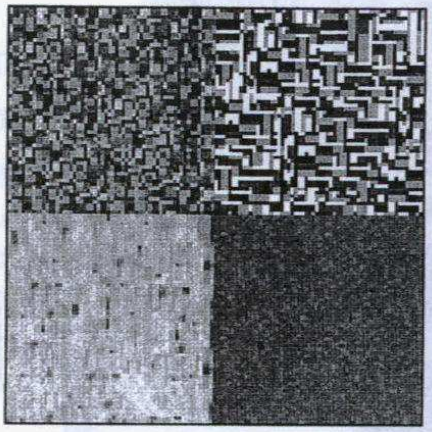


Рис. 1. Об'єднання текстур. Розмір 120 x 120.

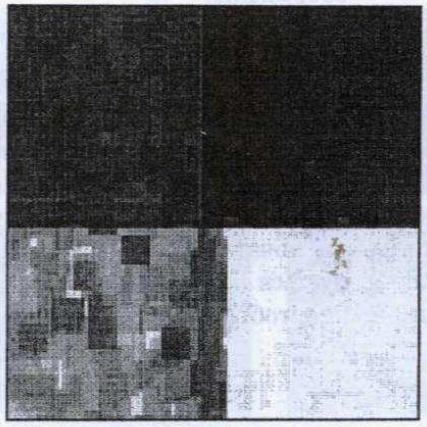


Рис. 2. Карта фрактальних розмірностей текстури

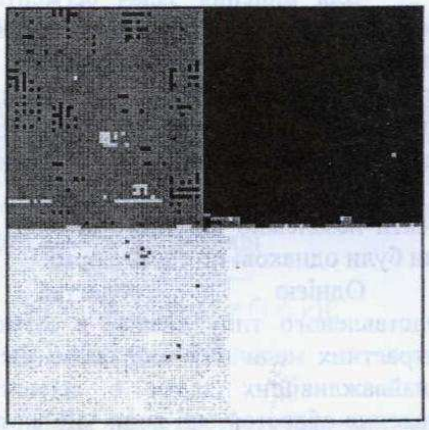


Рис. 3. Результат сегментації текстур.

Рис. 4. Низькоконтрастне зображення

Рис. 5. Карта фрактальних розмірностей

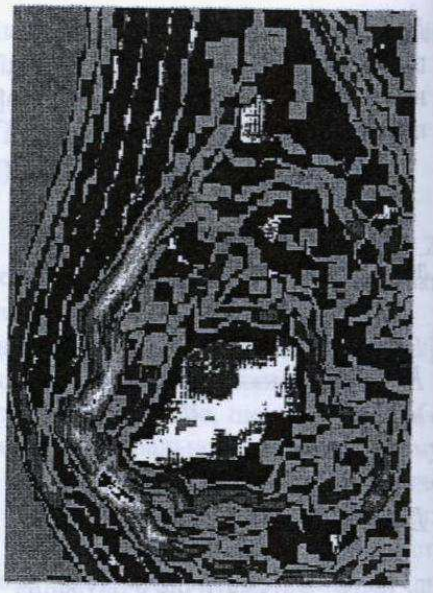


Рис. 6.Результат сегментації

По низькоконтрастному зображенню (рис. 4) було обчислено фрактальні розмірності (рис. 5) на основі яких зроблено сегментацію зображення (рис.6). На рис. 7 чітко виділені групи розмірностей. Бачимо, що початкове зображення є множина, яка характеризується набором фрактальних розмірностей, тобто є мультифракталом.

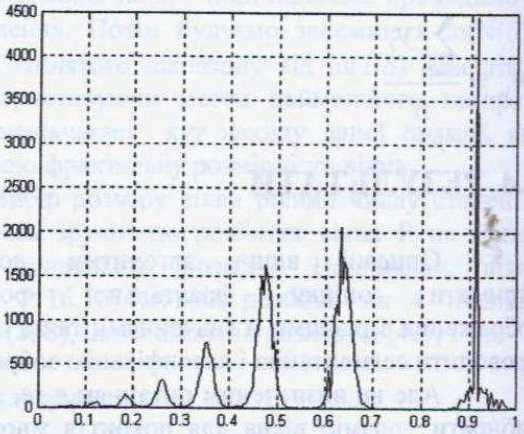


Рис. 7. Розподіл фрактальних розмірностей, розбитих в діапазоні [0..1]

ВИСНОВОК

В цій статті був розроблений метод мультифрактальної декомпозиції з допомогою якого була проведена сегментація текстур та низькоконтрастних зображень.

Для рішення задач сегментації зображень застосовується багато різних методів, але тільки з допомогою фрактального аналізу, застосовуючи метод мультифрактальної декомпозиції, при сегментації зображення та класифікації об'єктів, буде враховуватись їх структура (геометрія).

Даний метод відносить до одного класу об'єкти незалежно від його кольору, головне, щоб вони були однакові по своїй формі.

Однією з областей застосування представленого типу аналізу є сегментація низько контрастних медичних зображень. Оскільки одними з найважливіших задач є сегментація пухлин, виявлення областей, які вони займають та виділення країв цих областей.

ЛІТЕРАТУРА

1. Федер Е. "*Фракталы*": Пер. с англ.-М.; Мир, 1991.-254 с.
2. Горобець Ю.І., Кучко А.М. "*Вступ до фізики фрактальних структур*".-Київ; 2000.
3. Потапов А.А. Герман А.А. //"*Радиотехника и электроника*".-М.; 2000. Т.-45 N8, с 946-953.
4. Kenkel N.C. Walker D.J."*Fractal in the biological sciences*"
5. Вентцель Е.С. "Теория вероятностей".-М.;1969.-576 с.
6. Chen, C. C., J. S. Daponte, and M. D. Fox, "*Fractal Feature Analysis and Classification in Medical Imaging*" IEEE Transactions on Medical Imaging, **8**, pp. 133-142, 1989.
7. Коваленкова Н.В.//"*Оптико механическая промішленность*".-1991.-N4 с.42-46.

Класифікація текстурних зображень в інваріантному просторі моментів Цернакі аналітичних автокореляційних функцій.

Довженко Оксана Володимирівна, Олександр Мубаркович Ахметшин

Дніпропетровський національний університет, Радіофізичний факультет, вул. Наукова 13, корпус 12, м. Дніпропетровськ 49050. тел. 056-7769092. e-mail: dovzhenko@mail.dsu.dp.ua, akhm@mail.dsu.dp.ua.

В роботі розглянуто новий метод класифікації текстурних зображень, представлених аналітичними автокореляційними функціями (АКФ) в інваріантному просторі ознак моментів Церніке на основі апарату векторної підтримки. В рамках екпериментів було досліджено інваріантну стійкість ознак АКФ еталонних текстурних зображень до перетворень зсуву, масштабу і повороту.

1. Вступ

Виділення стійких інформаційних ознак текстурних зображень представляє собою одну з найбільш важливих і актуальних проблем на перетині областей штучного інтелекту і "комп'ютерного бачення", що обумовлено її великою практичною значимістю для задач ранньої медичної діагностики, дистанційного зондування і контролю якості матеріалів.

З необхідності автоматизації процедур класифікації текстур постає вимога переконливих теоретичних гарантій інваріантності наборів описових ознак до зсуву, масштабу і повороту, у зв'язку з чим у цій області було досліджено цілий ряд методів, до найбільш популярних з котрих можна віднести статистичний, спектрально-кореляційний підходи [1, 2], а також поширені останнім часом методи виділення інформаційних ознак багаторівневої вейвлет-декомпозиції [3], однак проблема залишається далекою від розв'язку.

Остання обставина стимулює до відшукання нових підходів, в тому числі основаних на стандартних методах спектрально-кореляційної класифікації.

Візуальний аналіз АКФ вихідного зображення або енергетичного Фур'є-спектра (рис. 1) показує, що головним недоліком даних підходів є недостатня чутливість до виділення характерних інформативних ознак кожної з текстур, крім того, ці характеристики не мають властивості інваріантності.

Успішне використання математичного апарату кореляційного аналізу, в першу чергу, пов'язане з відшуканням методу підвищення чутливості АКФ до виділення характерних ознак зображень текстур, подруге, вирішенням проблеми інваріантності описових характеристик, для чого в рамках даної роботи було використано метод моментів Церніке.

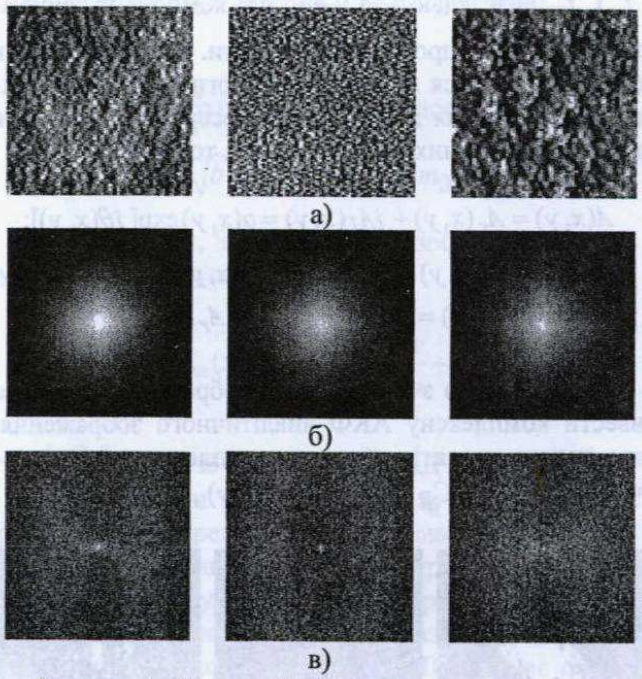


Рисунок 1. Кореляційні і спектральні характеристики, обчислені на основі стандартних підходів: а – вихідні зображення текстур; б – АКФ; в – енергетичні Фур'є спектри.

2. Аналітичні АКФ як апарат виділення інформаційних ознак текстур.

Задачу підвищення чутливості АКФ було вирішено на основі використання концепції "аналітичного зображення", що відома як метод амплітудно-частотної демодуляції [4], в рамках котрої зображення $S(x, y)$ представляється у вигляді

$$S(x,y) = a(x,y)\cos[2\pi f(x,y)];$$
$$f(x,y) = f_{cp} + \Delta\theta(x,y), \qquad (1)$$

де припускається, що функція $a(x,y)$ повільно змінюється у зрівнянні з $cos[2\pi f(x,y)]$. Функція $f(x,y)$ розглядається у вигляді аналогу "віртуальної" несучої частоти, а $\theta(x,y)$ – компонента, що відповідає фазовій модуляції. В цьому розумінні $a(x,y)$ представляє собою компоненту амплітудної модуляції, а $f(x,y)$ –

частотної. Оскільки fcp в (1) не являє інтересу, виділення інформативних характеристик $a(x,y)$ та $\theta(x,y)$ базується на формуванні нового комплексного зображення $z(x,y)$

$$z(x,y) = S(x,y) + jq(x,y) = A(x,y)\exp[-j(xf_x + yf_y)], \quad (2)$$

де $q(x,y)$ розглядається як зображення, сполучене за Гільбертом з вихідним $S(x,y)$, або

$$q(x,y) = \frac{1}{\pi^2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\frac{S(x,y)}{(u-x)(v-y)}dudv,$$

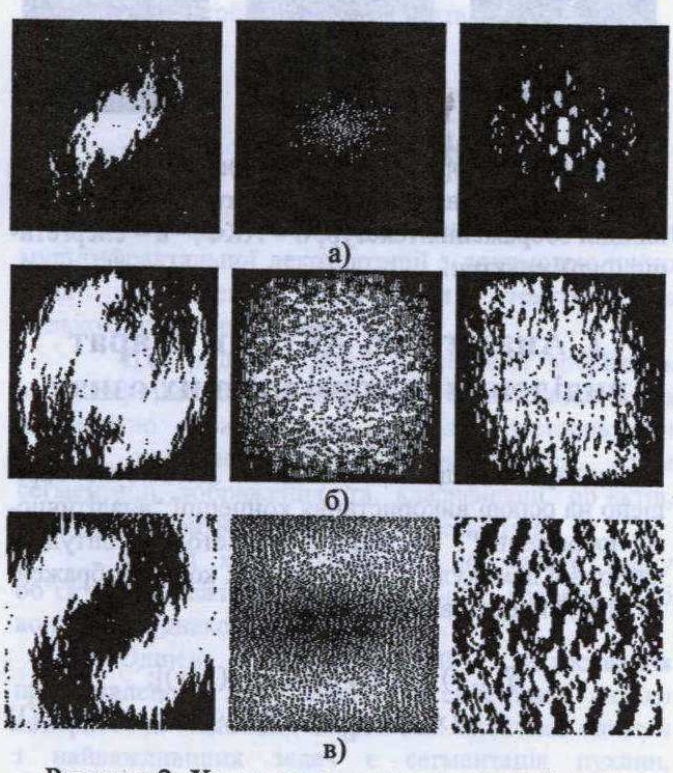
f_x і f_y розглядаються у вигляді компонент "віртуальної" несучої просторової частоти. Функція $A(x,y)$ в (2) розглядається як комплексна огинаюча аналітичного зображення $z(x,y)$, що безпосередньо залежить від інформативних характеристик, тобто

$$A(x,y) = A_r(x,y) + jA_I(x,y) = a(x,y)\exp[j\theta(x,y)];$$

$$a(x,y) = (A_r(x,y) + A_I(x,y))^{1/2};$$

$$\theta(x,y) = arctg[A_I(x,y)/A_r(x,y)].$$

За аналогією з аналітичним зображенням, можна ввести комплексну АКФ аналітичного зображення, що також характеризується комплексною огинаючою $C(x,y) = C_R(x,y) + jC_I(x,y)$.



а)

б)

в)

Рисунок 2. Характеристики комплексної огинаючої АКФ текстурних зображень: а – бінаризовані $C_R(x,y)$; б - амплітудно-просторові характеристики $C_R(x,y)$; в - фазо-просторові характеристики $C_R(x,y)$.

На рис. 2 представлені бінаризовані характеристики АКФ, також її амплітудно-просторові і фазо-просторові характеристики, що відповідають вихідним текстурним зображенням, показаним на рис. 1,а, візуальний аналіз котрих дозволяє зробити висновок про високу чутливість нового підходу до виділення значних інформативних характеристик.

3. Представлення аналітичних АКФ в інваріантному просторі моментів Церніке.

Моменти і функції моментів в якості інформаційних характеристик широко використовуються в цілому ряді практичних задач [4]. Один з найбільш поширених підходів базується на відшуканні так званих регулярних моментів, що в дискретній формі можуть бути представлені

$$m_{pq} = \sum_x\sum_y x^p y^q f(x,y), \quad (3)$$

де $(p+q)$ – порядок моменту дискретної функції зображення $f(x,y)$. На основі виразу (3) грунтуються сім інваріантних до зсуву, масштабу і повороту функцій, відомих як інваріантні моменти, однак, базис $x^p y^q$ не має властивості ортогональності, у зв'язку з чим реконструкція зображення на основі цих моментів вимагає значного обсягу обчислень. Для вирішення цієї проблеми було запропоновано ортогональний базис поліномів Церніке [5], коефіцієнти котрого можуть бути використані як інваріантні ознаки зображень.

Набір поліномів Цернакі $\{V_{nm}(x,y)\}$, визначених на одиничному колі $x^2+y^2=1$ в загальному вигляді може бути представлений

$$V_{nm}(x,y) = V_{nm}(\rho,\theta) = R_{nm}(\rho)\exp(jm\theta),$$

де n – порядок поліному, m – параметр повтору, ρ - відстань від центру одиничного кола до пікселя (x,y), θ - кут між вектором ρ та віссю x в напрямку проти годинникової стрілки. Радіальні поліноми $R_{nm}(\rho)$ визначаються

$$R_{nm}(\rho) = \sum_{s=0}^{n-|m|/2}(-1)^s\frac{(n-s)!}{s!\left(\frac{n+|m|}{2}-s\right)!s!\left(\frac{n-|m|}{2}-s\right)!}\rho^{n-2s},$$

причому $R_{nm}(\rho) = R_{nm}(-\rho)$.

Ці поліноми ортогональні і задовільняють

$$\iint_{x^2+y^2\le 1}[V_{nm}(x,y)]^* V_{pq}(x,y)dxdy = \frac{\pi}{n+1}\delta_{np}\delta_{mq}$$

$$\text{при } \delta_{ab} = \begin{cases} 1 & a = b \\ 0 & \text{в іншому випадку.} \end{cases}$$

Моменти Церніке – це проекція функції зображення на базис ортогональних поліномів; для дискретного зображення вираз для комплексних коефіцієнтів розкладення записується

$$A_{nm} = \frac{n+m}{\pi} \sum_x \sum_y f(x,y) V^*_{nm}(\rho,\theta), \quad x^2 + y^2 \le 1$$

В якості інформативних ознак (характеристик Церніке) використовуються амплітудні характеристики $|Anm|$, обчислені для різних порядків поліномів.

Припустимо, вихідне текстурне зображення повертається на довільний кут α, в такому разі зв'язок між вихідним і перетвореним зображенням може бути представлений у вигляді виразу:

$$f^*(\rho,\theta) = f(\rho,\theta-\alpha).$$

Моменти Церніке повернутого зображення, представлені в полярній системі координат, записуються

$$A^*_{nm} = \frac{n+1}{\pi} \int_0^{2\pi} \int_0^1 f(\rho,\theta-\alpha) R_{nm}(\rho) \exp(-jm\theta) \rho d\rho d\theta. \quad (4)$$

Якщо позначити

$$\theta_1 = \theta - \alpha, \quad p = \frac{n+1}{\pi},$$

то (4) набуде вигляду:

$$A^*_{nm} = p \int_0^{2\pi} \int_0^1 f(\rho,\theta_1) R_{nm}(\rho) \exp(-jm(\theta_1+\alpha)) \rho d\rho d\theta_1 =$$

$$= \left[p \int_0^{2\pi} \int_0^1 f(\rho,\theta_1) R_{nm}(\rho) \exp(-jm\theta_1) \rho d\rho d\theta_1 \right] \exp(-jm\alpha) =$$

$$= A_{nm} \exp(-jm\alpha). \quad (5)$$

Рівняння (5) показує, що перетворення повороту зображення може бути виражене через зсув фазової характеристики в той час, як амплітудна характеристика $|Anm|$ лишається незмінною.

Інваріантність до зсуву та масштабу досягається попередньою нормалізацією зображення на основі регулярних інваріантних моментів mpq [6]. Спочатку вихідне зображення $f(x,y)$ трансформується за допомогою моментів першого порядку до вигляду $f(x+\bar{x}, y+\bar{y})$, де $\bar{x}$ і $\bar{y}$ представляють собою центроїди $f(x,y)$

$$\bar{x} = \frac{m_{10}}{m_{00}}, \quad \bar{x} = \frac{m_{01}}{m_{00}}.$$

Іншими словами, початок координат переміщується в точку з координатами центроїдів.

Властивість масштабної інваріантності отримуємо збільшенням/зменшенням зображення так, щоб регулярний момент нульового порядку дорівнював змінній масштаба β.

Нехай $f(x/a, y/a)$ представляє собою масштабовану версію зображення. Тоді регулярні моменти порядку pq вихідного і перетвореного зображення будуть пов'язані на основі виразу:

$$m^*_{pq} = \iint_{xy} x^p y^p f\left(\frac{x}{a}, \frac{y}{a}\right) dxdy =$$

$$= a^{(p+q+2)} \iint_{xy} x^p y^q f(x,y) dxdy = a^{(p+q+2)} m_{pq}.$$

Оскільки необхідно досягти рівності $m_{00} = \beta$, то

$$a = \sqrt{b / m_{00}}, \quad m^*_{00} = a^2 m_{00} = \beta.$$

Таким чином, нормалізоване зображення в результаті має наступний вигляд

$$f_{norm}(x,y) = f\left(\frac{x}{a} + \bar{x}, \frac{y}{a} + \bar{y}\right).$$

4. Метод векторної підтримки.

Класифікатори на основі методу векторної підтримки (МВП) представляють собою чутливий і потужний в розумінні спроможності працювати в просторі параметрів дуже високої розмірності апарат для вирішення задач розпізнавання образів [7].

В рамках спрощеного опису, МВП являє собою задачу побудови гіперплощини, що розділяє вхідні дані $\{xi, yi\}$, $i=1,...,m$; $yi \in \{-1,+1\}$, $xi \in R^n$ таким чином, що забезпечується максимальна відстань між найближчими точками.

В загальному (лінійно нероздільному випадку) задача побудови гіперплощини базується на вирішенні оптимізаційної проблеми виду:

$$\min\left(\frac{|w|^2}{2} + C\left(\sum_i \xi_i\right)\right),$$

$$\text{за обмежень } y_i((w \cdot x_i) + b) \ge 1, \text{ для } \forall i.$$

де w – нормаль до гіперплощини, b – зміщення, ξ_i – «ковзаюча» змінна, що разом з константою C визначає припустиму похибку класифікації. Для вирішення оптимізаційної проблеми використовується математичний апарат Лагранжа [8].

Узагальнення методу до нелінійного випадку досягається відображенням вхідних ознак x_i в деякий багатовимірний простір, розмірність котрого вище, ніж розмірність простору вхідних значень на основі виразу $\Phi: X \to F$; після чого застосовується стандартний алгоритм МВП.

Для здійснення відображення використовуються так звані функції ядра, для котрих виконується умова Мерсера. Аналіз публікацій показує, що найбільш поширеними практично є наступні функції ядра: ядра на основі радіальних базисних функцій, поліноміальні ядра, гіперболічні тангенціальні ядра.

4. Експериментальна частина.

Експерименти проводилися на прикладах 61-го зображення різних фізичних текстур з міжнародної бази даних.

Кожна бінаризована характеристика аналітичної АКФ текстур представлялася в просторі інваріантних моментів Церніке, обчислених для різних порядків поліномів і представлених амплітудними характеристиками.

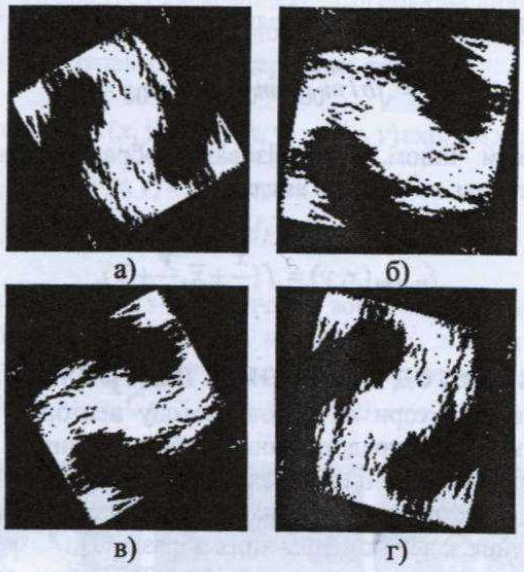


Рисунок 2. Поворот бінаризованої аналітичної кореляційної функції на а) 30°; б) 80°; в) 300°; г) 345°.

В табл. 1 ілюструється інваріантність амплітудних характеристик Церніке порядків n=2, 3, 4 до перетворення повороту бінаризованої АКФ, окремі приклади котрого представлені на рис. 3. В табл. 2 наведені значення стандартного відхилення кожного з наборів характеристик і його відношення до маточікування, аналіз котрих показує дуже низький ступінь розсіювання характеристик для різних кутів повороту АКФ відносно середнього значення.

Таблиця 1. Амплітудні характеристики моментів аналітичної кореляційної функції для перетворень повороту і масштабу.

Перетворення	Амплітудні характеристики моментів			
Поворот	$\lvert A_{20}\rvert$	$\lvert A_{22}\rvert$	$\lvert A_{31}\rvert$	$\lvert A_{42}\rvert$
0°	1605.4	2192.2	7506.9	2005.0
30°	1593.2	2225.9	7644.8	1997.0
45°	1614.0	2229.8	7657.5	1974.1
60°	1611.4	2230.0	7660.2	1989.6

Таблиця 1. Продовження.

Перетворення	Амплітудні характеристики Моментів			
Поворот	$\lvert A_{20}\rvert$	$\lvert A_{22}\rvert$	$\lvert A_{31}\rvert$	$\lvert A_{42}\rvert$
70°	1594.9	2221.2	7629.9	1991.8
80°	1615.5	2220.0	7620.8	1999.4
300°	1593.2	2225.9	7644.8	1997.0
345°	1604.5	2226.2	7649.6	1970.8

Для класифікації, котра проводилася на основі характеристик Церніке до тридцятого порядку (256 ознак для кожної текстури; як було показано в [5], саме такого порядку поліномів достатньо для відновлення зображення, візуально подібного вихідному) було вибрано метод векторної підтримки, що обумовлено його високою точністю, на основі котрого була досягнуто 99% класифікація текстур.

Таблиця 2. Аналіз амплітудних характеристик моментів Церніке.

Параметри	$\lvert A_{20}\rvert$	$\lvert A_{22}\rvert$	$\lvert A_{31}\rvert$	$\lvert A_{42}\rvert$
σ	9.3013	12.308	50.215	12.16
σ/μ	0.0058	0.0055	0.0065	0.0061

4. Висновки

На основі проведених експериментів можна заключити, що метод класифікації на основі інваріантних характеристик Церніке аналітичних АКФ представляє чутливий і стійкий спосіб виділення інформативних ознак текстурних зображень.

Література

1. Haralick R. H. *Statistical and structural approaches to texture.* // Proc. IEEE, 1979, v. 67, pp. 786-804.
2. Tomita F., Tsuji S. *Computer analysis of visual textures.* – Kluwer Academic Publishers, Boston, 1990.
3. Simoncelli E. P. and Portilla J. Texture *Characterization via Joint Statistics of Wavelet Coefficient Magnitudes.* // 5th IEEE Int'l Conf on Image Processing. - Chicago, IL. Oct 4-7, 1998.
4. S. A. Dudani, K. J. Breeding, and R. B. McGhee, *Aircraft identefication by moment invariants*, IEEE Trans. Comput. , vol. C-26, no. 1, pp. 39-45, Jan. 1983.
5. A. Khotanzad and Y. H. Hong, *Rotation invariant pattern recognition using Zernike moments*, in Proc. 9th ICPR, Rome, Italy, Nov. 1998, pp. 326-328.
6. M. K. Hu, *Visual pattern recognition by moments invariants*, IRE Trans. Inform. Theory, vol. IT-8, pp. 179-187, Feb. 1962.
7. Ахметшин О. М. , Довженко О. В., *Підвищення ефективності класифікації в багатовимірному просторі ознак великої розмірності на основі методу векторної підтримки,* Збірка праць МКІМ-2002, т. 4, сс. 36-41, 2002.
8. V. Vapnik, *The nature of statistical learning theory,* Springer – Verlag, N. Y., 1995.

Перетворення палітри кольорових зображень

Роман Воробель, Ігор Журавель, Костянтин Сущик

Фізико-механічний інститут НАН України, 79601, Львів, Наукова5,
vorobel@ah.ipm.lviv.ua, тел. (0322) 654803

Transformation of the color image palette. Roman Vorobel, Igor Zhuravel, Kostantyn Suschyk. The article is devoted to methods, connected with transformation palette of image. The method of optimum transformation palette of image by using a theory of clasterization. The results of test of proposed image processing are shown.

1. Вступ

Одним з важливих чинників, які суттєво впливають на швидкість обробки та передачі інформації, є її об'єм. В даній роботі об'єктом досліджень є цифрові зображення, як один з найбільш зручних способів представлення інформації при неруйнівному контролі, медичній діагностиці та дистанційному зондуванні Землі.

Графічні файли таких зображень, в більшості випадків, потребують значних об'ємів пам'яті, що зумовлює певні незручності при обробці та передачі інформації. Тому актуальною є задача зменшення об'єму графічних файлів зображень при мінімальному погіршенні їх візуальної якості. Вирішення цього питання є також важливим у задачах сегментації кольорових зображень.

Усі зображення можна умовно розділити на чотири типи - бінарні, напівтонові, палітрові та повноколірні. Бінарні та напівтонові зображення представляють собою двохвимірні масиви чисел, які є еквівалентами яскравостей. Повноколірні зображення зберігаються у вигляді тривимірних масивів. Для доступу до значень інтенсивностей складових кольору піксела зображення потрібно вказати рядок - k, стрічку l та номер складової: 1 - для R, 2 - для G та 3 - для B. Палітрові зображення зберігаються у вигляді двохвимірних масивів індексів. Для кожного палітрового зображення існує двовимірний масив палітри, в трьох стовпцях якого розміщені значення інтенсивностей R, G, B.

Кожний графічний файл складається із двох основних частин - заголовка та власне даних. Заголовок містить інформацію про структуру графічного файлу. Здебільшого після заголовку у файлі палітрового зображення міститься палітра кольорів. В даній роботі ми розглядатимемо палітрові повноколірні зображення з метою зменшення кількості кольорів при мінімізації втрати їх якості. Тому з аналізу структури графічного файлу випливає, що одним із шляхів зменшення його об'єму, є оптимізація палітри зображень. Розглянемо відомі підходи до вирішення цього питання [1].

Для більшості зображень, кількість кольорів, яка необхідна для їх відтворення, менша за 256. Забираючи невикористані кольори із палітри сприяємо зменшенню об'єму файлу зображення. Однак даний підхід неефективний для зображень, які містять велику кількість кольорів та відтінків.

Нами запропоновано інший підхід, який базується на тому, що палітра зображення вибирається фіксованою або оптимізується. У випадку фіксованої палітри кожному елементу зображення вибирається у відповідність найближчий за значенням колір. Основною вимогою до такої палітри є те, що вона повинна містити основні кольори, які присутні на зображенні. Виконання цієї вимоги буде в значній мірі впливати на якість відтворення зображення. В більшості випадків використовують так звану "безпечну палітру", яка містить набір часто використовуваних кольорів та їх відтінків. Оскільки розмір палітри є фіксований, а колірне наповнення зображень є найрізноманітнішим, то не завжди вдається провести перетворення палітри зображення із допустимою мінімальною візуальною похибкою. Тому доцільно оптимізовувати палітру.

У випадку оптимізованої палітри, формується список усіх кольорів, які містяться на зображенні. Далі на підставі аналізу частоти появи цих кольорів будується конкретна палітра.

Розглянемо процес аналізу та вибору кольорів оптимізованої палітри детальніше.

2. Оптимізація палітри зображення

Однією з найбільш зручних видів представлення яскравісних характеристик зображення є його гістограма. У випадку кольорових (RGB) зображень гістограма представляється у вигляді куба (рис. 1), всередині якого розміщена інформація про частоту появи кольорів на зображенні.

Побудова тривимірної гістограми зображення є важливим елементом при формуванні оптимальної палітри. Далі на підставі аналізу будують нову палітру зображення. Розглянемо ряд методів визначення нової палітри зображення.

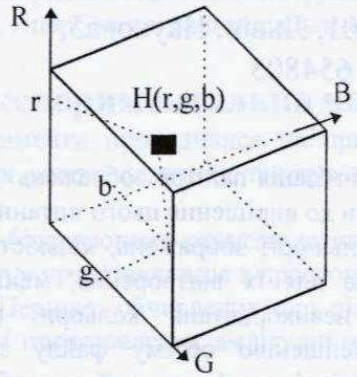


Рис. 1. Структура даних тривимірної гістограми *H(r, g, b)* зображення.

Один із підходів базується на тому, що формується нова палітра розмірності *n×3*, яка складається з найбільш вживаних на зображенні кольорів. Такий алгоритм ефективний для побудови нової палітри зображень, які не містять відтінків основних кольорів. Але більшість зображень містить велику кількість відтінків, тому використання даного методу призводить до того, що палітра перетвореного зображення складатиметься здебільшого з відтінків одного кольору, який найчастіше зустрічається на вхідному зображенні. В результаті такого перетворення, візуальна різниця вхідного та перетвореного зображень є суттєвою, що зумовлює його низьку якість.

Наступним підходом до створення нової палітри, використання якої призвело б до зменшення об'єму графічного файлу зображення, є створення "умовно фіксованої палітри". Суть цього підходу

а)

б)

в)

г)

Рис. 2. Результат комп'ютерного моделювання оптимізації палітри зображення: *а)* вхідне зображення (розмір графічного файлу зображення 131126 байт); *б)* зображення *а)*, оброблене не методом бінаризації з використанням алгоритму Флойда-Стейнберга (8254 байт); *в)* зображення *а)*, палітра якого зменшена до 8 кольорів (32886 байт); *г)* зображення *а)*, палітра якого зменшена до 8 кольорів та використано алгоритм Флойда-Стейнберга (32886 байт).

полягає у тому, що нова палітра потрібного (довільного) розміру формується на основі використання білого, чорного, *RGB*-кольорів та їх відтінків. Цей метод доцільно застосовувати щодо перетворення палітри зображень, які містять всі сім основних кольорів. Недоліком цього методу є те, що у ньому не проводиться аналіз кольорів зображення. Це призводить до того, що побудована палітра в більшості випадків не є оптимальною.

Як було зазначено вище, більшість зображень містить невелику кількість чистих кольорів та велику кількість їх відтінків. В результаті цього елементи тривимірної гістограми розподілені нерівномірно, а певними скупченнями (кластерами). Ця властивість і лягла в основу запропонованого нами методу. У цьому методі для аналізу гістограми зображення використовують один з методів кластеризації. Він полягає у тому, що всі елементи тривимірної гістограми розбиваються на *k*-кластерів таким чином, що мінімізується евклідова відстань між об'єктами всередині кластерів та максимізується евклідова відстань між ними. Вибираючи координати центра мас заданої кількості *k* найбільших кластерів, проводиться побудова оптимальної палітри із заданою розмірністю *k*.

Зменшення об'єму графічного файлу досягається через оптимізацію палітри призводить до погіршення візуальної якості сприйняття зображення. Для зменшення помітності похибки та створення враження, що на зображенні присутня більша кількість різноманітних кольорів, ніж є насправді, використовується дифузійне псевдозмішування кольорів.

3. Дифузійне псевдозмішування кольорів

Потрібний ефект зменшення помітності похибки досягається за рахунок групування в певній локальній області зображення елементів з кольорами, представленими в палітрі, змішування яких дає близький до необхідного колір. Для цього використовується відомий алгоритм Флойда-Стейнберга [1]. В цьому алгоритмі елемент зображення замінює свій колір на найближчий в евклідовій метриці, який є у палітрі. Також обчислюється похибка колірності $Error\_RGB$ для кожної складової як різниця між значеннями величин складових кольорів палітри $MAP\_new(k,:)$ та вхідного зображення $L(i,j,:)$. У випадку *RGB*-складових, алгоритм представляється наступним чином:

$$Error\_RGB(i,j,:) = L(i,j,:) - MAP\_new(k,:). \quad (1)$$

Далі похибка розповсюджується на сусідні елементи зображення шляхом додавання деякої долі похибки до значень відповідних складових і відбувається так звана "дифузія похибки".

$$L(i,j+1,:) = L(i,j+1,:) + \frac{3 \cdot Error\_RGB(i,j,:)}{8}, \quad (2)$$

$$L(i+1,j,:) = L(i+1,j,:) + \frac{3 \cdot Error\_RGB(i,j,:)}{8}, \quad (3)$$

$$L(i+1,j+1,:) = L(i+1,j+1,:) + \frac{Error\_RGB(i,j,:)}{4}. \quad (4)$$

Алгоритм застосовується рекурсивно до всіх пікселів зображення. Використання псевдозмішування кольорів дозволяє покращати колірне сприйняття за рахунок погіршення роздільної здатності та підвищення рівня високочастотного шуму.

4. Висновки

Запропонований у роботі метод побудови оптимальної палітри на основі кластерного аналізу гістограми зображення та використання дифузійного псевдозмішування кольорів дозволяє представляти будь-яке кольорове чи напівтонове зображення заданою кількістю кольорів з мінімальною візуальною похибкою. Похибка перетворення зображення обернено пропорційна кількості елементів оптимальної палітри.

Області застосування методу різноманітні. Важливе місце серед них займає передача відеоданих по каналах зв'язку, де об'єми графічних файлів повинні бути невеликі. Приклад ефективності запропонованого методу ілюструє рис. 2. Тут на рис. 2-а показано вхідне зображення з розмірами файлу у 128 КВ. На рис. 2-б наведено застосування класичної технології дифузійного псевдозмішування кольорів через бінаризацію і застосування алгоритму Флойда-Стейнберга [1-3]. На рис. 2-в показано вхідне зображення 2-а зі зменшеною до восьми кольорів палітрою. Результуюче зображення *г*) отримане із вхідного шляхом оптимізації палітри та застосуванням алгоритму псевдозмішування кольорів Флойда-Стейнберга. За візуальною якістю воно практично не поступається вхідному зображенню, а об'єм його графічного файлу становить лише 32 КВ, що майже у чотири рази менше, ніж у вхідного зображення. Вибором розміру оптимальної палітри шукається компроміс між об'ємом файлу та якістю перетворюваного зображення. При кількості кольорів оптимальної палітри більше ніж 16, втрати якості стають візуально непомітними, а виграш у об'ємі значним.

1. Роджерс Д. Алгоритмические основы машинной графики. М.: Мир, 1989.
2. Шлихт Г.Ю. Цифровая обработка изображений. М.: Эком, 1997.
3. Красильников Н.И. Цифровая обработка изображений. М.: "Вузовская книга", 2001.

Керування п'єзоелектричною компенсацією вимірювальної камери за допомогою нестаціонарного фільтра Калмана

В.П. Савчук, І.М. Глушко

Одеський національний політехнічний університет
м. Одеса, пр. Шевченка, 1, тел. 22-34-74
E. mail: postmaster@opi.odessa.ua

Розглядається задача нестаціонарної фільтрації відгуку складної коливальної системи на зміну амплітуди і фази сигналу її збудження. Рішення цієї задачі зводиться до мінімізації функціонала втрат, що залежить від стану системи, за допомогою фільтра Калмана. У результаті поліпшується якість керування п'єзоелектричною компенсацією коливань унаслідок зменшення впливу перехідних процесів.

ВСТУП

Метод п'єзоелектричної компенсації має широке поширення при градуюванні гідрофонів у діапазоні частот 0,1...1000 Гц [1]. Як правило, цей метод реалізується в камерах малого об'єму, характерний розмір яких менше довжини звукової хвилі у вільному просторі. Камера, показана на рис.1, являє собою два коаксіальних п'єзоелектричних циліндри 1,2, торці яких закриті кришками. Між циліндрами існує пружний зв'язок. Усередині камери, заповненою водою 3, розміщається випробуваний гідрофон 4. В одну з кришок умонтований допоміжний випромінювач 5. Цей випромінювач створює в камері перемінний тиск на заданій частоті, що викликає радіальні коливання внутрішнього циліндра 2 – нуль-індикатора. Коливання нуль-індикатора компенсуються п'єзоелектричною силою випромінювача 1 – нуль-органа. Обидва випромінювачі збуджується спільним генератором 6, що дозволяє змінювати в блоці 7 відносну фазу й амплітуду двох сигналів, а, отже, сил, що діють на нуль-орган. Моменту компенсації відповідає перетворення в нуль амплітуди коливань нуль-індикатора. У цей момент виміряється електрична напруга на виході випробуваного гідрофона пристроєм 10. Тиск у камері розраховується по обмірюваній електричній напрузі збудження нуль-органа в момент компенсації і сталої камери, знайденої в режимі компенсації відомого тиску при її атестації.

Для автоматизації процесу вимірів використовується контур зворотного негативного зв'язку (на рис.1 показаний пунктирною лінією), що складається з пристрою порівняння 8 і блоку формування сигналів керування 9, який реалізує алгоритм регулювання по відхиленню. Об'єкт керування являє собою складну електромеханічну коливальну систему, що має кілька резонансних частот. Навіть у тому випадку, коли ці частоти лежать удалині від робочого діапазону частот, при зміні амплітуди чи фази коливань нуль-органа з'являються перехідні процеси, що вносять збурення в систему керування компенсацією. Для зниження цих збурень був використаний лінійно-квадратичний гауссів регулятор (LQG-регулятор), що складається з лінійно-квадратичного оптимального регулятора (LQ-регулятор) по стану системи і фільтра Калмана, який дає оцінку цього стану.

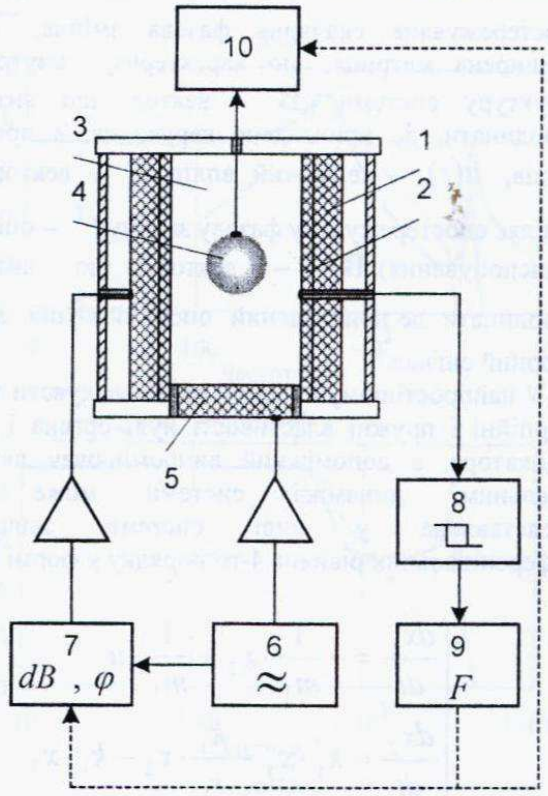


Рис. 1. *Схема градуювання гідрофонів методом п'єзоелектричної компенсації камери малого об'єму.*

1. АЛГОРИТМ КЕРУВАННЯ

Об'єктом керування в нашому випадку є електромеханічна система, що складається з трьох пружно зв'язаних коливальних підсистем. У робочій області частот і на резонансних частотах модель системи можна представити у виді еквівалентної лінійної системи з зосередженими параметрами [2].

Простір станів системи формується з фазових змінних, що представляють еквівалентні механічні переміщення (напруги) і швидкості (струми) її елементів. Об'єкт керування відноситься до класу систем з багатьма входами та одним виходом (MISO). Перший вхід системи (нуль-орган) є керованим, другий вхід (допоміжний випромінювач) є опорним і "заморожується" гармонійним сигналом робочої частоти. Спостерігається тільки одна фазова змінна – відгук нуль-індикатора. Слід зазначити, що вимірник цієї змінної сам має інерційність і пружність. Тому підпростір його станів включено в розширений простір станів об'єкта керування. Динаміка об'єкта керування в просторі станів характеризується наступними матричними рівняннями [3]:

$$\begin{aligned} \dot{\mathbf{X}}(t) &= \mathbf{A}\cdot\mathbf{X}(t)+\mathbf{B}\cdot u(t)+\mathbf{B}_0\cdot s(t) \\ y(t) &= \mathbf{C}^T\cdot\mathbf{X}(t) \end{aligned}, \quad (1)$$

де $\mathbf{X}(t)$ – вектор фазових змінних, що характеризують стан у момент часу t, $y(t)$ – спостережувана скалярна фазова змінна, $\mathbf{A}$ – розширена матриця, що характеризує внутрішню структуру системи, $\mathbf{B}$ – вектор, що визначає координати, де прикладене керування, в просторі станів, $u(t)$ – керуючий вплив, $\mathbf{C}$ – вектор, що виділяє спостережувану фазову змінну (^{T} – операція транспонування), $\mathbf{B}_0$ – вектор, що визначає координати, де прикладений опорний вплив, $s(t)$ – опорний сигнал.

У найпростішому випадку, якщо врахувати тільки інерційні і пружні властивості нуль-органа і нуль-індикатора, а допоміжний випромінювач вважати ідеальним, динаміка системи може бути представлена у виді системи звичайних диференціальних рівнянь 4-го порядку у формі Коші:

$$\begin{cases} \dfrac{dx_1}{dt} = -\dfrac{1}{m_1}\cdot x_2 + \dfrac{1}{m_1}\cdot u \\ \dfrac{dx_2}{dt} = k_1\cdot x_1 - \dfrac{k_1}{r_1}\cdot x_2 - k_1\cdot x_3 \\ \dfrac{dx_3}{dt} = \dfrac{1}{m_2}\cdot x_2 - \dfrac{1}{m_2}\cdot x_4 \\ \dfrac{dx_4}{dt} = k_2\cdot x_3 - \dfrac{k_2}{r_2}\cdot x_2 + \dfrac{1}{m_2}\cdot s \end{cases}, \quad (2)$$

де x_1, x_3 – фазові змінні, що визначають швидкість інерційних елементів (m_1, m_2 – їхні еквівалентні маси) нуль-органа і нуль-індикатора, x_2, x_4 – фазові змінні, що визначають переміщення пружних елементів (k_1, k_2 – їхні еквівалентні коефіцієнти пружності) нуль-органа і нуль-індикатора, r_1, r_2 – опір дисипативних елементів нуль-органа і нуль-індикатора.

Задача поліпшення якості керування полягає в такій корекції керуючого впливу, що дозволяє усунути вплив перехідних процесів у момент зміни амплітуди чи фази гармонійного сигналу, який подається на вхід нуль-органа. Рішення цієї задачі зводиться до мінімізації функціонала втрат, що залежить від стану системи:

$$J = \int_0^\infty \left(\sigma\cdot\mathbf{X}^T\mathbf{X} + \tilde{u}^T\tilde{u}\right)dt, \quad (3)$$

де σ – ваговий коефіцієнт, що визначає значимість помилки спостереження за станом перехідного процесу, $\tilde{u}$ – сигнал керування перехідним процесом (у даному випадку вага помилки керування передбачається рівною одиниці).

Оптимальна стратегія регулювання визначається сигналом керування

$$u^*(t) = u(t) - \mathbf{L}\cdot\mathbf{X}(t), \quad (4)$$

де $\mathbf{L}$ – матриця коефіцієнтів зворотного зв'язку LQ-регулятора, яка мінімізує функціонал

$$j = \int_0^\infty \left(\sigma\cdot y^2 + u^T u\right)dt. \quad (5)$$

Оптимальна оцінка змінних стану при випадкових умовах забезпечується фільтром Калмана. У результаті замкнута система керування, що усуває вплив перехідних процесів, визначається наступними рівняннями:

$$\frac{d}{dt}\begin{bmatrix}\mathbf{X}(t)\\ \hat{\mathbf{X}}(t)\end{bmatrix} = \begin{bmatrix}\mathbf{A}-\mathbf{B}\cdot\mathbf{L} & \mathbf{B}\cdot\mathbf{L}\\ \mathbf{0} & \mathbf{A}-\mathbf{G}\cdot\mathbf{C}\end{bmatrix}\begin{bmatrix}\mathbf{X}(t)\\ \hat{\mathbf{X}}(t)\end{bmatrix} + \begin{bmatrix}\mathbf{B}\\ \mathbf{0}\end{bmatrix}u(t) + \begin{bmatrix}\mathbf{B}_0\\ \mathbf{0}\end{bmatrix}s(t) + \begin{bmatrix}\mathbf{1}\\ \mathbf{1}\end{bmatrix}\xi(t) + \begin{bmatrix}\mathbf{0}\\ -\mathbf{G}\end{bmatrix}\eta(t) \quad (6)$$

де $\hat{\mathbf{X}}(t) = \mathbf{X}(t) - \tilde{\mathbf{X}}(t)$, $\tilde{\mathbf{X}}(t)$ – оцінка фазових змінних фільтром Калмана, $\mathbf{G}$ – матриця коефіцієнтів зворотного зв'язку фільтра Калмана, $\xi(t), \eta(t)$ – шум, що діє на нуль-орган і нуль-індикатор відповідно.

Оптимальна система з замкнутим контуром показана на рис. 2.

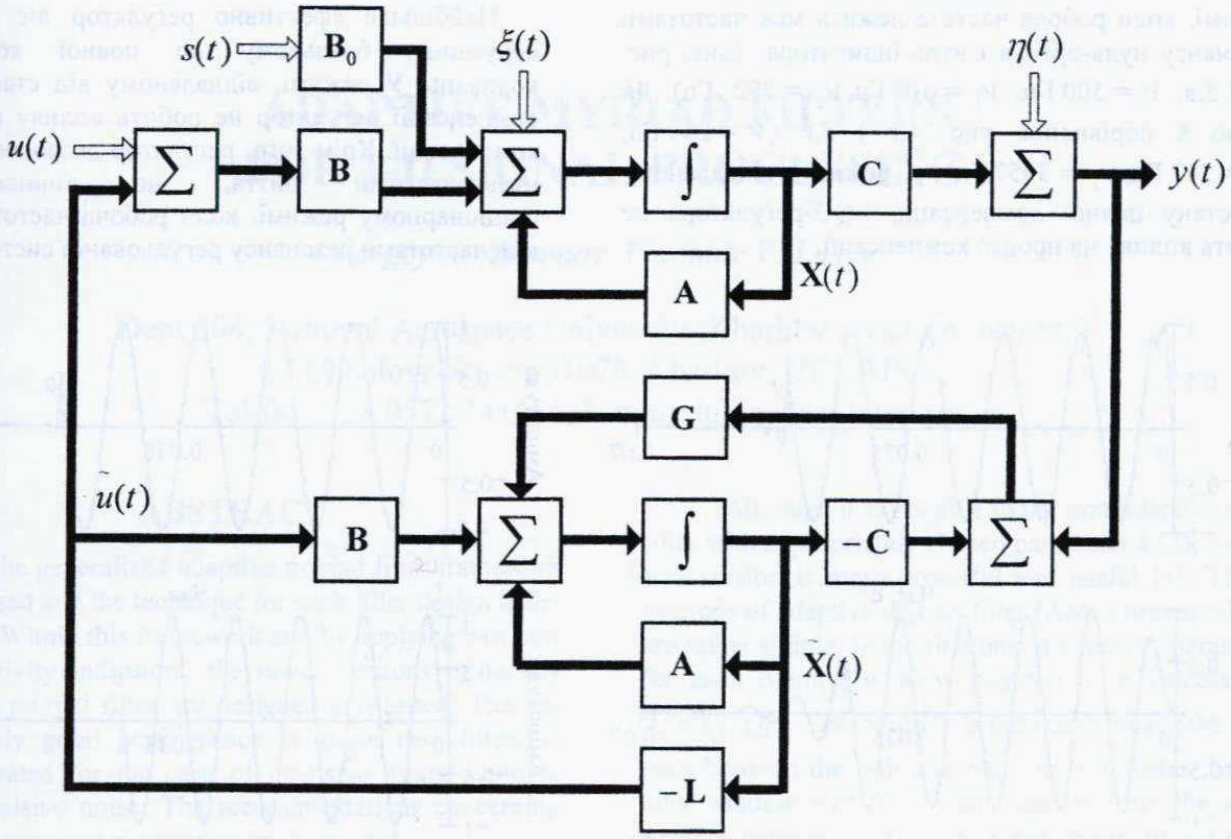


Рис. 2. *Блок-схема замкнутої системи керування компенсацією камери малого об'єму.*

2. РЕЗУЛЬТАТ МОДЕЛЮВАННЯ

Моделювання керування камерою проводилося для системи, динаміка якої характеризується співвідношеннями (2). АФЧХ розімкнутої системи показана на рис. 3,а,в. Резонансна частота нуль-органа $\nu_1 = 547$ Гц, нуль-індикатора – $\nu_2 = 3957$ Гц, їх добротності $Q_1 \approx 4.6$ і $Q_2 \approx 6.4$. На рис. 3,б зображена АЧХ тієї ж системи, але з контуром зворотного зв'язку по стану, підключеному до входу керування. ФЧХ замкнутої і розімкнутої систем практично ідентичні. З порівняння представлених даних видно, що LQG-регулятор усуває низькочастотний полюс системи, практично не змінюючи її фазових властивостей. У даному випадку значимість помилки спостереження $\sigma = 10$.

На рис. 4,а і 5,а зображені відгуки системи при компенсації, проведеної на частоті $\nu = 100$ Гц. У момент часу t_0 зовнішнім керуючим сигналом система переводилася в режим повної компенсації зміною фази сигналу керування на -90^0. Видно, що LQG-регулятор послабляє перехідний процес, що виникає при стрибку фази сигналу керування. Як видно з рис. 4,б і 5,б ($\nu = 120$ Гц, $\nu_1 = 164$ Гц, $\nu_2 = 1189$ Гц) ефект застосування LQG-регулятора зростає при наближенні робочої частоти до частоти резонансу системи. Більш того, LQG-регулятора усуває також низькочастотні биття, що виникають при роботі камери в стаціонарному

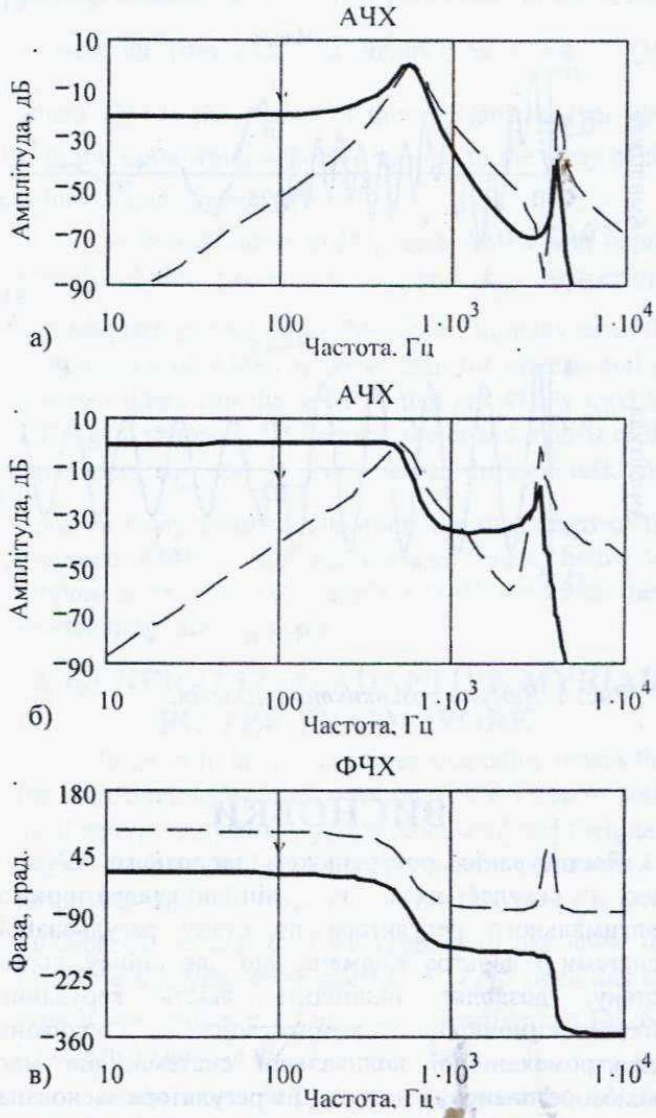


Рис. 3. *АФЧХ (а, в) розімкнутої і АЧХ (б) замкнутої систем: — вхід керування, --- опорний*

режимі, коли робоча частота лежить між частотами резонансу нуль-органа і нуль-індикатора (див. рис. 4,в і 5,в, $\nu = 500$ Гц, $\nu_1 = 109$ Гц, $\nu_2 = 792$ Гц). Як видно з порівняння рис. 4,г і 5,г ($\nu = 200$ Гц, $\nu_1 = 547$ Гц, $\nu_2 = 3957$ Гц) у режимі, віддаленому від стану повної компенсації, LQG-регулятора не робить впливу на процес компенсації.

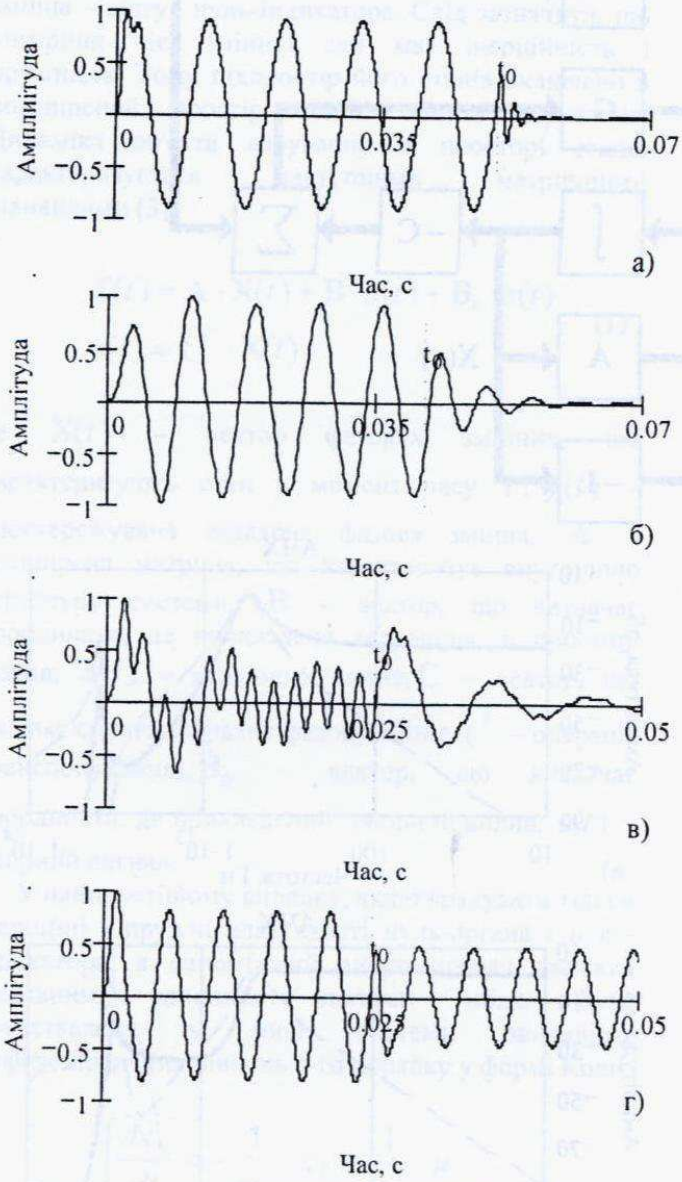


Рис. 4. *Відгуки розімкнутої системи.*

Найбільше ефективно регулятор діє в режимі керування, близькому до повної компенсації коливань. У режимі, віддаленому від стану повної компенсації регулятор не робить впливу на процес компенсації. Крім того, регулятор дозволяє усувати низькочастотні биття, що виникають в стаціонарному режимі, коли робоча частота лежить між частотами резонансу регульованої системи.

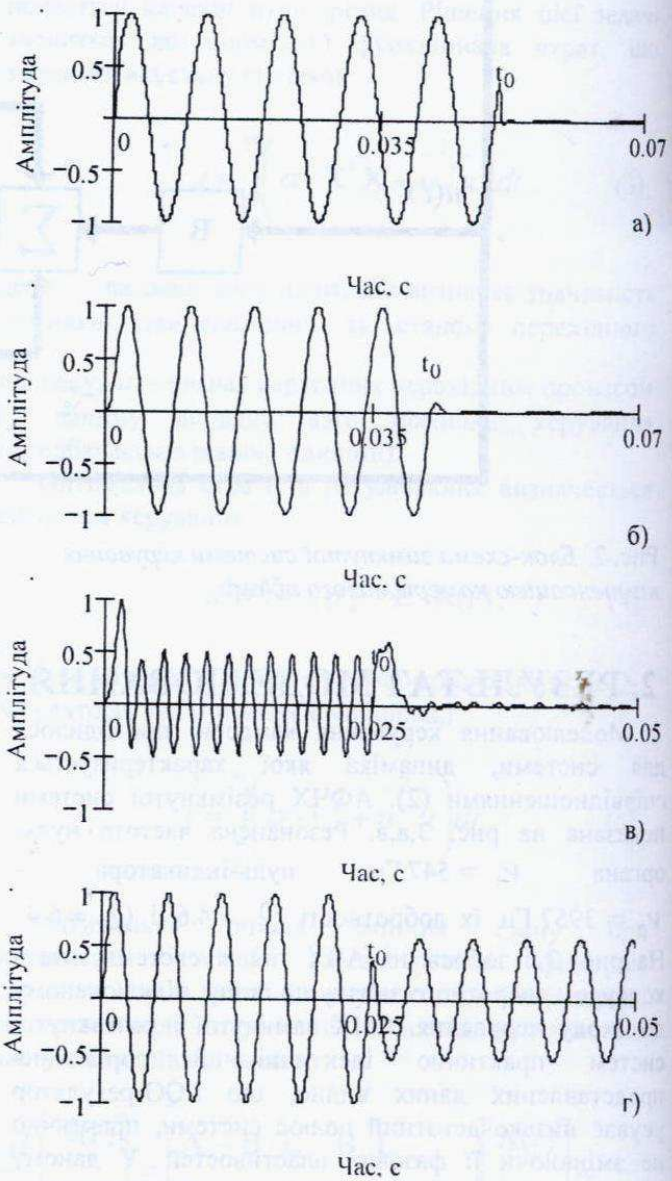


Рис. 5. *Відгуки замкнутої системи.*

ВИСНОВКИ

Застосування рекурсивного зворотного зв'язку, що складається з лінійно-квадратичного оптимального регулятора по стану регульованої системи і фільтра Калмана, що дає оцінку цього стану, дозволяє підвищити якість керування п'єзоелектричною компенсацією коливань електромеханічної коливальної системи, що має кілька резонансних частот. Дія регулятора заснована на зниженні впливу перехідних процесів, що виникають у момент зміни амплітуди чи фази гармонійного сигналу на вході керування.

ЛІТЕРАТУРА

1. Справочник по гидроакустике/ А. П. Евтютов, А. Е. Колесников, Е. А. Корепин и др. Л.: Судостроение, 1988. 522 с.
2. Маццола К. Дж., Бердуэлл Дж. Д., Атанс М. Регулировка излучаемого давления с помощью обратной связи, зависящей от состояния системы.//Подводная акустика и обработка сигналов/Под ред.Л. Бьёрнё. М.: Мир, 1979. С. 251-255.
3. Браммер К, Зиффлинг Г. Фильтр Калмана-Бьюси. М.: Наука, 1982. 200 с.

ADAPTIVE MYRIAD FILTERS FOR 1-D SIGNAL PROCESSING

Sergey K. Abramov, Vladimir V. Lukin

Dept 504, National Aerospace University (Kharkov Aviation Institute)
17 Chkalova Street, 61070, Kharkov, UKRAINE,
Tel/fax + 38 0572 441186, E-mail: lukin@xai.kharkov.ua

ABSTRACT

The generalized adaptive myriad filter framework is proposed and the technique for such filter design is described. Within this framework and by applying two new local activity indicators, the novel versions of locally adaptive myriad filter are designed and tested. The appropriately good performance of these new filters is demonstrated for the case of intensive mixed additive and impulsive noise. The recommendations concerning the filter parameters selection are presented.

1. INTRODUCTION

For many practical applications of 1-D signal processing it is typical that informational signal is corrupted by mixed noise, i.e. simultaneously present additive and impulsive noise. For improving the extraction of useful information from signal or for enhancing the accuracy of its post-processing in these applications it is necessary to reduce the influence of that type of noise. For these purposes linear filters cannot be applied due to presence of impulses, and, furthermore, transitions in informational signal of its own. Therefore, nonlinear filters should be used since they are able to simultaneously remove spikes, suppress the additive noise component and preserve transitions in signals well enough [1].

But nonlinear filters do not possess these "contradictory" properties simultaneously. Filters can only provide some trade-off between additive noise suppression efficiency, transition preservation and impulse rejection; besides, this tradeoff for each nonlinear filter is fixed. This shortcoming can be overcome in the recently introduced myriad filter [2].

The main feature of standard myriad filter (SMF) is that it has a tunable parameter k that controls its behavior. For rather small k values the SMF works in essentially nonlinear mode and possesses good robust properties. On the contrary, when k is rather large, SMF practically performs as mean filter and, consequently, it has good additive noise suppression. Thus, for SMF (in case of the fixed scanning window size N) the trade-off between the additive noise suppression efficiency and the ability to remove spikes can be easily changed by means of varying the tunable parameter k.

Although it is possible to use non-adaptive myriad filter with appropriately chosen parameter k [3], its adaptive version is more powerful and useful [4]. The two versions of adaptive myriad filter (AMF) presented in [4] are rather simple. In the first one, the tunable parameter k for each running window position i is calculated as $k_i = k_Q \cdot Q_i^{pq}$, where Q_i^{pq} is the quasi-range (the difference between the q-th and p-th order statistics for scanning window size N) of data sample from the current running window, and k_Q is a parameter. In the second version, the parameter k is obtained as $k_i = k_{Qf} \cdot Q_i^f$, where Q_i^f is the output of the standard median filter with the same window size N applied to the array of the values of quasi-range Q_i^{pq}.

As it was shown in [4], these AMFs with appropriately chosen parameters k_Q and k_{Qf} outperform non-adaptive myriad filters. Moreover, in many cases the performance of AMFs is better than for median and α-trimmed filters (for the same N) that are widely used for 1-D signal processing. However, the determination of the parameters k_Q and k_{Qf} is a rather difficult task. Besides, in many practical situations the robustness of the proposed AMFs is not appropriate. Hence, below we propose some new versions of the AMF for which these shortcomings are overcome.

2. GENERALIZED ADAPTIVE MYRIAD FILTER FRAMEWORK

In general, the myriad filter adaptation means that for each running window position i the value of some local activity indicator (LAI) is calculated and then, taking into account the LAI value, the tunable parameter k is adaptively adjusted, thus, $k_i = f(LAI_i)$ where f is some function. As a function f we propose to use linear dependence implying some coefficient k_{LAI} since this approach was shown itself as rather effective one [4]. So, let's further suppose that $k_i = k_{LAI} \cdot LAI_i$.

Furthermore, SMF is scale non-invariant [2], i.e. its properties depend upon data scale, hence we propose to use some normalized tunable parameter k/σ_a instead

of parameter k, where σ_a is the standard deviation (STD) of additive noise. So, as will be denoted below, we assume that additive component of the mixed noise obeys Gaussian distribution with the STD σ_a.

Summarizing the aforesaid, our design of the proposed AMFs consists of two stages. At the first one, an optimal value of $(k/\sigma_a)_{opt}$ that provides the best trade-off between additive noise suppression efficiency, signal transitions preservation and spike rejection abilities is found. At the second stage a coefficient $k_{SC} = \sigma_a / LAI$ that is simply the proportionality coefficient between STD and LAI values for the constant level signal fragment is determined. Finally, the k_{LAI} value is calculated as $k_{LAI} = (k/\sigma_a)_{opt} k_{SC}$.

The requirements to the LAI are the following. First, its value should characterize the scale of data samples in the scanning window that are not corrupted by spikes, i.e. LAI should be robust to impulsive noise. Second, for neighborhoods of sharp transitions in signal the LAI values have to be approximately the same as for the fragments of constant signal.

3. PROPOSED LOCAL ACTIVITY INDICATORS

One disadvantage of the already used quasi-range (QR) and median filtered quasi-range (MQR) LAIs [4] is their insufficient robustness (in cases of rather large probabilities of spike occurrence). Another drawback of QR and MQR is that their values are greatly influenced by sharp signal transitions (for example, the neighborhoods of samples with indices 50, 250, 300 of the test signal shown in Figure). So, now we propose to use two new LAIs: 1) the median of the absolute deviation from the median (MAD) [5], and 2) the so-called "minimized" quasi-range (MQR) specially designed by us.

The MAD parameter for the running window position i is defined as $MAD_i = \mathrm{med}\{x_j - \mathrm{med}\{x_i\}\}$, where $\{x_i\}$ is the data sample formed from the values within the current running window and $j = \overline{i - N/2, i + N/2}$. The MQR parameter is expressed as $MQR_i^m = \min_l (Q_l^m)$, $l = \overline{m+1, N-m}$, where $Q_l^m = x^{(l+m)} - x^{(l-m)}$, and m is the parameter that corresponds to the half of the inter-quantile distance $(q-p)/2$ in the usual quasi-range Q_i^{pq}.

The MAD parameter is widely used as robust local scale estimator [5] and, as known, it is more robust than quasi-range. Beside, MAD is less sensitive to the influence of signal sharp transitions.

The use of the minimization operation in the proposed procedure of the MQR allows increasing robustness of the standard QR in the case of identical sign spikes. In such cases QR can reject only p-1 spikes while MQR can reject $2m = q - p = N - 2p + 1$ spikes. Since, for the same considered window size N=9 and for p=3 and q=N-p+1=7 QR can reject only two spikes while MQR can reject four. Because of the same reason, this modification has an improved insensitiveness of LAI to sharp signal transitions.

All the LAI parameters are calculated for the same scanning window position and size N as the proposed AFM output is derived.

4. CONSIDERED SIGNAL/NOISE MODEL AND FILTER PERFORMANCE CRITERIA

As has been mentioned above, we suppose that noise is mixed, and it contains additive and impulsive components. Furthermore, we consider the most complicated case of spikes with identical sign. So the model of a 1-D sampled data sequence we used is

$$y(i) = \begin{cases} S(i) + n_a(i), & \text{with probability } 1 - P_{imp} \\ S(i) + n_{imp}(i), & \text{with probability } P_{imp} \end{cases}, \quad (1)$$

where $S(i)$ denotes the true signal value of the i-th sample; $n_a(i)$ is the zero mean Gaussian additive noise with the variance σ_a^2; $n_{imp}(i)$ defines the amplitude of impulsive noise that occurs with the probability P_{imp}. The test signal shown in Figure consists of different fragments peculated to the most of real-live situations: 1) a constant signal (for example, the fragment with indices from 5 to 45); 2) a step edge (positioned on the 50-th sample); 3) piecewise linear curves (indices 90-110 and 190-210); 4) linearly increasing and decreasing signals (for example, the fragments with indices 110-140 and 160-190), 5) a peak-like maximum (90-110); 6) a polynomial maximum (265-285), and 7) piecewise linear and polynomial curves (the junction points have the indices 250 and 300).

As the general characteristic of the AMFs and others concerned filters efficiency we used the MSE criterion χ_t calculated for the entire test signal. Besides, in order to analyze noise suppression efficiency and transitions preservation separately we calculate some local MSE values χ_{loc} for different signal regions (subindices of χ_{loc} in Table 2 correspond to the fragment margins).

5. AMF PARAMETER SELECTION

In order to determinate the k_{SC} value, the special analysis has been performed. For the constant level signal fragment corrupted by only additive Gaussian noise with several variance values σ_a^2=0.01; 0.03; 0.05; 0.1 (σ_a=0.1; 0.17; 0.22; 0.32) the mean values of the LAI parameters calculated for entire signal were obtained. As it has been established, these mean MAD and MQR LAI values relate to σ_a as $\sigma_a \approx 1.5\,\overline{MAD}$ (see also [5]) and

$\sigma_a \approx 1.22\,\overline{MQR}$, i.e. k_{SC} =1.5 for MAD and 1.22 for MQR.

Generalizing the previously obtained results concerning myriad filter parameter selection [3,4] one can conclude that the optimal $(k/\sigma_a)_{opt}$ value is within the range 0.7÷1.0. So, the optimal value of k_{LAI} is expected to be within the range 1.05÷1.5 for MAD and 0.85÷1.22 for MQR, respectively.

6. NUMERICAL SIMULATION RESULTS

Besides the new AMF versions based on MAD (*Adapt III*) and MQR with m=2 (*Adapt IV*) we also simulated two old AMF versions [4] based on QR with p=3 and q=7 (*Adapt I*) and MQR (*Adapt II*). For comparison we obtained numerical simulation results for the standard median (*Median*) and α-trimmed (*α-Trim*) filters. These results for four different noise environments (different parameters in model (1) - *Cases A, B, C, and D*) are presented in Table 2. The best results (minimal χ values) for each AMF for different k_{LAI} are marked by **bold** formatting and the best result among all considered AMFs is marked by ***bold italic*** formatting.

As seen from Table, for each AMF the optimal value of the k_{LAI} coefficient that provides the best general characteristic χ_t of the filter can be found. The increasing of k_{LAI} value with respect to the value k_{LAI}^{opt} brings to noise suppression improving but signal transition preservation becomes worse. On the contrary, the decreasing of k_{LAI} value leads to noise suppression degradation but signal transition preservation improves.

The best integral characteristics χ_t for all *Cases* are obtained for *Adapt IV* filter (for optimal k_{LAI}^{opt}). As the analysis of the local MSE values shows, in *Cases A*, *B*, and *C*, when the level of the additive and impulsive noise is relatively small, it happens only due to perfect preservation of piecewise linear and polynomial curve (indices 240-260) but for the rest test signal fragments the best results are obtained for *Adapt II* filter. As for the *Case D*, when the destructive influence of mixed noise is very strong, the *Adapt IV* filter is the best for almost all test signal fragments, i.e. this filter is the most robust among the considered AMFs.

The *Adapt III* filter occupies the intermediate place between *Adapt II* and *Adapt IV* filters. This means that MSE χ_t for *Adapt III* filter is usually a little bit worse than for *Adapt IV*, but it is better than for *Adapt II*. On the other hand, almost all χ_{loc} values (except $\chi_{240-260}$) for the *Adapt III* filter are better than for *Adapt IV*, but they are worse than for *Adapt II*.

It is especially interesting that the proposed AMFs provide better preservation of step edge than the standard median filter commonly considered the best in this sense.

CONCLUSIONS

The proposed AMFs based on MAD and MQR LAIs possess good noise suppression efficiency and, at the same time, accurate signal transitions preservation. For the considered test signal and noise properties they improve input SNR by 7...11 dB. These filters (in the sense of total MSE χ_t) outperform the median and α-trimmed filters that are widely used for 1-D signal processing. Moreover, the characteristics of the proposed AMFs are better than for previously designed AMFs, especially for the case of very intensive additive and impulsive noise (σ_a^2=0.03, P_{imp}=0.1).

REFERENCES

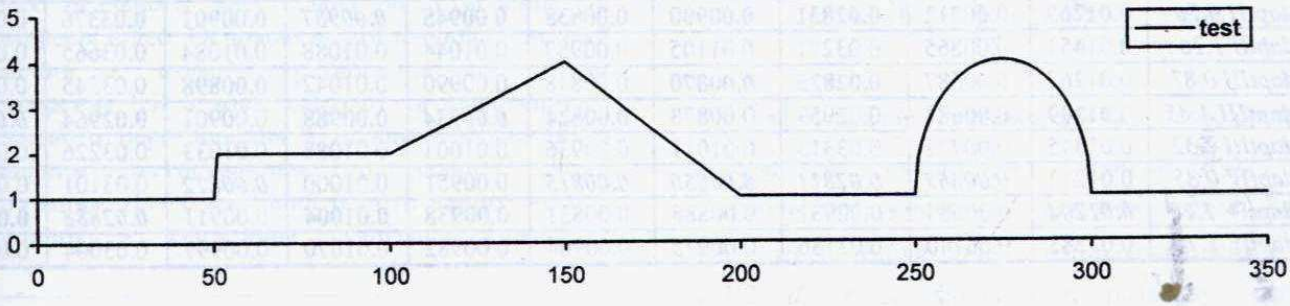
[1] Pitas I., Venetsanopoulos A.N. *Nonlinear Digital Filters: Principles and Applications*. New York: Kluwer Academic Publishers, 1990.

[2] Kalluri S, Arce G. *Adaptive weighted myriad filter algorithms for robust signal processing in α-stable noise environments*. IEEE Trans. On Signal Processing, Vol. 46, pages 322-334, Feb. 1998.

[3] Abramov S.K., Lukin V.V., Astola J.T. *Myriad filter properties and parameter selection*. - Proc. of the Fifth All-Ukrainian international conference UkrOBRAZ'2000, 2000, Kyjiv, Ukraine, p.p.59-62.

[4] Abramov S., Lukin V., Astola J. *Adaptive myriad filter*. - CD-ROM Proceedings of NSIP'2001, June 2001, Baltimore MD USA, 5 p.

[5] Astola J., Kuosmanen P. *Fundamentals of Nonlinear Digital Filtering*. CRC Press LLC, USA, 1997.



The considered noise-free test signal

Table. The entire and local MSEs for the considered filters (N=9).

	χ_t	χ_{10-40}	χ_{40-60}	χ_{90-110}	$\chi_{110-140}$	$\chi_{140-160}$	$\chi_{160-190}$	$\chi_{190-210}$	$\chi_{240-260}$	$\chi_{265-285}$
Case A: σ_a^2=0.01; P_{imp}=0; n_{imp}=0										
Noise	0.01000	0.00995	0.01016	0.01000	0.01005	0.00998	0.01008	0.00981	0.00991	0.01001
Median	0.00273	0.00166	0.00407	0.00213	0.00237	0.00369	0.00309	0.00244	0.00542	0.00228
α-Trim	0.00441	0.00129	0.02206	0.00153	0.00145	0.00256	0.00154	0.00163	0.01485	0.00160
AdaptI 0.50	0.00316	0.00166	0.00795	0.00212	0.00232	0.00358	0.00291	0.00235	0.00723	0.00223
AdaptII 0.70	**0.00316**	0.00166	0.00795	0.00212	0.00232	0.00358	0.00291	0.00235	**0.00723**	0.00223
AdaptII 1.26	0.00333	***0.00124***	**0.00288**	***0.00150***	***0.00146***	***0.00254***	***0.00147***	***0.00159***	0.01581	***0.00156***
AdaptIII 0.87	0.00290	0.00182	**0.00258**	0.00236	0.00272	0.00406	0.00357	0.00270	0.00581	0.00251
AdaptIII 1.45	**0.00229**	0.00152	0.00434	0.00186	0.00197	0.00314	0.00230	0.00203	**0.00392**	0.00197
AdaptIII 2.32	0.00266	**0.00133**	0.00929	**0.00157**	**0.00154**	**0.00264**	**0.00162**	**0.00169**	0.00657	**0.00167**
AdaptIV 0.85	0.00256	0.00171	***0.00246***	0.00217	0.00243	0.00371	0.00307	0.00244	0.00462	0.00229
AdaptIV 1.28	***0.00218***	0.00148	0.00361	0.00181	0.00190	0.00305	0.00218	0.00197	***0.00374***	0.00191
AdaptIV 1.70	0.00227	**0.00136**	0.00573	**0.00162**	**0.00161**	**0.00273**	**0.00174**	**0.00175**	0.00483	**0.00172**
Case B: σ_a^2=0.01; P_{imp}=0.03; n_{imp}=1.0										
Noise	0.03993	0.03949	0.03783	0.04029	0.03934	0.04253	0.04102	0.03964	0.04015	0.04072
Median	0.00413	0.00185	0.00971	0.00244	0.00271	0.00399	0.00385	0.00304	0.01232	0.00235
α-Trim	0.00531	0.00159	0.02439	0.00192	0.00189	0.00293	0.00232	0.00231	0.01846	0.00179
AdaptI 0.50	0.00423	0.00178	0.01212	0.00222	0.00247	0.00375	0.00328	0.00268	0.01310	0.00226
AdaptII 0.70	**0.00388**	0.00159	***0.00835***	0.00203	0.00226	0.00341	0.00279	0.00243	**0.01373**	0.00201
AdaptII 1.26	0.00443	***0.00142***	0.00962	***0.00169***	***0.00171***	***0.00280***	***0.00204***	***0.00207***	0.01989	***0.00169***
AdaptIII 0.87	0.00412	0.00189	**0.00919**	0.00244	0.00280	0.00415	0.00384	0.00296	0.01200	0.00249
AdaptIII 1.45	**0.00334**	0.00160	0.01034	0.00197	0.00211	0.00332	0.00269	0.00236	**0.00873**	0.00197
AdaptIII 2.32	0.00359	**0.00145**	0.01381	**0.00176**	**0.00180**	**0.00295**	**0.00228**	**0.00215**	0.01068	**0.00173**
AdaptIV 0.85	0.00372	0.00176	**0.00915**	0.00223	0.00251	0.00380	0.00333	0.00269	0.01043	0.00226
AdaptIV 1.28	***0.00322***	0.00155	0.00987	0.00192	0.00205	0.00324	0.00260	0.00231	***0.00838***	0.00191
AdaptIV 1.70	0.00326	**0.00146**	0.01134	**0.00180**	**0.00186**	**0.00302**	**0.00236**	**0.00219**	0.00899	**0.00176**
Case C: σ_a^2=0.03; P_{imp}=0.03; n_{imp}=1.0										
Noise	0.06012	0.06009	0.06176	0.06121	0.06079	0.06154	0.05925	0.06031	0.05879	0.05974
Median	0.00844	0.00559	0.01628	0.00602	0.00638	0.00755	0.00724	0.00673	0.01945	0.00601
α-Trim	0.00852	0.00447	0.02839	0.00481	0.00481	0.00574	0.00508	0.00521	0.02332	0.00467
AdaptI 0.50	0.00817	0.00535	0.01754	0.00564	0.00605	0.00734	0.00680	0.00635	0.01854	0.00583
AdaptII 0.70	**0.00758**	0.00475	***0.01379***	0.00512	0.00553	0.00682	0.00627	0.00574	**0.01947**	0.00527
AdaptII 1.26	0.00776	***0.00424***	0.01850	***0.00453***	***0.00463***	***0.00569***	***0.00496***	***0.00495***	0.02346	***0.00456***
AdaptIII 0.87	0.00852	0.00577	**0.01542**	0.00612	0.00667	0.00806	0.00762	0.00692	0.01893	0.00635
AdaptIII 1.45	**0.00755**	0.00489	0.01890	0.00516	0.00545	0.00664	0.00597	0.00575	**0.01655**	0.00529
AdaptIII 2.32	0.00780	**0.00447**	0.02361	**0.00475**	**0.00487**	**0.00590**	**0.00520**	**0.00523**	0.01974	**0.00477**
AdaptIV 0.85	0.00795	0.00538	**0.01554**	0.00568	0.00615	0.00746	0.00691	0.00639	0.01729	0.00589
AdaptIV 1.28	***0.00736***	0.00478	0.01826	0.00505	0.00533	0.00649	0.00582	0.00562	***0.01608***	0.00518
AdaptIV 1.70	0.00745	**0.00452**	0.02111	**0.00480**	**0.00497**	**0.00603**	**0.00534**	**0.00531**	0.01760	**0.00485**
Case D: σ_a^2=0.03; P_{imp}=0.1; n_{imp}=1.0										
Noise	0.13052	0.12980	0.13029	0.13455	0.13083	0.13012	0.12653	0.13243	0.12454	0.12877
Median	0.01398	0.00784	0.02967	0.01007	0.00978	0.01053	0.01183	0.01063	0.03580	0.00848
α-Trim	0.01513	0.00926	0.03775	0.01168	0.01048	0.01061	0.01128	0.01153	0.03551	0.00893
AdaptI 0.50	0.01266	0.00711	0.02872	0.00891	0.00852	0.00966	0.01010	0.00909	0.03274	0.00778
AdaptII 0.70	**0.01267**	**0.00712**	**0.02831**	**0.00900**	**0.00838**	**0.00945**	**0.00987**	**0.00901**	**0.03376**	**0.00754**
AdaptII 1.26	0.01451	0.00865	0.03271	0.01105	0.00987	0.01044	0.01088	0.01084	0.03665	0.00858
AdaptIII 0.87	0.01267	0.00687	**0.02825**	**0.00870**	0.00848	0.00990	0.01042	**0.00898**	0.03345	0.00793
AdaptIII 1.45	**0.01209**	**0.00681**	0.02959	0.00878	**0.00824**	***0.00934***	**0.00988**	0.00901	**0.02964**	***0.00740***
AdaptIII 2.32	0.01345	0.00772	0.03313	0.01011	0.00936	0.01003	0.01088	0.01033	0.03226	0.00805
AdaptIV 0.85	0.01213	***0.00667***	***0.02811***	***0.00850***	***0.00815***	0.00951	0.01000	***0.00872***	0.03101	0.00759
AdaptIV 1.28	***0.01204***	0.00683	0.02937	0.00888	0.00831	**0.00938**	**0.01004**	0.00911	***0.02888***	**0.00742**
AdaptIV 1.70	0.01283	0.00740	0.03136	0.00975	0.00904	0.00982	0.01070	0.00997	0.03004	0.00781

Хвилькове та Карунена-Лоева перетворення зображень у реальному часі

Мусатенко Ю. С., Курашов В. Н.

Лабораторія оптичної та мікрохвильової обробки інформації та теорії середовищ,
радіофізичний факультет, Київський національний університет імені Тараса Шевченка
Тел. (044) 266-0483; E-mail: ysm@mail.univ.kiev.ua

В роботі представлений розроблений авторами програмний пакет для здійснення хвилькового (wavelet)[1] та Карунена-Лоева(КЛ) перетворень зображень в реальному часі на персональному комп'ютері. При розробці пакету максимальна увага приділялась отриманню високої швидкодії. Для забезпечення високих швидкісних показників реалізацію перетворень було здійснено в цілих числах з використанням технології ММХ. Розроблений пакет призначений для забезпечення роботи в реальному часі системи стиснення наборів зображень, розробленої авторами раніше.

ВСТУП

Ортогональні перетворення широко використовуються в системах стиснення зображень. Найбільш широко використовуваним, мабуть, є блочне косинусне перетворення, на основі якого побудований популярні стандарти стиснення JPEG та MPEG-2. Проте, на сьогоднішній день, було запропоновано більш ефективні системи стиснення на основі хвилькового перетворення [1-2] та перетворення КЛ [3] (в тому числі авторами даної роботи [4]). Як правило, базовим алгоритмом для даних систем стиснення є алгоритм EZW[4] і хвилькове перетворення. Багато інших систем стиснення використовують модифікації даного алгоритму з деякими покращеннями. Одним з основних недоліків таких систем є їх вища складність реалізації та низька швидкодія, що не дозволяє їм конкурувати з відомими стандартами в задачах реального часу таких як стиснення та декомпресія відео. Основою нашої існуючої системи стиснення є схема перетворення на основі перетворення КЛ вздовж осі часу та двовимірного хвилькового перетворення в площині зображень з подальшим застосуванням оригінальної схеми квантування. Для ефективної роботи даного методу необхідною є ефективна реалізація кожної з складових. В даній роботі представлений пакет, який з високою швидкодією реалізує частину системи стиснення, яка відповідає перетворенню КЛ та хвильковому перетворенню.

Для хвилькового перетворення використовується відомий фільтр 9/7, який є одним з найкращих з точки зору використання в задачах стиснення з втратами.

Робота має таку структуру. В першому розділі коротко розглянуто перетворення КЛ, хвилькове перетворення та використовувані алгоритми їх здійснення. В ньому ж представлено схему розвинення для якої необхідно забезпечити високу швидкодію. В розділі 2 розглянуто реалізацію згаданих перетворень в цілих числах з використанням технології ММХ. В третьому розділі представлено результати тестування швидкодії та точності відновлення. Наприкінці приведені висновки з роботи.

1. ПЕРЕТВОРЕННЯ СИГНАЛІВ ТА ЗОБРАЖЕНЬ

1.1 Перетворення Карунена-Лоева

Зміст перетворення КЛ полягає в мінімізації середньоквадратичної похибки ε^2 для K-вимірної проекції сигналу $\{U \in R^d\}$ для всіх K. Іншими словами, ми маємо знайти ортогональний базис $\left\{\phi_k^0\right\}_0^d$ такий, що задовольняє умові:

$$\varepsilon^2 = \left\langle \left\| U_n - \sum_{k=1}^{K} a_n^k \phi_k^0 \right\|^2 \right\rangle_n = \min . \qquad (1)$$

Тут $|*|$ позначає L^2 норму, а <*> - усереднення по генеральній сукупності. Функції ϕ_n^0, що задовольняють умові (1.1), утворюють базис КЛ для заданого ансамблю $\{U\}$; а набір коефіцієнтів $\left\{a_n^k\right\}$, k=1,2,...d, є коефіцієнтами розкладу сигналів ансамблю в базис. Варіаційна задача (1) дає рівняння на власні значення, котре відоме, як теорема КЛ [3]:

[1] Термін "хвилькове перетворення" є перекладом з англійської виразу "wavelet transform", даний переклад автори вважають найбільш вдалим.

$$G\phi_k^0 = \lambda_k \phi_k^0 , \qquad (2)$$

де G – кореляційна матриця вибірки $\{U\}$, а λ_k – власні числа матриці G. Оскільки кореляційна матриця після перетворення КЛ є діагональною, то координати $\{\mu_n^k\}$ векторів U_n в просторі, породженому базисними векторами $\{\phi_k^0\}_0^K$, є некорельованими випадковими величинами.

Основною проблемою, пов'язаною з перетворенням КЛ, є громіздкість чисельного розв'язку рівняння (2), яке має складність $O(Nd^2 + d^3)$; N – обсяг навчаючої вибірки.

1.1 Хвилькове перетворення

Останнім часом в системах обробки інформації широко використовується хвилькове перетворення. Воно забезпечує т.зв. мультирозрізнювальну декомпозицію, подає ансамбль через набір промасштабованих "копій" і є потужним засобом для аналізу та стиснення зображень, виділення ознак та розпізнавання образів. Однією з важливих переваг цього перетворення є низька складність, вона лінійно зростає з кількість відліків. Нижче ми розглянемо основні принципи хвилькового перетворення[1-2].

Поняття мультирозрізнювального аналізу дозволило відшукати метод побудови хвилькового ортонормованого базису для простору сигналів скінченної енергії $L_2(R)$. У 1989 році Маллат[2] представив цей математичний метод, основна ідея якого полягає в такому. Ми маємо послідовність підпросторів $V_j \subset L_2(R)$, $V_j \subset V_{j-1}$, $j \in Z$ і функцію $\varphi(x)$, таку, що послідовність функцій $\left\{2^{-j/2}\varphi\left(2^{-j}x-n\right)\cdot\right\}$, $n \in Z$, утворює ортонормований базис у V_j. Традиційно W_j позначає ортогональне доповнення V_j в V_{j-1}, так що $V_{j-1} = V_j \oplus W_j$, де символ $\oplus$ позначає суму підпросторів. Підпростори V_j обрано таким чином, що підпростори W_j є ортогональні між собою. W_j містить деталі j-го рівня аналізованого сигналу, тому, що він отримується як різниця підпросторів V_{j-1} і V_j, які містять апроксимації сигналу j-1-го і j-го рівня відповідно. Можна показати, що можливо створити материнську функцію $\psi(x)$, таку, що послідовність функцій $\left\{2^{-j/2}\psi\left(2^{-j}x-n\right)\cdot\right\}$, $n \in Z$, утворює ортонормований базис у W_j. Звідси очевидно, що набір функцій $\left\{2^{-j/2}\psi\left(2^{-j}x-n\right)\cdot\right\}$, $n \in Z$, утворює ортонормований базис у $L_2(R)$. Апроксимація і деталі сигналу отримуються його ортогональним проектуванням у апроксимуючий і різнецевий підпростори. Інформація, котра міститься у підпросторах V_j та W_j може розглядатись як апроксимація та деталі j-го рівня сигналу $f(x)$ відповідно. Проекції сигналу в j-тий апроксимуючий та різнецевий підпростори є:

$$c_n^j = \int_{-\infty}^{+\infty} f(x)\cdot 2^{-j/2}\varphi(2^{-j}x-n)\cdot dx;$$
$$d_n^j = \int_{-\infty}^{+\infty} f(x)\cdot 2^{-j/2}\psi(2^{-j}x-n)\cdot dx. \qquad (3)$$

В дійсності коефіцієнти подання сигналу, що відповідають підпросторам V_j та W_j обчислюються з попереднього рівня апроксимації V_{j-1} за допомогою пари фільтрів h і g за алгоритмом Маллата:

$$c_n^j = \sum_k \tilde{h}(2n-k)\cdot c_k^{j-1},$$
$$d_n^j = \sum_k \tilde{g}(2n-k)\cdot c_k^{j-1}. \qquad (4)$$

Відновлення сигналу відбувається з використанням майже тих самих фільтрів і зворотнього ходу алгоритму:

$$c_n^{j-1} = \sum_k h(n-2k)\cdot c_k^j + \sum_k g(n-2k)\cdot d_k^j . \qquad (5)$$

Тут відгуки пари фільтрів традиційно позначаються символами h та g для відновлюючої частини і $\tilde{h}(n)$ $\tilde{g}(n)$ для аналізуючої частини. Вони між собою пов'язані простими рівностями.

$$\tilde{h}(n) = h(-n),$$
$$\tilde{g}(n) = g(-n) . \qquad (6)$$

Якщо оператори $\hat{H}$ і $\tilde{G}$ позначають згортку сигналу з низькочастотним та високочастотним відгуками h та g відповідно, то декомпозиція сигналу з (j-1)-го рівня в j-й рівень здійснюється застосуванням операторів $\hat{H}$ і $\tilde{G}$ до коефіцієнтів (j-1)-го рівня. В результаті цього процесу ми отримаємо бінарне дерево для обчислення коефіцієнтів розвинення в набір підпросторів, котрий визначається хвильковим мультирозрізнювальним аналізом. Це дерево зображено на Рис.1. На кожному кроці ми здійснюємо розкладання коефіцієнтів розвинення сигналу в дві частотні смуги за допомогою фільтрів $\hat{H}$ і $\tilde{G}$ та рекурсивно продовжуємо цю фільтрацію для низькочастотної частини розвинення сигналу. Коефіцієнти розкладання по хвильковому базису лишатимуться у блоках, котрі виділені на рисунку. Відновлення здійснюється за подібною схемою, тільки в зворотній бік.

Крім ортогональних хвилькових фільтрів часто користуються неортогональними, серед яких найпоширенішими є т.зв. біортогональні фільтри. До таких фільтрів відноситься і використовуваний нами відомий фільтр 9/7. Особливістю цих фільтрів є те, що не зважаючи на відсутність ортогональності вони дозволяють повне відновлення, крім того на відміну від ортогональних фільтрів вони можуть бути симетричними для довжин фільтра >2. Повне

відновлення можливе за рахунок того, що фільтрація вперед є проектуванням в деякий афінний базис, а зворотня фільтрація являє собою відновлення за допомогою взаємного базису. В зв'язку з цим фільтри h та g пов'язані з $\widetilde{h}(n)$ та $\widetilde{g}(n)$ зовсім інакшим чином, але фільтрація проводиться так само. Свою назву (біортогональні) ці фільтри отримали через те, що попарно ортогональними є фільтри h та $\widetilde{g}(n)$ і g та $\widetilde{h}(n)$. В багатьох роботах відмічено вищі показники біортогональних фільтрів поряд з ортогональними в задачах стиснення зображень.

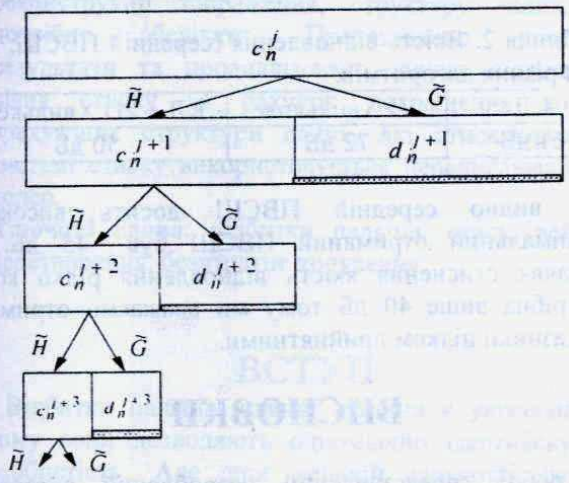


Рис. 1 Дерево хвилькових розвинень.

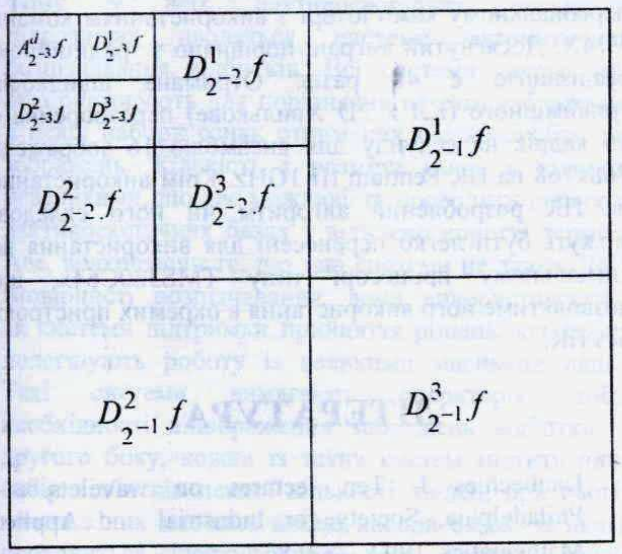


Рис. 2 Схема двовимірного хвилькового перетворення.

При застосуванні до зображень часто користуються простим сепарабельним узагальненням одновимірного хвилькового перетворення на двовимірний випадок. Як можна показати, схема перетворення, тоді має вигляд представлений на Рис. 2. При цьому оператор $D^k_{2^{-j}}f$, k=1,2,3 відповідає відшуканню деталей j-го рівня з k-ю орієнтацією, а оператор $A^d_{2^{-j}}f$ є оператором побудови апроксимації сигналу. В одновимірному випадку є два оператори, а в двовимірному – чотири, які можна подати як добутки одновимірних.

Для задач стиснення наборів зображень нами проводиться перетворення фактично тривимірного набору даних. При цьому здійснюється перетворення КЛ по вісі часу, і двовимірне хвилькове перетворення в площині зображень. При цьому схема перетворення виглядає так як на Рис.3.

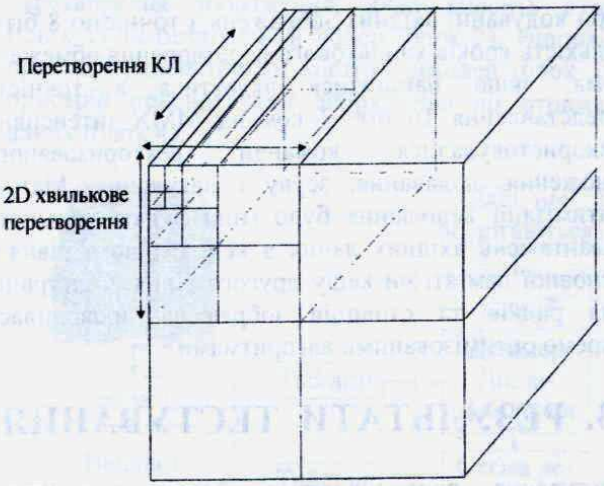


Рис. 3. Схема розрізнення при застосуванні перетворення КЛ по часовій осі і 2D хвилькового перетворення в площині зображень.

Далі нами розглядатиметься яким чином забезпечити здійснення представленого тривимірного перетворення в реальному часі.

2. РЕАЛІЗАЦІЯ ПЕРЕТВОРЕНЬ

Побудована ефективна реалізація описаного тривимірного перетворення містила такі складові.

1. У всіх найважчих ділянках використовувалися цілі числа з точністю 16 біт.
2. Найважчі обчислення здійснювалися з використанням команд MMX.
3. Для перетворення КЛ і хвилькового перетворення проводилася грунтовна оптимізація стратегії кешування.

Типовим ансамблем, для якого необхідно здійснювати перетворення є 16 зображень 768x768 з глибиною сірого 8 бітів. Зрозуміло, що оптимізація проводилася саме для подібних конфігурацій. Найважчими ділянками, які вимагали суттєвих оптимізацій були такі:

- Обчислення кореляційної матриці часових послідовностей пікселів. (Векторів часових послідовностей стільки, скільки пікселів у зображенні, тобто дуже багато).
- Отримання коефіцієнтів перетворення КЛ для кожного вектора часової послідовності пікселів.

Таблиця 1. Порівняльна таблиця швидкодії алгоритмів перетворень реалізованих з командами MMX та з плаваючою комою для ансамблю 16 зображень 768x768 на ПК Pentium III 1GHz.

Перетво-рення	Для алгоритму з командами MMX			Для алгоритму з плаваюч. комою			Пришвид шення, разів
	Час всього, мс	Час на одне зображ., мс	Тактів на піксель	Час всього, мс	Час на одне зображ., мс	Тактів на піксель	
КЛ	336	21	35	20480	1280	2170	62
2D Хилькове	672	42	70	27200	1700	2940	42
КЛ+Хвильк.	1008	63	105	47680	2980	5110	48

- Фільтрація за виразами (4)-(5) окремо для рядків та стовпців зображень.

При кодуванні вхідних зображень з точністю 8 бітів кількість кроків хвилькового перетворення обмежена 4-ма, якщо намагатися вкластися в точність представлення 16 біт. З команд MMX інтенсивно використовувалися команди векторизованого множення, додавання, зсуву і пакування. Метою оптимізації кешування було мінімізувати кількість завантажень вхідних даних в кеш першого рівня з основної пам'яті чи кешу другого рівня. Фільтрація для рядків та стовпців зображень проводилася окремо оптимізованими алгоритмами.

3. РЕЗУЛЬТАТИ ТЕСТУВАННЯ

Тестування проводилося з метою порівняння швидкодії з існуючими реалізаціями в числах з плаваючою комою та визначення точності відновлення. Для оцінки швидкодії використовувався ансамбль 16 зображень 768x768. Проводилося порівняння розробленого алгоритму реального часу з таким самим алгоритмом реалізованим в числах з плаваючою комою на ПК Pentium III 1GHz. Крім цього проводилося порівняння для складових – окремо для перетворення КЛ та для 2D хвилькового перетворення. Результати порівняння представлені в Таблиці 1. Видно, що проведені оптимізації дозволили досягнути великого виграшу в швидкості, а саме, в 40-60 разів. Як видно час обробки в перерахунку на одне зображення є 63мс, що дозволяє обробляти 15 зображень за секунду, а для зображень 512x512 вдасться обробляти 30 зображень на секунду. Подібні показники дозволяють використовувати даний алгоритм в системах реального часу. Зрозуміло, що підвищення частоти процесора, та перехід від команд MMX до SSE2 дасть ще додаткове підвищення швидкодії.

Оскільки перетворення здійснювалися в цілих числах з обмеженою точністю, то проводилося тестування якості відновлення. Якість оцінювалася на основі Пікового Відношення Сигнал Шум(ПВСШ) в децибелах, яке визначається так:

$$ПВСШ = 20 \log(255/MSE) \quad (7)$$

де 255 - максимальна амплітуда значень пікселів в зображенні, а MSE - величина середньоквадратичного відхилення пікселів відновленого зображення від оригіналу. Як тестовий використовувався набір зображень Mobile.

Таблиця 2. Якість відновлення (середній ПВСШ, дб) для різних алгоритмів.

КЛ	2D Хвилькове	КЛ + 2D Хвилькове
61 дБ	72 дБ	50 дБ

Як видно середній ПВСШ досить високий. Мінімальний отриманий ПВСШ був ~45 дБ. В задачах стиснення якість відновлення рідко коли потрібна вище 40 дБ тому ми вважаємо отримані показники цілком прийнятними.

ВИСНОВКИ

В роботі представлений розроблений авторами програмний пакет для здійснення хвилькового та КЛ перетворень зображень в реальному часі на персональному комп'ютері з використанням команд MMX. Досягнутий виграш порівняно з традиційною реалізацією є 48 разів. Отримана швидкодія тривимірного (КЛ x 2D Хвилькове) перетворення є 15 кадрів на секунду для ансамблю 16 зображень 768x768 на ПК Pentium III 1GHz. Крім використання на ПК розроблений алгоритм чи його складові можуть бути легко перенесені для використання на сигнальному процесорі типу TMS320C64x, що дозволятиме його використання в окремих пристроях без ПК.

ЛІТЕРАТУРА

1. Daubechies I. Ten lectures on wavelets. — Philadelphia, Society for Industrial and Applied Mathematics, 1992. — 358c.
2. Mallat S.G. A Theory for multiresolution signal decomposition: the wavelet representation // IEEE Trans. Pattern Anal. Machine Intell. — 1989. — v.31. — P. 674-693.
3. Фрэнкс Л. Теория сигналов. — М.: Сов. радио, 1974. — 344с.
4. O.M. Soloveyko , Musatenko Yu.S., V.N. Kurashov, V. A. Dubikovskiy Video compression based on enhanced EZW scheme and Karhunen-Loeve transform. — SPIE Proceedings. — v.4041. 2000. — pp.180-191.

СИСТЕМА СТИСКУ ДАКТИЛОСКОПІЧНИХ ЗОБРАЖЕНЬ НА ОСНОВІ ВЕЙВЛЕТНОГО ПЕРЕТВОРЕННЯ

Олег Капшій[1], *Богдан Русин*[2]

[1]ФМІ НАН України, Львів 79601, вул.Наукова 5а, т.(0322)65-45-30, dep32@ipm.lviv.ua

[2]ФМІ НАН України, Львів 79601, вул.Наукова 5а, т./ф.(0322)63-41-09, rusyn@ipm.lviv.ua

Анотація В даній статті розглянуто метод втратного стиску зображень відбитків пальців людини, що базується на вейвлет перетворені зображення. Описано структуру пристроїв стиску та реконструкції зображення, структуру даних, які потрібно зберігати. Представлено досягнуті результати та проаналізовано шляхи підвищення рівня стиску за рахунок покращення кодера, врахування структури даних, які стискаються. В системі стиску використовується перечислювальний кодер.

Ключові слова: відбитки пальців, стиск, вейвлет перетворення, безвтратне кодування.

ВСТУП

Відбитки пальців кожної людини є унікальними, тому вони дозволяють однозначно ідентифікувати особистість. Але при великій кількості людей, масиви дактилоскопічних зображень також стають дуже великими, а їхня ручна обробка ускладненою. Тому у всіх дактилоскопічних експертних підрозділах вводяться системи автоматичного розпізнавання відбитків. Всі системи такого типу використовують для порівняння не самі зображення, а певні набори ознак отриманих після обробки цих зображень. Кількість і розміри ознак є відносно невеликими, що дає можливість проводити пошук по дактилоскопічних базах у відносно короткі терміни. Але, враховуючи те, що такі системи не дають 100% імовірності розпізнавання, вони використовуються як системи підтримки прийняття рішень, які значно полегшують роботу із великими масивами даних. Такі системи вимагають операторів, тобто необхідності відображення зображень відбитків. З другого боку, кожна із таких систем містить лише набір відбитків певної кількості людей, при цьому, кожна з них має свої власні набори ознак, за якими проводиться ідентифікація, тому виникає задача передачі зображень відбитків між системами. Якщо зберігати повні зображення відбитків в базі, це буде займати величезні об'єми пам'яті. Так, для бази дактилокарт на 100 тис. осіб, при розмірах зображень 512x512 точок, 256 рівнях градації яскравості, зображення будуть займати близько 260 Гб простору. Тому постає задача стиску зображень. Безвтратний стиск дозволяє зменшити необхідний об'єм максимум в 2 рази. Цього є недостатньо, тому використовують втратні методи компресії [2].

1. СТРУКТУРА СИСТЕМИ

Найбільш вживана структурна схема втратної компресії представлена на рис.1. Схема містить блок перетворення зображення, який вносить втрати, (блок 1) та безвтратний кодер (блок 2). Відповідно пристрій реконструкції містить декодер (блок 3) та пристрій реконструкції зображення по отриманих даних (блок 4).

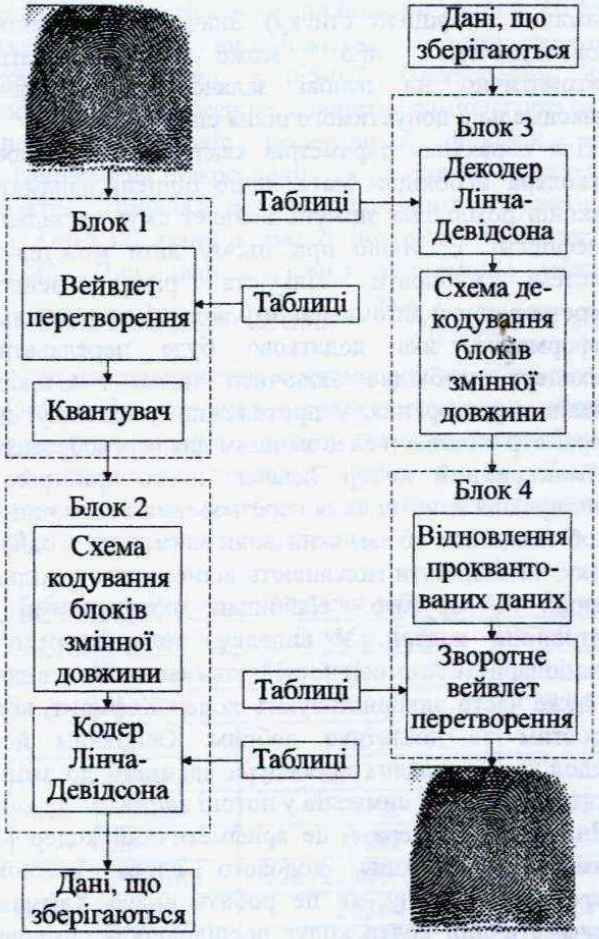


Рис.1. Структурна схема системи компресії

Робота блоку 1 полягає у перетворенні зображення у дані, спотворення яких менше впливає на саме зображення, аніж спотворення самого зображення напряму. При цьому такі дані після перетворення і контрольованого спотворення мають мати надлишковість, яка пізніше може бути усунена в блоці 2. Дуже популярним перетворювачем такого виду є пара вейвлет перетворювач – квантувач.

Вейвлет перетворення [1] переводить зображення у набір слабокорельованих частотно-просторових смуг, які дозволяють однозначно відновити зображення. У нашому алгоритмі використовується біортогональна пара фільтрів. Для роботи перетворювача як при стиску, так і при відновленні, відбувається синтез фільтрів на основі їхніх збережених імпульсних характеристик. Ці фільтри є незмінними при роботі зі всіма зображеннями.

Квантувач – це пристрій, який вносить втрати у дані за рахунок відображення неперервної множини можливих значень на обмежену кількість підінтервалів. У роботі використано нерівномірний квантувач із мертвою зоною. Параметри квантувача розраховуються для кожної смуги вейвлет перетворення окремо, на основі параметрів законів розподілів даних смуг після вейвлет перетворення та на основі бажаної швидкості створення повідомлення r (фактично цим визначається бажаний коефіцієнт стиску). Значення r задається користувачем, або може обраховуватися автоматично на основі заданого користувачем максимально допустимого рівня спотворень.

Для обрахунку параметрів квантувача на стороні декодера необхідно мати лише оцінені параметри законів розподілів значень вейвлет смуг та заданий коефіцієнт r. Якщо при цьому дати можливість системі вибирати кількість рівнів вейвлет перетворення в автоматичному режимі, то в загальну інформацію, яка додатково буде передаватися декодеру необхідно включити також і кількість рівнів перетворення, у протилежному випадку, цей параметр залишається незмінним для всіх зображень.

Безвтратний кодер Задачею цього пристрою є маніпуляція даними та їх перетворення таким чином, щоб зменшити об'єм, який вони займають, з одного боку, та залишити можливість відновлення вихідних даних, з другого. Найбільш популярними є ентропійні кодери. У випадку, коли джерело є стаціонарним та імовірності появи символів є відомі – дуже часто використовують кодер Хафмана, він є простим та достатньо добрим. Основним його недоліком є складна процедура адаптації до змін у статистиці появи символів у потоці даних.

Інший вид кодера – це арифметичний кодер [4]. Замість присвоєння кодового слова кожному окремому символу, як це робить кодер Хафмана, арифметичний кодер кодує послідовності символів. Він оперує з кумулятивною складеною імовірністю, обрахованою по послідовності символів, і виводить її з точністю, достатньою для відновлення послідовності. Цей кодер може досягнути теоретичної границі Шенона при компресії даних. Основними недоліками методу є його обчислювальна складність та складність реалізації, а також те, що метод на даний момент закритий для використання патентами.

Зів-Лемпел кодування. Це дуже популярна методика, яка має велику кількість варіантів. Дана методика при компресії використовує динамічні словники, які формуються на основі вхідних даних. Цим досягається висока степінь адаптації кодера до потоку вхідних даних. Серед кодерів цього сімейства є LZ77, LZ78, LZW та багато їхніх модифікацій. Кодери цього типу можуть бути дуже ефективними для кодування довгих стаціонарних сигналів з невідомою статистикою, але вони мають обмежену ефективність для нестаціонарних сигналів.

Перечислювальний кодер (enumerative coder) [3] В алгоритмі стиску зображень відбитків пальців використано кодер, який робить перенумерацію стрічок. Базовий принцип роботи кодера полягає у наступному: Нехай дано стрічку символів довжиною n, яка містить рівно w одиниць, а решта символів - нулі. Нехай w є відомим, тоді існує $C(n,w)=\frac{n!}{w!\cdot(n-w)!}$ таких стрічок і номер цієї стрічки у впорядкованому наборі таких стрічок унікально її визначає. При цьому номер стрічки можна передавати за допомогою лише $\log_2(C(n,w))$ біт. Існує багато швидких алгоритмів для генерації і декодування номера, особливо для малих n.

На даному етапі розробки методу компресії використовується схема кодування Лінча-Девідсона. Алгоритм роботи кодера виглядає наступним чином – побітно проглядається вхідна стрічка, що містить n символів, з яких w - одиниці. Якщо текучий символ рівний одиниці – до лічильника додається число стрічок, які мають такі самі символи у всіх попередніх позиціях, що й стрічка, яка проглядається, і нуль в текучій позиції.

Декодер, знаючи n та w, зчитує номер запису, який має довжину $\log_2(C(n,w))$, після чого працює за наступним алгоритмом: присвоює кожному наступному символу вихідної послідовності 0, якщо текучий номер стрічки є меншим за $C(n-i-1,left\_1s)$, де i - текуче положення вихідної стрічки, $left\_1s$ - кількість одиниць, які ще не було записано у вихідний потік. Якщо текучий номер стрічки не є меншим за $C(n-i-1,left\_1s)$, тоді у вихідний потік записується 1, зменшується текучий номер на $C(n-i-1,left\_1s)$.

Але для такого кодера необхідно для кожної послідовності з n символів передавати при кодуванні у вихідний потік кількість одиниць, присутніх в блоці. Для усунення цієї проблеми, вхідні для кодування дані попередньо обробляються за допомогою схеми кодування блоків змінної довжини у блоки однакової довжини. Принцип роботи кодера такої схеми полягає в наступному: задається n та w. При цьому кількість нульових символів у вхідних стрічках буде рівною $z=n-w$. Вхідний потік побітно зчитується і передається на вихід до тих пір, поки кількість переданих одиниць чи нулів не стане рівною w чи z відповідно. Послідовність на виході доповнюється нулями у першому випадку, чи одиницями у другому, щоб її

довжина була рівною n. Таким чином, у всіх стрічок на виході буде довжину n та кількість одиниць w. Тобто на сторону декодера необхідно лише раз передати n та w. Далі використовується алгоритм Лінча-Девідсона, що був описаний вище.

Декодер для схеми кодування блоків змінної довжини працює наступним чином: з вхідного потоку читають символи до тих пір, поки кількість прочитаних одиниць чи нулів не буде рівна w чи z відповідно. Цю частину стрічки пересилають у вихідний потік, а решту символів із входу, що доповнюють стрічку до довжини n опускають.

2. РЕАЛІЗАЦІЯ СИСТЕМИ КОМПРЕСІЇ

Для реалізації вейвлет перетворення на стороні компресії та на стороні відновлення зображення зберігаються імпульсні характеристики для біортогональних фільтрів. Також, для пришвидшення процедур безвтратного кодування та декодування зберігаються таблиці значень $C(n,w)$, для всіх можливих значень n та w. Вхідне зображення передається на вейвлет перетворювач, після чого для кожної смуги розраховуються параметри закону розподілу значень смуги. В якості апроксимуючого закону розподілу використано узагальнений гаусівський розподіл. Його параметр β передається в якості додаткової інформації на сторону відновлення для кожної смуги. Також в якості додаткової інформації передається дисперсія та розміри кожної смуги. На основі розрахованих параметрів, для кожної смуги розраховується свій квантувач. Після квантувача на вихід блоку 1 передаються не проквантовані значення, а номери рівнів квантування, які відповідають значенням в кожній точці смуги. Це дозволяє в подальшому уникнути операцій з числами з плаваючою комою, що збільшує швидкість системи. Після цього для кожної смуги складається бінарна матриця ненульових елементів, бінарна матриця знаків ненульових елементів, та матриця модулів ненульових елементів. Для кожної з матриць визначається імовірність появи одиниць в бінарному представленні матриці. На основі цієї інформації визначається n та w, виходячи з максимально можливого числа $C(n,w)$, яке можна порахувати без округлень. Кожна з отриманих матриць обробляється за допомогою описаної вище схеми кодування блоків змінної довжини, а результат – за допомогою перечислювального кодера. Отримані дані потоком передаються на систему збереження. Також, в якості додаткової інформації, передається кількість одиниць у бінарному представленні матриці та кількість блоків даних отриманих для кожної матриці після схеми кодування блоків змінної довжини. Для зменшення кількості біт, які необхідно кодувати для матриць модулів ненульових елементів, для кожної такої матриці визначається максимальне значення модуля і, відповідно, кількість біт, яку вимагається для кодування такого числа. Кількість біт передається в якості додаткової інформації.

На стороні відновлення на основі інформації про кількість одиниць у кожній бінарній матриці визначається n та w. Після цього дані проходять через декодер перечислювального кодера та декодер схеми кодування блоків змінної довжини (блок 3). При цьому використовуються передані значення кількості блоків даних для кожної матриці. В результаті отримують відновлені номери рівнів квантування для кожної точки всіх смуг.

Знаючи параметри законів розподілу значень кожної смуги формуються квантувачі для кожної смуги, що дозволяє перейти від номерів рівнів квантування до дійсних значень. Ці дані передаються на блок зворотного вейвлет перетворення, на виході якого отримують реконструйоване зображення.

Експериментальні дослідження системи Перевірка візуальної якості відновленого зображення для різних швидкостей створення повідомлення r (тобто кількості біт, які витрачаються на один символ) показало, що при r=0.5-0.7, спотворення, які вносяться в зображення, повністю компенсуються з допомогою методів попередньої обробки, що розраховані на підкреслення на зображені структури відбитка. Приклад реконструйованого зображення при r=0.5 показано на рис.2г (вихідне зображення - рис.2б). При значені r=0.5 теоретично можна досягнути стиску зображення в $8bps/0.5bps = 16$ раз.

Результати дослідження системи компресії подано у таблиці 1. Для порівняння приведені розміри наборів даних отриманих після стиску даних на виході блоку 1 за допомогою програм WinRar та zip. Також подано значення пікового співвідношення сигнал – шум для відновлених зображень

Таблиця 1. Результати стиску зображень (r=0.5)

Рисунок	Компресор	Розмір, байт	PSNR, дБ
Рис.2а	Власний	556+34061=34617	28.04
	WinRar	29258	
	Zip	31546	
Рис.2б	Власний	556+24131=24687	28.93
	WinRar	29369	
	Zip	32882	
Рис.2в	Власний	556+22346=22902	35.22
	WinRar	26364	
	Zip	28967	

Цікавим може бути використання результату вейвлет перетворення, а саме найбільш низькочастотної смуги. За рахунок зменшення роздільної здатності зображення, відображення папілярних ліній згладжуються і на місці власне відбитка залишається практично однорідна пляма, яка по інтенсивності відрізняється від навколишнього фону. Це можна використати для відділення власне відбитка від фону. Можна навіть використати звичайну операцію порогового

розділення, обробивши після цього отримане бінарне зображення медіанним фільтром.

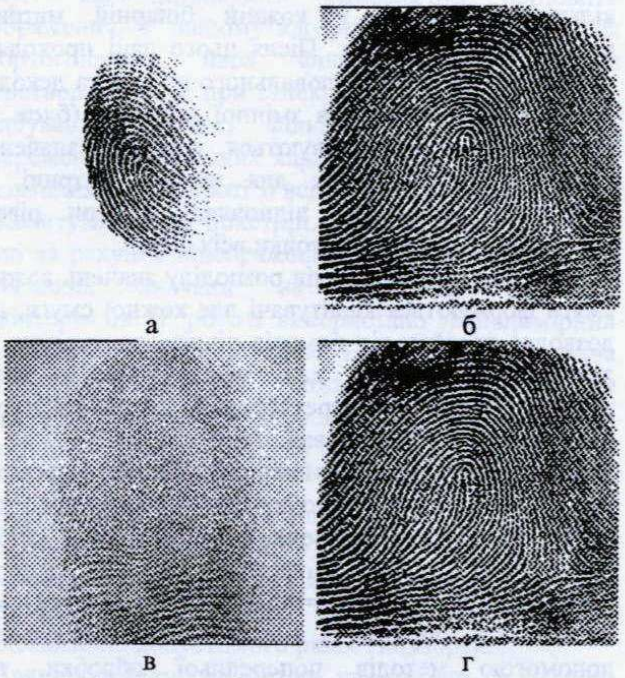


Рис.2 Приклади зображень відбитків (а,б,в) та відновленого після компресії зображення (г) при r =0.5

Шляхи покращення рівня компресії Якщо розглянути гістограми розподілу номерів проквантованих рівнів в кожній смузі, то типова гістограма буде мати вигляд показаний на рис.4. Ця гістограма відповідає припущенню про те, що закон розподілу значень в межах смуги відповідає узагальненому гаусівському розподілу. Але всі номери рівнів при передачі на схему кодування

Рис.3 Зображення вейвлет смуг для відбитка рис.2в

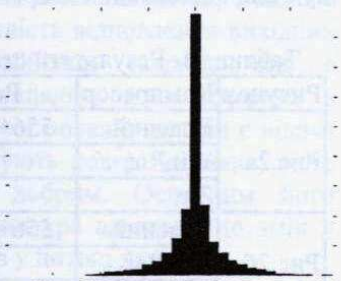


Рис.4 Типова гістограми розподілу значень вейвлетних смуг

блоків змінної довжини кодуються однаковою кількістю біт. Враховуючи те, що менші номери зустрічаються частіше, їх можна закодувати за допомогою меншої кількості біт, а ніж більші номери. Це добавить ще один блок на початку безвтратної частини, але дозволить збільшити рівень компресії. З другого боку, бінарне представлення номерів рівнів є різним по кількості та рівномірності розміщення одиниць (для прикладу 7 в бінарному представлені це 0111, а 8 – 1000). Тут виникає ще один шлях покращення кодування, а саме – кожному символу у відповідність ставиться нове бінарне представлення так, щоб в межах смуги всі символи мали приблизно однакову рівномірно розміщену кількість одиниць у бінарному представленні. В цьому випадку покращиться робота схеми кодування блоків змінної довжини. Ще один шлях зменшення розміру необхідного об'єму полягає у використанні раніше згаданої особливості вейвлетного перетворення, яке дозволяє робити сегментацію власне відбитка. Тоді фон можна заміняти довільними даними, які роблять зображення в цілому більш рівномірним, або взагалі відкидати фон і кодувати тільки область відбитка.

Також можна отримати зменшення більше ніж у два рази і для додаткових даних. Перш за все можна усунути інформацію про кількість блоків даних після схеми кодування блоків змінної довжини за рахунок додавання додаткових блоків обчислення. Оскільки розміри окремих смуг вейвлетного перетворення є відомими, то при декодуванні можна спочатку зчитати матрицю ненульових елементів, яка має розмір рівний кількості елементів смуги. З неї можна визначити розміри матриць від'ємних елементів та модулів елементів, вони рівні кількості одиниць в матриці ненульових елементів. Також на даному етапі розробки зберігається кількість одиниць для кожної матриці в бінарному представлені. Але особливістю процедури обрахунку параметрів квантувача є те, що паралельно обраховуються теоретичні імовірності попадання символів у інтервали квантувача. Знаючи ці імовірності та бітове представлення символів можна обрахувати і кількість одиниць у бінарних представленнях матриць.

ВИСНОВКИ

Розроблено систему стиску зображень відбитків пальців, забезпечує стиск зображень розміром 512x512 приблизно в 9.6 раза при збереженні якості зображення відбитка в межах, необхідних для збереження можливості подальшої роботи із відбитком. Показано шляхи покращення рівня стиску на основі аналізу гістограми розподілу значень проквантованих вейвлетних смуг. В подальшому планується досягнути ще більшого рівня стиску, за рахунок моделювання структури відбитка.

ЛІТЕРАТУРА

1.Antonini M., Barlaud M., Mathieu P., and Daubechies I. *Image coding using wavelet transform* // IEEE Trans. Im. Proc., vol. 1(2), pp. 205- 220, April 1992;

2.Brislawn Ch., Bradley J., Onyshczak R. and Hopper T. *The FBI compression standard for digitized fingerprint images* //Proc. SPIE, vol.2847, Denver, Aug. 1996;

3. Oktem L. *Hierarchical enumerative coding and its application in image compression* //Ph.D. dissertation, Tampere University, Finland, Oct. 1999;

4. Howard P. and Vitter J. *Practical implementations of arithmetic coding* //Image and Text Compression, James A. Storer, ed., Kluwer Academic Publishers, Norwell, MA, 1992, pages 85--112.

ПІДВИЩЕННЯ ЯКОСТІ ВІДНОВЛЕННЯ ЗОБРАЖЕННЯ ІМПУЛЬСНИХ ПРОТЯЖНИХ ОБ'ЄКТІВ

О.В.Антонов, Ю.М.Герасимов, М.В.Руженцев
Національна Академія Наук України
Радіоастрономічний інститут
61002, Харків, МСП, Червонопрапорна, 4
Тел.: 44-83-51, Fax: (380)-572-476506, E-mail: ruzh@rian.kharkov.ua

У роботі розглядається якість відновлення зображення імпульсного протяжного об'єкта нелінійним ітераційним методом з обмеженнями - обліком апріорної інформації про ненегативність рішення і просторовий розмір об'єкта. Показано, що введення обмежень приводить до ефекту надвисокої роздільної здатності при відновленні границь об'єкта. Однак деталі його поверхні відновлюються недостатньо чітко. Крім того, у відновленому зображенні поблизу границь об'єкта виникають паразитні коливання. Розглянуто причини появи зазначених недоліків і запропонований алгоритм їхнього усунення. Попереднє виключення з відновлюваного зображення основної - детермінованої компоненти зображення дозволяє усунути відзначені недоліки. Інформація про основну компоненту зображення об'єкта встановлюється по його реальному запису. Математичне моделювання підтверджує значне підвищення якості відновлення зображення запропонованим алгоритмом.

1. ВСТУП

У різних областях науки та техніки, таких як астрономія, дистанційне зондування і т.п., предметом досліджень є зображення об'єктів. Викликано це тим, що детальна структура зображення даних об'єктів дозволяє виділяти важливу інформацію про їхні властивості, механізми утворення й еволюції. Повнота отриманої інформації у першу чергу залежить від якості реєстрації об'єктів, що спостерігаються, тобто від роздільної здатності, яку має прилад, що здійснює реєстрацію. У такий спосіб роздільна здатність, що її має прилад для реєстрації, є чинником, здатним істотно обмежувати можливості різних досліджень і експериментів. Підвищення якості реєстрації зображень можна забезпечити різними методами відновлення зображення. Існують різні лінійні та нелінійні методи відновлення зображення [1]. Лінійні методи, засновані на корекції зображення за допомогою частотного фільтра, характеристика якого зворотна до апаратної функції приладу, що реєструє, дозволяють поліпшити якість зображення. Однак у цьому випадку відновлене зображення не має підвищеної роздільної здатності, що визначається шириною апаратної функції приладу, в межах якої виконується відновлення.

Інформація про структуру зображення, що загублена у вихідному зображенні приладу, що реєструє, може бути відновлена, якщо в процесі відновлення зображення врахувати апріорну інформацію про досліджуваний об'єкт. Подібна інформація є граничною, тому методи відновлення зображення, які враховують апріорну інформацію, є нелінійними.

Облік обмежень приводить до підвищення просторових градієнтів у відновленому зображенні, тобто до розширення спектра просторових частот зображення за межі смуги частот, які пропускаються апаратною функцією приладу, що еквівалентно одержанню ефекту надвисокої роздільної здатності. Фактично облік апріорної інформації в процесі відновлення приводить до генерації спектра просторових частот, що був подавлений апаратною функцією приладу.

Очевидно, чим більше апріорної інформації про об'єкт вивчення можна врахувати в процесі відновлення зображення, тим вище ефект надвисокої роздільної здатності і тим з більшою точністю відновлюється зображення.

Різні нелінійні методи, що враховують однотипну апріорну інформацію, досить еквівалентні між собою за одержуваними результатами і розрізняються лише більш-менш складним способом реалізації.

Найбільш простим з погляду обліку апріорної інформації способом реалізації є ітераційний нелінійний метод відновлення, у якому проводиться послідовна покрокова корекція відновлюваного зображення. При цьому з'являється можливість контролювати процес відновлення, визначати якість відновлення в залежності від числа ітерацій. Аналіз ефективності відновлення зображення нелінійним ітераційним методом

показує, що вона залежить від типу об'єкта. Найбільш складним є відновлення зображення імпульсних протяжних об'єктів. Типовим представником такого об'єкта може служити диск Сонця, на поверхні якого розташовані активні утворення з підвищеною яскравістю випромінювання. Відновлення такого об'єкта ітераційним методом з урахуванням інформації про його розмір приводить до появи на його поверхні паразитних коливань.

Задачею даної роботи є підвищення якості відновлення зображення шляхом розробки алгоритму застосування ітераційного методу, що усуває паразитні коливання.

2. ГОЛОВНА ЧАСТИНА

Розглянемо результати застосування нелінійного ітераційного методу відновлення зображення до задачі відновлення імпульсного протяжного об'єкта.

Ітераційний метод відновлення є розвитком лінійного методу інверсної фільтрації, у якому зображення відновлюється шляхом корекції його спектра просторових частот за допомогою оператора зворотного виду апаратної функції приладу, що реєструє. Розкладання зазначеного оператора в ряд, що сходиться, дозволяє перевести разову процедуру інверсної фільтрації в послідовний ітераційний процес покрокової корекції, що сходиться, спектра просторових частот відновлюваного зображення. За допомогою зворотного Фур'є перетворення даний процес ітераційного відновлення можна перевести в просторову область. Перехід від інверсної фільтрації до ітераційного процесу є лінійним і в такому виді обидва методи є еквівалентними по якості відновлення зображення. Однак на відміну від інверсної фільтрації в ітераційному методі з'являється можливість на кожному кроці відновлення одержуваний результат порівнювати з апріорною інформацією як по структурі зображення, так і по його спектру. Результат порівняння враховується на наступному кроці відновлення. У підсумку процес відновлення виявляється нелінійним і відновлюване зображення направляється в клас об'єктів, заданих апріорною інформацією.

У просторовій області ітераційний алгоритм відновлення зображення має наступний вигляд [1]:

$$T_я^{k+1}(\vec{r}) = \left\{ T_я^k(\vec{r}) + \lambda \left[T_a(\vec{r}) - T_я^k(\vec{r}) * A(\vec{r}) \right] \right\} F,$$

де $\vec{r}$ - вектор координат елементів поверхні об'єкта, $T_a(\vec{r})$ - яскравісна температура зображення на виході приладу, що реєструє, з нормованою просторовою характеристикою - діаграмою напрямленості (ДН) $A(\vec{r})$, $T_я^k(\vec{r})$, - яскравісна температура зображення після k - ої ітерації, λ - коефіцієнт швидкості збіжності процесу відновлення, * - знак згортки функцій, F - оператор обмежень, що враховує апріорну інформацію в просторовій області. Оператор F приймає значення 0 або 1 у залежності від відповідності обчислених значень $T_я^{k+1}(\vec{r})$ апріорним умовам.

У даному алгоритмі швидкість відновлення регулюється коефіцієнтом λ. Його максимальне значення обмежується величиною, перевищення якої приводить до процесу відновлення, що розходиться. Ця величина зв'язана з коефіцієнтом передачі апаратної функції приладу, що реєструє. Якщо апаратна функція нормована, то максимальне значення коефіцієнта λ дорівнює двом.

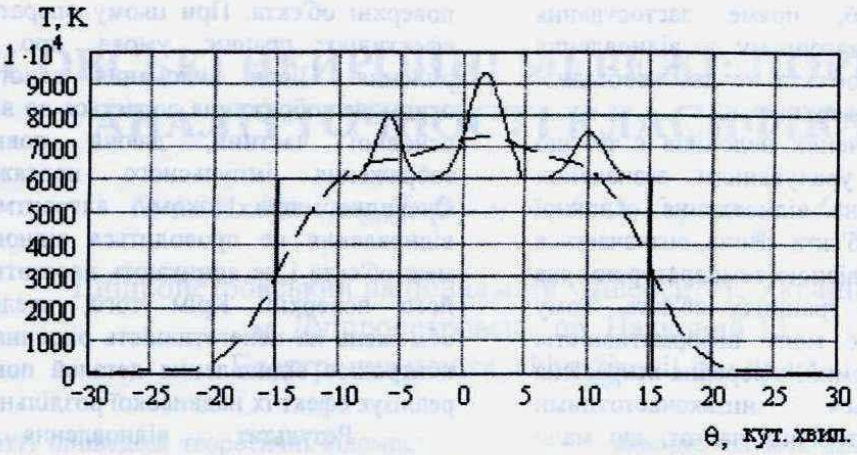
Найбільш очевидною апріорною інформацією є умова ненегативності відновленого зображення. Її застосування дозволяє значно поліпшити якість відновленого зображення, забезпечити виділення згладжених деталей.

У випадку відомого розміру відновлюваного об'єкта процедура відновлення його зображення обмежується в заданих розмірах, що також призводить до підвищення якості зображення. В ітераційному методі відновлення оператор F може відбивати облік і будь-якого іншого виду апріорної інформації.

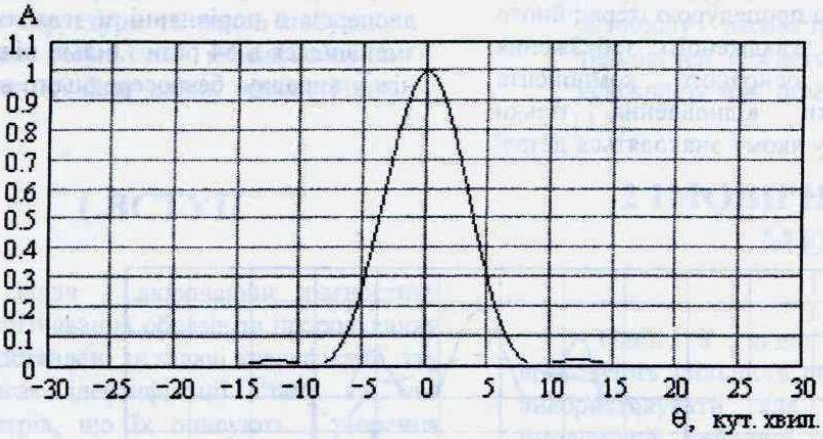
Пряме застосування нелінійного ітераційного методу з різними операторами обмежень, що контролюють покрокову процедуру відновлення, дозволяє успішно відновлювати широкий клас об'єктів. Однак при відновленні імпульсних протяжних об'єктів якість відновлення виявляється недостатньо високою. При будь-яких операторах обмежень у відновленому зображенні даних об'єктів спостерігаються паразитні коливання, що погіршують якість зображення. Крім того невисоким виявляється якість відновлення деталей поверхні об'єкта.

Виявити природу недоліків прямого застосування нелінійного ітераційного методу до відновлення зображення імпульсних протяжних об'єктів і визначити шляхи їх усунення дозволяє аналіз результатів математичного моделювання.

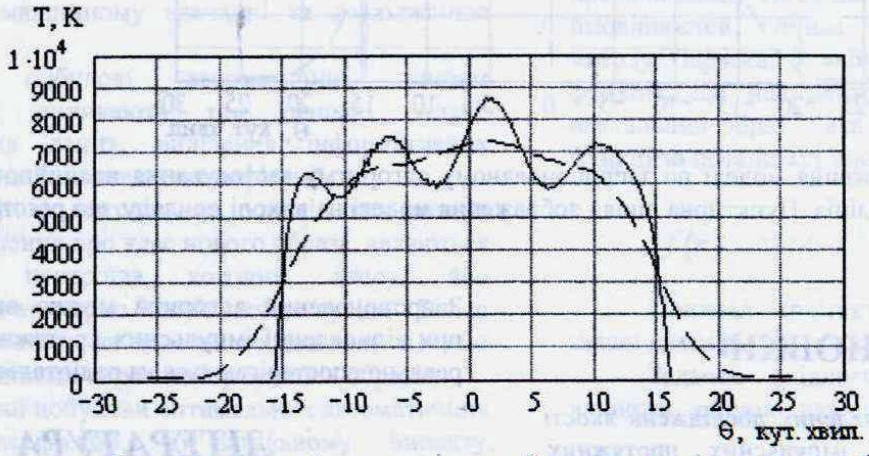
Як об'єкт моделювання виберемо зображення Сонця з кутовим розміром 32' і яскравісною температурою $6 \cdot 10^3$ К, на поверхні якого розміщені три яскравих активних утворення гаусового виду з кутовим розміром порядку 3' на рівні половинної потужності (мал. 1). Як діаграма напрямленості приладу, що реєструє - радіотелескопа використовується розрахункова ДН параболоїда з шириною на рівні половинної потужності, рівної 7' (мал. 2). Оскільки ширина ДН радіотелескопа істотно перевищує кутовий розмір деталей на поверхні моделі імпульсного протяжного об'єкта, то вид їх згортки - зображення на виході



Мал. 1. Модель імпульсного протяжного об'єкта в кутових координатах θ - суцільна лінія. Пунктирна лінія - зображення моделі на виході приладу, що реєструє.



Мал. 2. Нормована діаграма напрямленості приладу, що реєструє, у кутових θ координатах .



Мал. 3 Відновлене зображення моделі з оператором обмежень по відомому розміру об'єкта - суцільна лінія. Пунктирна лінія - зображення моделі на виході приладу, що реєструє.

радіотелескопа виявляється істотно згладженим (мал. 1).

Відновимо отримане згладжене зображення моделі об'єкта розглянутим ітераційним алгоритмом, у якому оператор обмежень контролює відомий заданий розмір об'єкта. Результат відновлення після 1000 ітерацій при максимальній збіжності процесу, що відповідає коефіцієнту $\lambda = 2$, приведений на мал. 3.

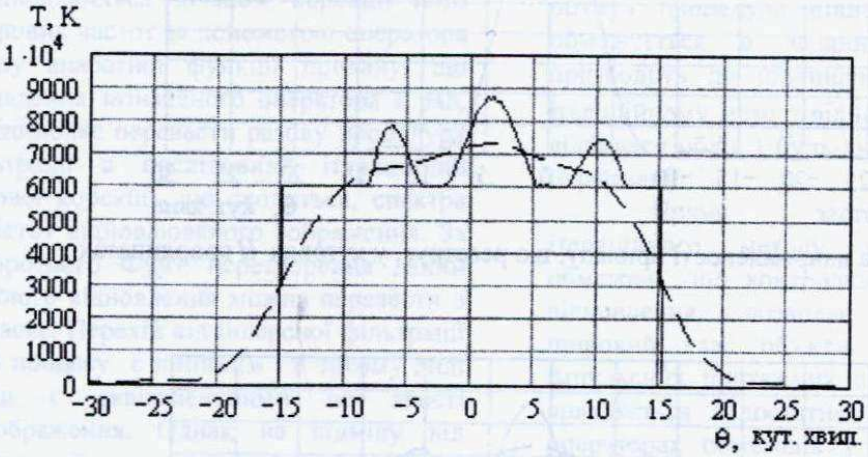
Оцінимо якість відновлення зображення моделі об'єкта як відношення дисперсій вихідного згладженого об'єкта щодо моделі об'єкта і відновленого зображення відносно моделі об'єкта. У даному випадку зменшення дисперсії відновленого зображення склало 4,6 рази. Аналогічні результати моделювання отримані при іншому операторі обмежень, що враховує ненегативність відновленого зображення.

У такий спосіб, пряме застосування нелінійного ітераційного алгоритму до відновлення імпульсного протяжного об'єкта не дає необхідної якості відновлення його поверхні.

Причиною відзначених недоліків є те, що процес відновлення з урахуванням зазначених обмежень орієнтований на відновлення основної енергетичної складової об'єкта. Вона визначається високою середньою яскравісною температурою, яка сконцентрована в різких границях об'єкта. Тому отриманий результат має малу інформативність. Розмір об'єкта апріорі відомий, а середня яскравісна температура описується низькочастотними складовими спектра просторових частот, що мало спотворюються апаратною функцією. Даний параметр об'єкта можна з гарною точністю оцінити по зображенню на виході приладу.

Це дозволяє перед процедурою ітераційного відновлення відняти зі згладженого зображення згортку встановленого основного компонента зображення і провести відновлення тільки додаткового компонента, у якому знаходяться деталі поверхні об'єкта. При цьому оператором обмежень ефективно працює умова про ненегативність рішення. Після виконання даного відновлення отримане зображення додається до виділеної раніше основної частини, даючи повне відновлене зображення імпульсного протяжного об'єкта. Очевидно, при такому алгоритмі ітераційного відновлення не проводиться відновлення відомих меж об'єкта і не виникають паразитні коливання на його поверхні. Крім того, введений оператор обмежень на ненегативність рішення безпосередньо контролює відновлення деталей поверхні об'єкта і реалізує ефект їх надвисокої роздільної здатності.

Результат відновлення згладженого зображення моделі по даному алгоритму при $\lambda = 2$ після 1000 ітерацій приведений на мал. 4. Відновлене зображення дуже близьке до вхідної моделі. Його дисперсія в порівнянні зі згладженим зображенням зменшилася в 54 рази і більш ніж у 10 разів менше, ніж у випадку безпосереднього відновлення на мал. 3.



Мал. 4. Відновлене зображення моделі по запропонованому алгоритму застосування нелінійного ітераційного метода - суцільна лінія. Пунктирна лінія - зображення моделі на виході приладу, що реєструє.

3. ВИСНОВКИ

У результаті проведених досліджень якості відновлення зображення імпульсних протяжних об'єктів нелінійним ітераційним методом запропонований алгоритм відновлення, який дозволяє істотно підвищити якість відновлення. Дані результати підтверджені модельними розрахунками.

Запропонований алгоритм можна використовувати при відновленні імпульсних протяжних об'єктів, які реально спостерігаються на радіотелескопі.

ЛІТЕРАТУРА

Василенко Г.И., Тараторин А.М. *Восстановление изображений*. - М.: Радио и связь, 1986.

БАЄСОВСЬКІ НЕЙРОННІ МЕРЕЖІ: ПОРІВНЯЛЬНИЙ АНАЛІЗ ТОЧНОСТІ КЛАСИФІКАЦІЇ

О. М. Ахметшин, Д. ф. - м. н., О.О. Киргизов, Магістр

Дніпропетровський національний університет, Україна, 49050,
м. Дніпропетровськ, пр. Науковий 13.
Електронна пошта: akhm@mail.dsu.dp.ua

В статті приведені теоретичні відомості про Баєсовські та Ймовірнісні нейронні мережі – класифікаційні мережі, засновані на баєсовськім вирішальнім правилі. Представлено структури та метод побудови мереж, а також результати класифікації на прикладі медичної бази даних онкологічних захворювань.

1.ВСТУП

Більшість задач – включаючи діагностику захворювань, розпізнавання образів та пророкування можуть бути представлені як задачі класифікації, так як кожна вимагає ідентифікації стану та має множину параметрів, що їх описують. Створення точних класифікаторів з предкласифікованих даних являється важливим та активно розвиваючимся розділом в машинному навчанні та дослідженню даних.

При побудові автоматичних систем класифікації, виникають три основні задачі: представлення даних, виділення інформативних ознак та побудова системи класифікації.

Одними з методів, на основі яких можна прийняти рішення про клас нового образа, являються обчислення центроїда кожного класу, або використання критерію "найближчого сусіда". Більш "витончені" методи на додаток до розгляду відстані враховують щільність розподілу сусідніх образів.

В основі побудови оптимальних автоматичних систем класифікації, в загальному випадку, використовується баєсовське класифікаційне правило або його різновиди. Цей підхід, заснований на припущенні, що задача вибору рішення сформульована в термінах теорії ймовірностей і відомі всі ймовірні величини, що представляють інтерес. Останнім часом бурхливо розвиваючимся методом є Баєсовські нейронні класифікаційні мережі, що враховують окрім окремого вмісту кожного параметру, що описують образи класифікації, ще й взаємозв'язки між ними.

Представимо дві нейронні мережі, що виконують класифікацію даних за баєсовським правилом [1]:

- Імовірнісні нейронні мережі (ІНМ) – використовують лише умовні спільні щільності розподілу параметрів кожного класу даних.
- Баєсовські нейронні мережі (БНМ) - використовують умовні спільні щільності розподілу параметрів як добуток умовних щільностей кожного з параметрів, враховуючи залежності між ними для кожного класу даних [2,3].

2 ІМОВІРНІСНІ НЕЙРОННІ МЕРЕЖІ

Один з видів нейронних мереж, що враховують щільність розподілу образів, які можна використовувати для класифікації називаються ймовірнісні нейронні мережі (ІНМ). В основу класифікації у мережі ІНМ покладене використання правила Баєса. Це вимагає оцінки функції щільності ймовірностей. Оцінка проводиться за допомогою методу Парзена, у якому використовується вагова функція, що має центр у точці, що представляє навчальний образ. Така вагова функція називається функцією потенціалу або ядром:

$$f(r) = \exp(-r^2) = \exp\left(-\frac{\|x - x_i\|^2}{\sigma^2}\right) \quad (1)$$

Приклад архітектури мережі для рішення задачі класифікації показаний на рис. 1.

Задачею вхідного шару є розподіл даних вхідного образа для шару нейронів. У даному

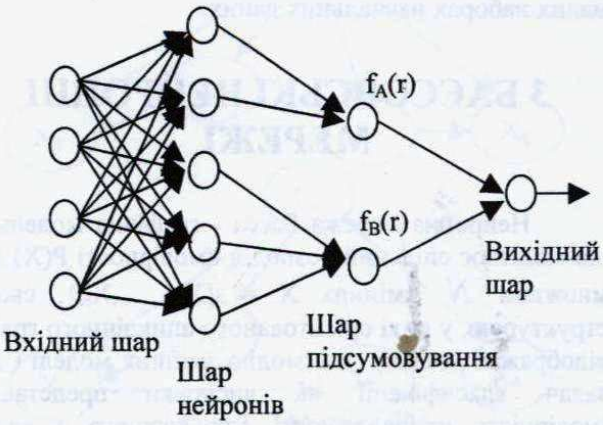


Рис. 1. Приклад архітектури ІНМ

випадку кожен вхідний набір даних має чотири параметри. Шар нейронів має по одному елементу для кожного образа з набору навчальних даних. Вхідний шар і шар нейронів утворять повнозв'язну

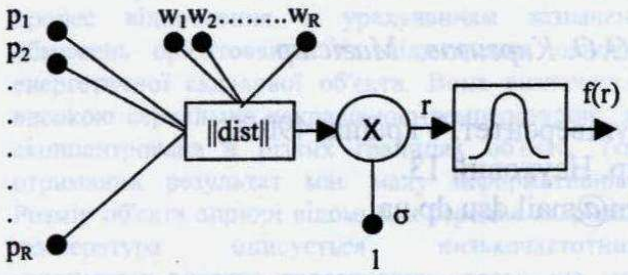


Рис.2. Приклад архітектури радіального нейрона

структуру. Для вхідних в елемент шару нейронів зв'язків вагові значення встановлюються рівними елементам відповідного образа. На рис. 2 представлений радіальний нейрон із входом p і вагами w розмірності R.

Якщо різниця між p і w зменшується, то значення функції збільшується. У такий спосіб нейрон діє як детектор відображаючий ступінь близькості вхідного вектора p до вектора w, зсув σ дозволяє регулювати чутливість нейрона [2].

Шар підсумовування має по одному елементу для кожного класу з навчальної області даних. До будь-якого елемента шару підсумовування йдуть зв'язку тільки від нейронів, що належать відповідному класу. Елемент підсумовування підсумовує вихідні значення нейронів. Ця сума дає оцінку значення функції щільності розподілу ймовірностей для сукупності екземплярів відповідного класу та апріорної ймовірності кожного класу. Вихідний елемент являє собою дискримінатор, що вказує елемент шару підсумовування з максимальні значенням активності. Для мережі ІНМ не потрібно навчання в тім змісті, яке потрібно для мереж із зворотним поширенням помилок, тому що всі параметри мережі ІНМ (число елементів і значення ваг) визначаються безпосередньо навчальними даними.

Мережі ІНМ обмежується задачами класифікації, допускають наявність помилкових даних і забезпечують корисні результати навіть на малих наборах навчальних даних.

3 БАЄСОВСЬКІ НЕЙРОННІ МЕРЕЖІ

Нейронна мережа Баєса - графічна модель S, що моделює спільний розподіл ймовірності P(X) для множини N змінних X = {X_1,.....,X_N}, своєю структурою, у виді орієнтованого ациклічного графа, відображає реальну взаємодію змінних моделі і для задач класифікації як висновки представляє імовірність приналежності для кожного з класів даних [2]. Вершини в S відповідають змінним X, їх також можна представити як мережу зв'язних нейронів. Нехай X_i - означає як змінну, так і відповідну вершину у S, а Pa_i - означає батьків вершини X_i, тобто змінних від який залежить X_i, тоді спільний розподіл ймовірностей для змінних X структури S:

$$P(X) = \prod_{i=1}^{N} P(X_i \mid Pa_i) \quad (2)$$

де $P(X_i \mid Pa_i)$ - умовний розподіл імовірності. Розглянемо довільний орієнтований (ациклічний) граф, рис. 3:

Для даного графа з правила спільного розподілу ймовірностей маємо:

$$P(A,B,D,E) = P(A)P(B)P(D \mid A) \cdot P(E \mid A,B) \quad (3)$$

Імовірність вершини A, враховуючи локальні розподіли ймовірностей, обчислюється:

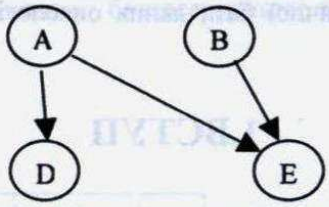


Рис 3 Приклад Баєсовської мережі.

$$P(A \mid B,D,E) = \frac{P(A)P(B)P(D \mid A)P(E \mid A,B)}{\sum_{A'} P(A')P(B)P(D \mid A')P(E \mid A',B)} = \frac{P(A)P(D \mid A)P(E \mid A,B)}{\sum_{A'} P(A')P(D \mid A')P(E \mid A',B)} \quad (4)$$

Області застосування Баєсовських мереж: оцінка щільності розподілу змінних, класифікація, регресія, кластеризація, експертні мережі, визначення ступеня впливу складових моделі і результатів її зміни.

3.1 БАЄСОВСЬКІ КЛАСИФІКАЦІЙНІ МЕРЕЖІ

Баєсовська мережа (БМ) класифікації

B = <N, A, Θ> - спрямований ациклічний граф <N,A>, де кожна вершина n ∈ N являє собою ознаку набору даних, а кожна дуга a ∈ A між вершинами, являє собою імовірнісну залежність. З кожною вершиною n_i зв'язана її умовна щільність розподілу імовірності θ_i (УЩРІ). БМ може використовуватися для обчислення умовної імовірності однієї вершини, з огляду на значення привласнені іншим вершинам, отже, БМ може використовуватися як класифікатор, що дає апостеріорну ймовірність вершини класу (4). Відповідно, класифікаційна система повертає номер класу з найбільшою апостеріорною ймовірністю. Одна з переваг БМ у порівнянні з іншим типам класифікаторів, типу нейронної мережі, - те, що БМ враховують взаємозв'язку серед набору ознак. Застосування БМ для задач класифікації включає дві

підзадачі:

- визначення графічної моделі БМ;
- визначення апріорних ймовірностей та умовних розподілів ймовірностей;
- виконання класифікації;

У задачах, що є оптимальними, тобто для повних наборів даних – для визначення УЩРІ використовує емпірично умовну частоту даних, інакше необхідно робити її відновлення. Дану задачу можна вирішити, скориставшись методом вікон Парзена.

Для визначення структури БМ маються два підходу:

- БМ - структура, що кодує спільний розподіл ознак. Це означає, що найкраща БМ - мережа, яка найкращим способом описує спільне розподілення даних моделі [2].
- БМ - структура, що кодує групу умовних відносин незалежності (УН) серед вершин. Використовуючи статистичний критерій взаємної інформації, можемо знаходити усі УН та використовувати їх як критерій побудови БМ [3].

3.2 ВИДИ КЛАСИФІКАЦІЙНИХ БНМ

Нижче представлено п'ять видів БНМ класифікаторів, які відповідають апріорному представленню взаємодії змінних системи: Проста БМ, Деревоподібна БМ, Розширена БМ, Мульти БМ і Звичайна БМ.

3.2.1 Проста БМ

Проста БМ [2] є структурою, що має вершину класу С як батьківську вершину всіх інших вершин змінні X_1, X_2, X_3, X_4 (Рис. 4). Ніякі інші з'єднання в Простий БМ не розглядаються.

На відміну від нижче перелічених класифікаторів, Проста БМ легка в побудові, оскільки структура дається апріорно і не треба використовувати процедуру побудови мережі.

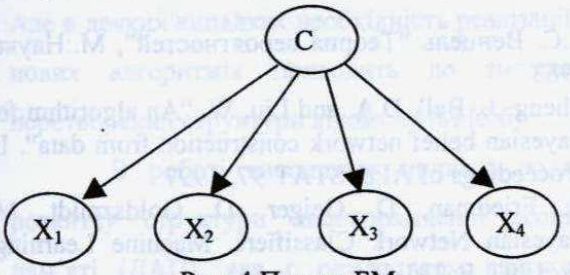


Рис. 4 Проста БМ

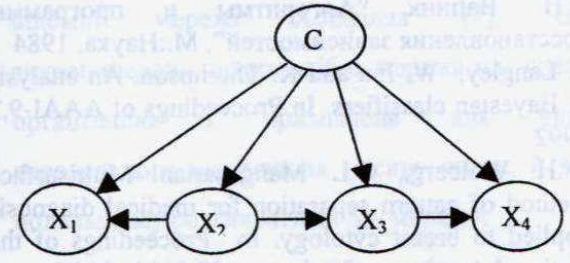


Рис. 5 Древоподібна БМ

Проста БМ припускає, що всі змінні незалежні один від одного, та представляє непогані результати в порівнянні зі складними класифікаторами на великій кількості даних, які не мають сильної кореляції [5].

3.2.2 Деревоподібна БМ

Деревоподібні БМ класифікатори розширюють Прості БМ, в припущені, що змінні мережі формують дерево (Рис.5). Такі мережі є частковим випадком Розширених БМ [2].

3.2.3 Розширена БМ

Розширені БМ класифікатори доповнюють Деревоподібні БМ, дозволяючи змінним мережі формувати довільний ациклічний граф (Рис .6).

Використання Розширених і Деревоподібних БМ менш ефективне в порівнянні з нижче викладеними БМ, оскільки для кожного класу

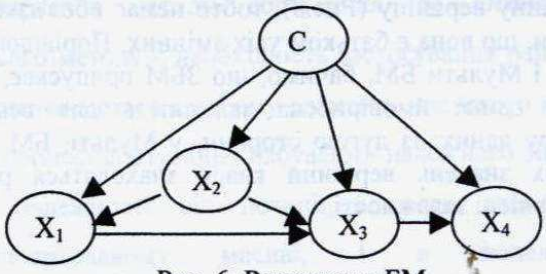


Рис. 6: Розширена БМ

передбачається однаковий вид залежності між змінними і вершиною класу даних.

3.2.4 Мульти БМ

Мульти БМ [3] складається з множини локальних мереж, кожна з який відповідає значенню, що може приймати вершина класу (Рис 7). Мульти БМ може розглядатися як узагальнення Розширених БМ. У Розширених БМ співвідношення серед змінними є однаковим для всіх значень вершини класу; з іншої сторони в Мульти БМ співвідношення серед змінними для різних класів можуть бути різні - тобто, для різних значень вершини класу, змінні

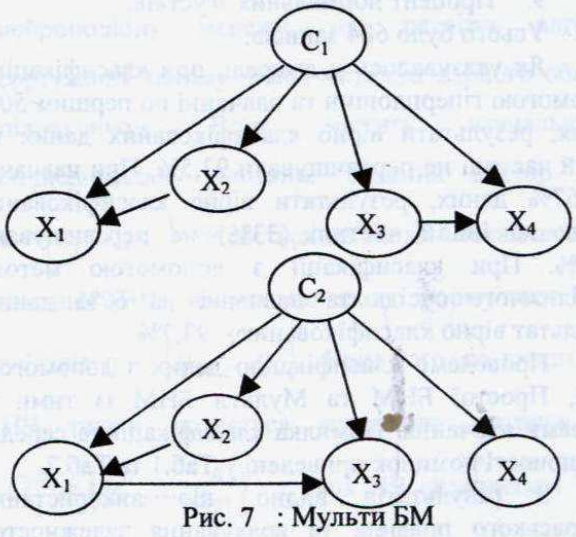


Рис. 7 : Мульти БМ

можуть формувати різні локальні мережі з різною структурою.

3.2.5 Звичайна БМ

Звичайна БМ (ЗБМ) - вид необмеженого БМ

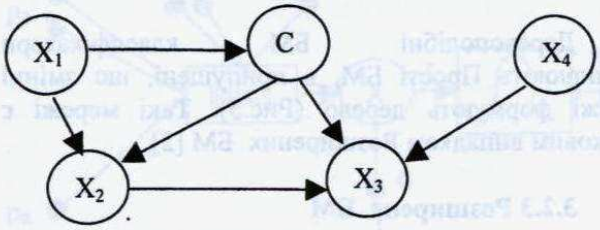


Рис. 8 : Звичайна БМ

класифікатора, що має особливість, не властиву перерахованим вище мережам. Загальна особливість перелічених БМ - те, що вершина класу трактується як спеціальна вершина – вона є батьком усіх змінних. Однак, ЗБМ трактує вершину класу як звичайну вершину (Рис.8), тобто немає обов'язкової умови, що вона є батьком усіх змінних. Порівнюючи ЗБМ і Мульти БМ, бачимо, що ЗБМ припускає, що існує єдина ймовірносна залежність для всього набору даних; із дугою сторони, у Мульти БМ для різних значень вершини класу знаходяться різні імовірнісні залежності.

4 РЕЗУЛЬТАТИ

Експерименти проводились на медичних даних ракових захворювань [6]. Записи описували зображення пункцій пацієнтів та мали наступні параметри:

1. Середня товщина згустків.
2. Однорідність розмірів згустків.
3. Однорідність форм згустків.
4. Густина згустків.
5. Епітеліальний розмірів згустків.
6. Процент згустків без оболонки.
7. Хроматин.
8. Мітоз.
9. Процент нормальних згустків.

Усього було 684 записів.

Як указувалось у джерелі, при класифікації з допомогою гіперплощин та навчанні по першим 50% даних, результати вірно класифікованих даних на іншій частині не перевищували 93,5%. При навчанні на 67% даних, результати вірно класифікованих даних на іншій частині (33%) не перевищували 95,9%. При класифікації з допомогою методу найближчого сусіда та навчанні на 67% даних, результат вірно класифікованих – 93,7%

Проведемо класифікацію даних з допомогою ІНМ, Простої БНМ та Мульти БНМ із тими ж умовами навчання. Помилка класифікації та середні ймовірності помилок приведені у Таб.1 та Таб.2.

З результатів видно, що використання баєсовського правила та врахування залежностей між параметрами приводе до суттєвого підвищення точності класифікації навіть при навчанні на половині набору даних.

Таблиця №1.

Навчання – 50% даних, тестування – інші 50%.

	Проста БНМ:	Мульти БНМ:	ІНМ:
Середня Ймовірність Помилки	0.0091	$2.9*10^{-6}$	0.0388
Помилка Класифікації, %	8.7977	1.7595	16.715

Таблиця №2.

Навчання – 67% даних, тестування – інші 33%.

	Проста БНМ:	Мульти БНМ:	ІНМ:
Середня Ймовірність Помилки	0.0067	$0.3*10^{-7}$	0.0425
Помилка Класифікації, %	7.8221	1.7699	15.929

5 ВИСНОВОК

В статті було продемонстровано новий метод класифікації, заснований на Баєсовській нейронній мережі, з допомогою якого проведено класифікацію даних онкологічних захворювань.

З порівнянь результатів видно, що використання взаємозв'язків між параметрами підвищує точність класифікації даних.

Таким чином, врахування залежностей між параметрами та щільностей їх розподілення, значно підвищує точність класифікації онкологічних захворювань.

ЛІТЕРАТУРА

1. Л.С. Венцель "Теория вероятностей", М.:Наука, 1983
2. Cheng, J., Bell, D.A. and Liu, W. "An algorithm for Bayesian belief network construction from data". In Proceedings of AI & STAT'97, 1997
3. N. Friedman, D. Geiger, D. Goldszmidt, M. Bayesian Network Classifiers. Machine Learning, 29, 1997, P. 131-161.
4. В.Н. Вапник "Алгоритмы и программы восстановления зависимостей", М.:Наука, 1984
5. P. Langley, W. Iba and K. Thompson. An analysis of Bayesian classifiers. In Proceedings of AAAI-9,2 1992
6. W.H. Wolberg,, O.L. Mangasarian. Multisurface method of pattern separation for medical diagnosis applied to breast cytology. In Proceedings of the National Academy of Sciences, 87,9193--9196

Сортувальна нейроподібна мережа

Т. Б. Мартинюк, В. В. Хом'юк, А. В. Кожем'яко, Н. В. Фофанова, О. Б. Мартинюк[1]

Вінницький державний технічний університет, 21021 Вінниця, Хмельницьке шосе, 95

тел. (0432) 325718, факс (0432) 465772, e-mail:psv@vstu.vinnica.ua

[1]*Кримський державний гуманітарний інститут, 98635 Ялта, вул. Севастопольська, 2*

тел/факс (0654) 322114, e-mail:office@cshi.ftrace.net

The mathematical model of the adaptive network for number array sorting using the pair exchange method was proposed. The structure of the sorting neural-like network was considered. The peculiarities of abstract structure of neural-like S-network model were shown.

Вступ

Лавиноподібне зростання кількості публікацій з нейроструктур та нейрообчислень свідчить про перспективність напрямку обчислювальної техніки, що пов'язане з моделюванням нейронних та нейроподібних мереж [1,2]. В багатьох випадках нові розробки стосуються нових сфер використання вже відомих нейроструктур або вдосконалення організації нейронних мереж (НМ) з орієнтацією на новітні елементні базиси, зокрема, на оптоелектроніку [3,4]. Але в деяких випадках необхідність реалізації на НМ нових алгоритмів приводить до змінення або перетворення структури відомих НМ [5,6].

В роботі наводиться приклад подальшого розвитку структури двонапрямленої асоціативної пам'яті (ДАП), яка є результатом вдосконалення відомої мережі Хопфілда [7]. Оскільки запропонована нейроподібна мережа має специфічну організацію і призначена для виконання паралельного сортування масиву чисел, її можна в подальшому розглядати як S – мережу.

Структура адаптивної нейроподібної мережі для алгоритму сортування

В роботі розглянуто реалізацію алгоритму сортування масиву чисел методом парного обміну з підрахунком, який вважається одним зі швидкісних апаратних методів сортування [8]. Особливістю цього методу є необхідність застосування рангів, які призначаються кожному елементу числового масиву, в процесі сортування набувають належного значення в залежності від позиції кожного елемента у відсортованому масиві, і в подальшому використовуються як адреса при послідовному зчитуванні елементів масиву [9]. В процесі досліджень з'ясувалось, що нетрадиційне кодування рангів, а саме, використання одиничного позиційного коду, дозволяє реалізувати просторово-розподілену особливість представлення числової інформації в цьому коді [10] при асоціативному обробленні інформації, до якого можна віднести процедуру сортування [11].

На рис. 1 представлено адаптивну нейроподібну мережу, що реалізує алгоритм сортування масиву чисел методом парного обміну з підрахунком. Вона містить навчальну і обчислювальну частини. Вхідний вектор даних $\mathbf{x} = \{x_1, \ldots, x_n\}$ подається на входи блока навчальної частини (НЧ), на виходах якого формується вектор зв'язків $\mathbf{q} = \{q_1, \ldots, q_n\}$. Крім того, на входи блока НЧ також подається початкова матриця ваги $G^0 = \{g_{11}^0, \ldots, g_{nn}^0\}$, а блок НЧ складається з двох

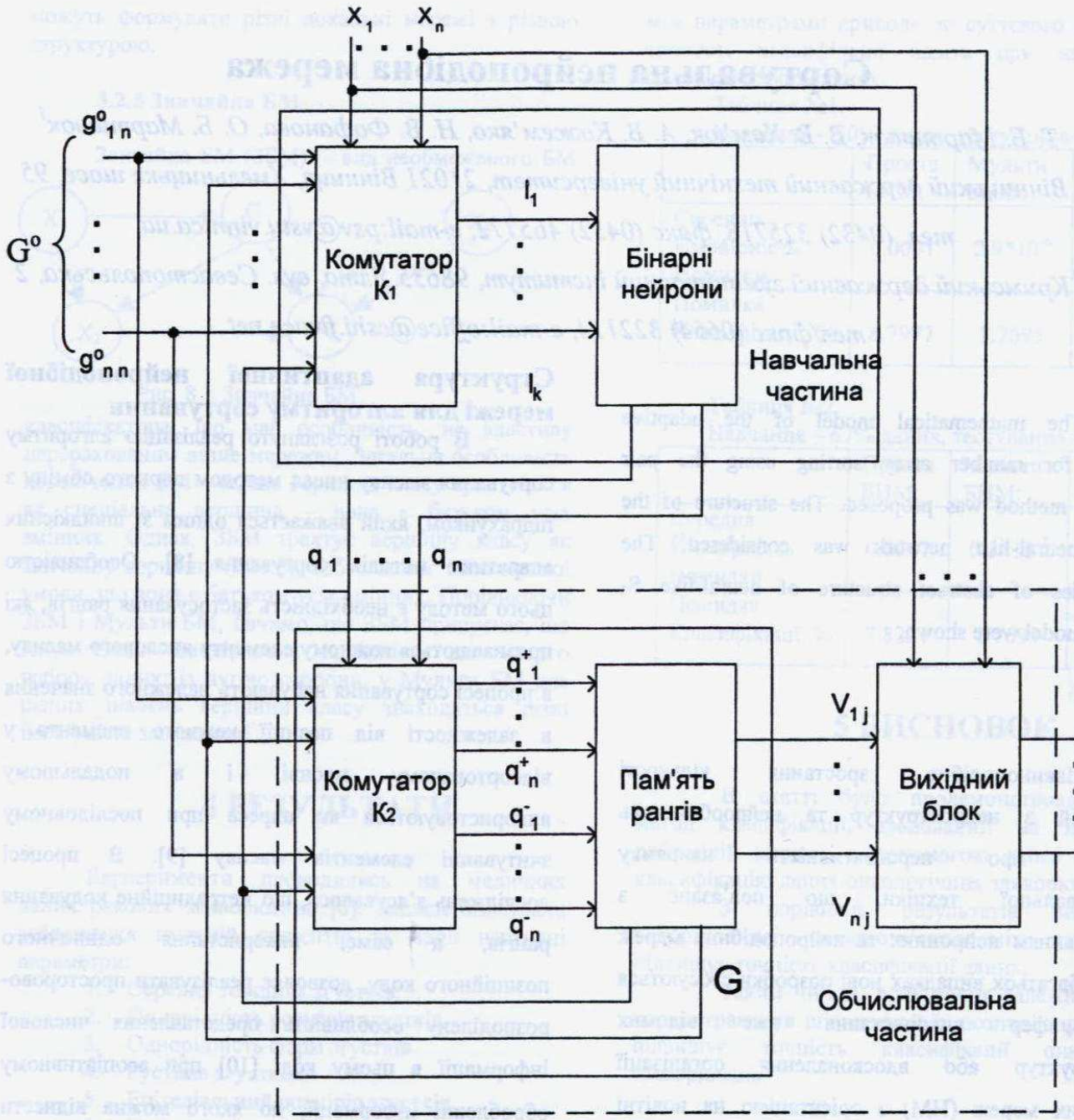


Рис. 1 Структура адаптивної мережі:

$x_1,...,x_n$ - вхідний вектор даних; $\mathbf{G}^0$, $\mathbf{G}$ - початкова і поточна матриці ваги відповідно; $l_1,...,l_k$ - вихідний вектор комутатора К1; $q_1,...,q_n$ - вектор зв'язків; $v_{1j},...,v_{nj}$ - вектор підстановки; $q_1^+,...,q_n^+$, $q_1^-,...,q_n^-$ - вихідні вектори комутатора К2; $x_{1j},...,x_{nj}$ - поточний вихідний вектор даних.

вузлів: комутатора К1 і групи k бінарних нейронів з пороговою функцією вигляду:

$$q_i = f(x_i) = \begin{cases} 1, & якщо\ x_{2k-1} < x_{2k}, \\ 0, & якщо\ x_{2k-1} \geq x_{2k}, \end{cases}$$

де $k = [n/2]$ - найближче більше число.

Комутатор К1 реалізує векторне-матричне множення з формуванням вихідного вектора **l** вигляду [12]:

$$\mathbf{l} = \mathbf{x} \cdot \mathbf{G}.$$

Обчислювальна частина (ОЧ) містить два блоки. На вхід першого блока подаються вектор зв'язків q і початкова матриця ваги G^0. В результаті ітеративного процесу на виходах цього блока формується матриця ваги G, а кінцевим результатом є вектор підстановки $v_j = \{v_{1j},...,v_{nj}\}, j=\overline{1,n}$, який відповідає адресі j – го компонента у відсортованому масиві. Матриця ваги G подається на входи блока НЧ і першого блока ОЧ на всіх ітераціях оброблення, крім першої, оскільки тоді на цих входах фіксується початкова матриця ваги G^0. На входи другого блока ОЧ, який є вихідним блоком, подаються вектори x і v , а на виходах формується послідовність векторів $x_j = \{x_{1j},...,x_{nj}\}$, де на кожному j – му кроці певне числове значення приймає лише один елемент x_i, $i=\overline{1,n}$, з вхідного вектора $x = \{x_1,...,x_n\}$, а всі інші мають нульове значення. Таким чином виконується послідовне зчитування елементів відсортованого масиву (за зростанням або спаданням числових значень елементів).

Перший блок ОЧ складається з двох вузлів: комутатора K_2 і пам'яті рангів. Причому комутатор K_2 реалізує векторне-матричне перемноження і формує два вектори: $q^+ = \{q_1^+,...,q_n^+\}$ і $q^- = \{q_1^-,...,q_n^-\}$ вигляду [12]:

$$q^+_p = q \cdot G^T_p,$$

$$q^-_p = q \cdot G^T_{p+1},$$

де G_p , G_{p+1} – матриці ваги G, які формуються відповідно з непарних (p=1) і парних (p=2) стовпців матриці ваги G; $p \in \{1,2\}$; q^+_p, q^-_p - вихідні вектори, які призводять відповідно до збільшення та зменшення на одиницю рангів у парному або непарному p – му циклі сортування сортування; T – символ транспонування.

Пам'ять рангів має матричну структуру, в якій основною операцією є реверсивний зсув інформації вздовж рядків матриці, а також передбачено як паралельне зчитування поточної матриці ваги G, так і зчитування вектора підстановки v_j у вигляді однорозрядного зрізу всіх рядків (слайзу) [13].

У праці [14] доведена слушність математичної моделі наведеного алгоритму сортування масиву чисел.

Висновки

1. Таким чином, в абстрактній структурі узагальненої моделі сортувальної мережі за аналогією з відомою моделлю [15] можна виділити дві частини: першу – обчислювальну і другу – навчальну. Відмінність полягає в тому, що лише НЧ містить шар нейронів і формує вектор зв'язків q, нульове значення всіх елементів якого припиняє динамічний процес у сортувальній мережі, а в ОЧ замість шару нейронів використовується просторово-розподілена пам'ять рангів, що дозволяє формувати не тільки вектор підстановки v, але й матрицю ваги G для шару нейронів НЧ.

2. Саме специфічність структури ОЧ призводить до того, що запропонована модель сортувальної мережі відрізняється від класичної структури ДАП [7]. Це підтверджує, що складна процедура сортування може бути апаратно реалізована на асоціативному процесорі, основним компонентом якого є асоціативна пам'ять [9,11]. З цієї причини наведена сортувальна мережа може бути визначена як нейроподібна.

3. Розглянута S– мережа досягає стаціонарного стану максимум за (n+1) ітерацій (циклів). Моделювання часових характеристик сортування масиву чисел даним методом показало пряму залежність середнього часу сортування від кількості елементів числового масиву [16]. Таким чином, тільки використання швидкодіючої оптоелектронної

елементної бази може забезпечити опрацьовування великих масивів чисел у реальному часі [13].

Література

1. Галушкин А. И. *Нейрокомпьютеры восьмидесятых (начало очередной революции в области нейрокомпьютеров)*//Зарубежная радиоэлектроника. – 1999. - № 1. – С. 3-16.
2. Hrytsyk V. V., Aizenberg N. N. at el. *The neural and neural-like networks: synthesis, realization, application and future*// Інформаційні технології і системи. – 1998. – Т. 1. - № 1/2. – С. 15 – 55.
3. Егоров В. М. *Трехмерные нейроподобные оптические вычислительные структуры*// Автометрия. – 1993. - № 3. – С. – 38 - 43.
4. Резник А. М., Куссуль М. Э. *Оптоэлектронный нейрокомпьютер*// УСиМ. – 1993. - № 5. – С. 6 – 12.
5. Кисіль Б. В., Стрямець С. П., Опотяк Ю. В. *Моделювання нейронних елементів на однорідному обчислювальному середовищі* // Інформаційні технології і системи. – 1998. – Т. 1. - № 1/2. – С. 110 – 116.
6. Загоруйко Л. В., Тимченко Л. І. *Семантичний підхід до створення просторових нейронних мереж*// Вимірювальна та обчислювальна техніка в технологічних процесах. – 1999. - № 1. – С. 23 – 29.
7. Уосермен Ф. *Нейрокомпьютерная техника: Теория и практика.* Пер. с англ. – М.: Мир, 1992. – 240 с.
8. Лорин Г. *Сортировка и системы сортировки*: Пер. с англ. – М.: Наука, 1983. – 384 с.
9. Мартынюк Т. Б. *Организация ассоциативного процессора с поразрядно-последовательной обработкой информации*// Электронное моделирование. – 1996. – Т. 18. - № 3. – С. 28 - 31.
10. Мартынюк Т. Б., Аль-Хияри М. М., Мартынович Е. Н., Гринчук А. Н. *Организация ассоциативной обработки информации с нетрадиционным кодированием*// Вимірювальна та обчислювальна техніка в технологічних процесах. – 1999. - № 3. – С. 114 – 118.
11. Кохонен Т. *Ассоциативные запоминающие устройства*: Пер. с англ. – М.: Мир, 1982. – 384 с.
12. Мартинюк Т., Кожем'яко В. *Реалізація алгоритму сортування на асоціативному процесорі*// Праці 2-ої Всеукраїнської міжнародної конференції УкрОБРАЗ'94. –Київ, 1994. – С. 235- 236.
13. Мартинюк Т. Б., Лисенко Г. Л., Суприган В. А., Аль-Хіярі М. М. *Процесор сортування чисел на базі оптоелектронних інтегральних схем*//Вісник Національного технічного університету України. Інформатика, управління та обчислювальна техніка. – К.: "Век +", 2000. - № 34. – С. 18 – 25.
14. А. с. 1793438, МКИ G06F7/06. *Устройство для сортировки чисел*/ В. П. Кожемяко, Т. Б. Мартынюк и др. - № 4735756/24: Заявлено 05. 09. 93; Опубл. 07. 02. 93, Бюл. № 5. – 14 с.
15. Григорьев В. Р., Наумов С. П. *Нейросетевая реализация алгоритма сортировки на трехмерном оптическом нейрочипе*// Автометрия. – 1993. - № 3. – с. 28 – 37.
16. Мартинюк Т. Б., Буда А. Г., Козлова В. І., Хом'юк В. В., Мартинюк О. Б. *Часові аспекти сортування великих масивів інформації*// Теорія і практика перебудови економіки: Збірник наукових праць. – Черкаси: ЧІТІ, 2001. – с. 240 – 244.

Fidelity and correlation measures of performance of digital imaging systems

Volodymyr Borovytsky
Optical Engineering Department, National Technical University of Ukraine
Tel. (+380-44) 441-10-22, E-mail: borovytsky@spie.org.ua

У статті досліджуються дві методики оцінювання якості цифрових оптичноелектронних систем з фотоприймачами зображення. Перша методика базується на обчисленні середньо-квадратичного відхилення вихідних сигналів реальної та ідеалізованої оптичноелектронних систем як показника якості. Друга методика використовує коефіцієнт кореляції вихідних сигналів як показник якості.

There are several requirements to an imaging system (IS) that defines the performance - energetic, spatial and temporal resolutions [1,2]. Energetic resolution is ability to distinguish small difference of output signal amplitudes that is limited by noise. Spatial resolution is ability to differ small objects or small parts of object image where the principal limitation factor is reduction of high spatial frequencies. Temporal resolution is ability to differ fast changing of output signal amplitude when object moves in a field of view. Its limitation factor is reduction of high temporal frequencies. Generally the energetic and spatial resolutions are the most important requirements in IS design.

The key problem of IS design is to maximize the integrate parameter of IS quality called performance [1,2]. It is obvious that the performance depends on the mentioned requirements. Thus the results of design procedures will be defined by the technique of performance evaluation. Now the classical approach for performance evaluation is calculation of MRTD and probability of target discrimination using the experimental data of human vision characteristics [1,2]. This technique has become FLIR90, FLIR92 standards [1,2,3]. But it has two principal disadvantages. First, we can apply it only for IS with human observer [1]. Now many IS can operates in automatic and semi-automatic mode when an embedded does target discrimination procedures. Second, it is difficult to apply this techniques for optimization of analog-digital or analog-digital-analog IS because effects connected with analog-digital conversion like aliasing, phasing effects make dramatic influence that can not be represented by a shape of MRTD [2,3]. Summarizing said above it is clear how it is important to investigate the objective techniques for performance evaluation that consider all principal effects of analog-digital and analog-digital-analog IS with or without human observer.

The objective techniques demands an objective measure of IS quality. An IS is used for image representation, for example, an analog-digital IS represents the input optical signal in form of a matrix of digital values, analog-digital-analog IS represents it in a form of a definite realization of analog signal. Thus the objective measure of IS quality will be defined by the distortions introduced by an IS into output signal. To calculate these distortions it is convenient to create mathematical abstraction as an idealized IS. The idealized IS is a model of an IS with infinity spatial, temporal and energetic resolutions. Infinity spatial resolutions mean that spatial an temporal MTF are equal to one and a detector and electronic do not produce noise. Mathematically we can calculate the output signal of this IS that is completely repeats input optical signal. Now we can introduce performance as a measure of similarity between output signals from an investigated IS and an idealized one. Of course, there are some possibilities to calculate signal difference. The least-square error shows the absolute value of the difference, coefficient of correlation clears the difference between linear combinations of the signal and the losses of information represents the difference of amplitude distributions [4,5,6]. But all these measures have two principal advantages. First, they are objective and we can use them for IS with human observer or without him/her. Second, they allow building for IS optimization a merit function which describes energetic and spatial resolutions.

The fidelity reflects the absolute difference between the output signals of investigated and idealized IS in sense of normalized least-square error [4,5]. When the input and output signals are analog the fidelity is calculated by the expression:

$$F = 1 - \frac{\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\left(U(x,y) - U_C(x,y)\right)^2 dxdy}{\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} U(x,y)^2 dxdy} \qquad (1)$$

where F - the fidelity,

x, y - the coordinates,

U(x,y), $U_C(x,y)$ - the output analog signals of the idealized IS and the real one, respectively.

In case of analog-digital IS when the output signal is digital it can be rewritten by the following ways [5]:

$$F = 1 - \frac{\sum_{i=1}^{N_X}\sum_{j=1}^{N_Y}(U(i,j) - U_C(i,j))^2}{\sum_{i=1}^{N_X}\sum_{j=1}^{N_Y}U(i,j)^2} \quad (2)$$

where i, j - the coordinates of pixels,

N_X, N_Y - the dimensions of the output signal matrix,

U(i,j), $U_C(i,j)$ - the output signals of the idealized IS and the real one that are two-dimensional arrays of pixels, respectively. The output signal of the real IS contains two parts: the signal losses after spatial transformations including spatial filtering, sampling, digital processing - $U_{SL}(i,j)$ and noise $U_N(i,j)$ [4,5]:

$$U_C(i,j) = U(i,j) - U_{SL}(i,j) + U_N(i,j) \quad (3)$$

We accept an assumption about absence of correlation between noise and the losses $U_{SL}(i,j)$ and noise $U_N(i,j)$. Now we can rewrite (2) in the following form doing some transformations:

$$F = 1 - \frac{\sum_{i=1}^{N_X}\sum_{j=1}^{N_Y}(U(i,j) - (U(i,j) - U_{SL}(i,j) + U_N(i,j)))^2}{\sum_{i=1}^{N_X}\sum_{j=1}^{N_Y}U(i,j)^2} = \quad (4)$$

$$= 1 - \frac{\sum_{i=1}^{N_X}\sum_{j=1}^{N_Y}U_{SL}(i,j)}{\sum_{i=1}^{N_X}\sum_{j=1}^{N_Y}U(i,j)^2} - \frac{\sum_{i=1}^{N_X}\sum_{j=1}^{N_Y}U_N(i,j)}{\sum_{i=1}^{N_X}\sum_{j=1}^{N_Y}U(i,j)^2} = 1 - \left(\frac{\sum_{i=1}^{N_X}\sum_{j=1}^{N_Y}U_{SL}(i,j)^2}{\sum_{i=1}^{N_X}\sum_{j=1}^{N_Y}U(i,j)^2} + \frac{\sum_{i=1}^{N_X}\sum_{j=1}^{N_Y}U_N(i,j)^2}{\sum_{i=1}^{N_X}\sum_{j=1}^{N_Y}U(i,j)^2}\right) =$$

$$= 1 - \left(NSL^2 + NEL^2\right)$$

where NSL – the normalized spatial losses coefficient that characterizes the relationship between the signal losses and energy of an output signal of an idealized IS:

$$NSL^2 = \frac{\sum_{i=1}^{N_X}\sum_{j=1}^{N_Y}U_{SL}(i,j)^2}{\sum_{i=1}^{N_X}\sum_{j=1}^{N_Y}U(i,j)^2}$$

NEL – the normalized energetic losses coefficient that characterizes the relationship between variance of noise and energy of the output signal of an idealized IS:

$$NEL^2 = \frac{\sum_{i=1}^{N_X}\sum_{j=1}^{N_Y}U_N(i,j)^2}{\sum_{i=1}^{N_X}\sum_{j=1}^{N_Y}U(i,j)^2}$$

It is possible to apply the expression (4) for optimization considering NSL and NEL as functions of IS parameters and coefficients of a digital filter:

$$F = 1 - \left(NSL^2 + NEL^2\right) = 1 - \left(NSL^2(p_1, p_2..p_M, c_1, c_2..c_N) + NEL^2(p_1, p_2..p_M, c_1, c_2..c_N)\right) \to max \quad (5)$$

where $p_1, p_2..p_M$ – the parameters of an IS,

$c_1, c_2...c_N$ – the coefficients of a digital filter.

According to (5) optimization changes a ratio between NSL and NEL to reach an optimum balance between energetic and spatial resolutions to guarantee the highest similarity of the output signals. Note, NSL and NEL can be found for any IS without human observer. But by adding characteristics of human vision into the process of calculation of NSL and NEL it is possible to get fidelity F for an IS with human observer [1,2]. It is obvious that F≡1 when an IS do not introduce any distortions into the output signal, other word when the IS acts as an idealized one. The case F=0 represents the situation when energy of the total distortions is equal to one of the output signal of idealized IS and the output signal of an IS is almost corrupted.

The coefficient of correlation between difference between the output signals of an idealized IS and a real one is an estimation of difference between its linear combinations [4]. The different gain

and average levels do no influence to the correlation and only differences of signal shapes reduce it. So the coefficient of correlation is also suitable measure of similarity of the output signals. For an analog-digital IS correlation coefficient can be written by the following way:

$$r = \frac{\sum_{i=1}^{N_X}\sum_{j=1}^{N_Y} U(i,j)\cdot U_c(i,j)}{\sqrt{\sum_{i=1}^{N_X}\sum_{j=1}^{N_Y} U(i,j)^2 \cdot \sum_{i=1}^{N_X}\sum_{j=1}^{N_Y} U_c(i,j)^2}} \quad (6)$$

where $r(U,U_{SL})$ – the coefficient of correlation between the output signals from an idealized IS and the real one.

We can do several transformation of (6) to define a relationship between correlation and fidelity:

$$r = \frac{\sum_{i=1}^{N_X}\sum_{j=1}^{N_Y} U(i,j)\cdot U_c(i,j)}{\sqrt{\sum_{i=1}^{N_X}\sum_{j=1}^{N_Y} U(i,j)^2 \cdot \sum_{i=1}^{N_X}\sum_{j=1}^{N_Y} U_c(i,j)^2}} = \frac{\sum_{i=1}^{N_X}\sum_{j=1}^{N_Y} U(i,j)\cdot\left(U(i,j)-U_{SL}(i,j)+U_N(i,j)\right)}{\sqrt{\sum_{i=1}^{N_X}\sum_{j=1}^{N_Y} U(i,j)^2 \cdot \sum_{i=1}^{N_X}\sum_{j=1}^{N_Y} \left(U(i,j)-U_{SL}(i,j)+U_N(i,j)\right)^2}} = \quad (7)$$

$$= \frac{\sum_{i=1}^{N_X}\sum_{j=1}^{N_Y} \left(U(i,j)^2 - U(i,j)\cdot U_{SL}(i,j) + U(i,j)\cdot U_N(i,j)\right)}{\sqrt{\sum_{i=1}^{N_X}\sum_{j=1}^{N_Y} U(i,j)^2 \cdot \sum_{i=1}^{N_X}\sum_{j=1}^{N_Y} \left(U(i,j)^2 - 2\cdot U(i,j)\cdot U_{SL}(i,j) + U_{SL}(i,j)^2 + U_N(i,j)^2\right)}} =$$

$$= \frac{1 - \frac{\sum_{i=1}^{N_X}\sum_{j=1}^{N_Y} U(i,j)\cdot U_{SL}(i,j)}{\sum_{i=1}^{N_X}\sum_{j=1}^{N_Y} U(i,j)^2} \cdot \sqrt{\frac{\sum_{i=1}^{N_X}\sum_{j=1}^{N_Y} U_{SL}(i,j)^2}{\sum_{i=1}^{N_X}\sum_{j=1}^{N_Y} U_{SL}(i,j)^2}}}{\sqrt{\left(1 - 2\cdot\frac{\sum_{i=1}^{N_X}\sum_{j=1}^{N_Y} U(i,j)\cdot U_{SL}(i,j)}{\sum_{i=1}^{N_X}\sum_{j=1}^{N_Y} U(i,j)^2} \cdot \sqrt{\frac{\sum_{i=1}^{N_X}\sum_{j=1}^{N_Y} U_{SL}(i,j)^2}{\sum_{i=1}^{N_X}\sum_{j=1}^{N_Y} U_{SL}(i,j)^2}} + NSL^2 + NEL^2\right)}} =$$

$$= \frac{1 - r(U,U_{SL})\cdot NSL}{\sqrt{\left(1 - 2\cdot r(U,U_{SL})\cdot NSL + NSL^2 + NEL^2\right)}} = \frac{1 - r(U,U_{SL})\cdot NSL}{\sqrt{2\cdot\left(1 - r(U,U_{SL})\cdot NSL\right) - F}}$$

where $r(U,U_{SL})$ – the coefficient of correlation between the output signal from an idealized IS and the spatial losses introduced by a real IS.

The expression (7) shows how to calculate the coefficient r using fidelity F and the coefficients $r(U,U_{SL})$ multiplied by NSL which describes spatial transformations of output signal. At the same time fidelity F represents influence of noise and distortion of spatial structure of the output signal. The coefficient r as measure of similarity of the output signals can also play a role of a merit function for IS optimization:

$$r = \frac{1 - r(U,U_{SL},p_1,p_2..p_M,c_1,c_2..c_N)\cdot NSL(p_1,p_2..p_M,c_1,c_2..c_N)}{\sqrt{2\cdot\left(1 - r(U,U_{SL},p_1,p_2..p_M,c_1,c_2..c_N)\cdot NSL(p_1,p_2..p_M,c_1,c_2..c_N)\right) - F}} \rightarrow max \quad (8)$$

Now we have two merit functions: the first one (5) characterizes the absolute difference and the second one (8)– relative difference of the output signals. As a fidelity (5) the correlation coefficient can be applied for optimization of IS with or without human observer. Both of them help to find an optimal balance between energetic and spatial resolutions but the weights of these resolutions will be different.

To illustrate the proposed technique we investigate performance of a typical staring thermal imager that has diffraction limited optics with numerical aperture 1:2, photon noise limited 256 x 256 focal plane array working in 8 – 12 mkm optical range with 0.04 sec frame period. Fig.1 shows dependence F (5) and r (8) from dimensions of a photosensitive element for a four bars test-object [1,2]. This test-object according to the international standards is characterized temperature difference ΔT, background temperature T and angular period of a bar α [1,2] (Fig.1). The growth of the dimensions causes bigger spatial losses due to high harmonics reduction and it increases signal-to-noise ratio due to the bigger absorbed radiant flux. Note, that maximum position of correlation coefficient is shifted into bigger values of the dimensions because it is less sensitive to spatial distortions than fidelity (Fig.1). It is important that in case of large and contrast test-objects dependence of different measures of performance from IS parameters is rather weak. In case of small dimensions and small contrast this dependence becomes strong and depended on the performance measure.

We introduced a model of an idealized imaging system as a powerful tool for performance evaluation. Now performance is considered as a measure of similarity of output signal from idealized IS and investigated one. The first possible measure is absolute - fidelity which represents absolute difference of amplitude values. The second one is relative - correlation coefficient which reflects only changes of signal shapes. These two interconnected measures can help to maximize performance by optimization balance between energetic and spatial resolutions.

References

1. J. Howe, "Electro-Optical Imaging System Performance Prediction", *Infrared and Electro-Optical Systems Handbook.* Editor Dudzik M., Volume 4, Chapter 2, SPIE Press, Bellingham, 1993.
2. G. Holst, *Electro-Optical Imaging System Performance.* SPIE Press, Bellingham, 1995.
3. W. Wittenstein, Minimum Temperature Difference Perceived- a new approach to assess undersampled thermal imagers. // Optical Engineering – Vol.38 – No. 5 – 1999 – p.773-781.
4. W.K.Pratt, Digital Image Processing, p.685-686, John Wiley & Sons Inc., New York, 1991.
5. S.K. Park, Z. Rahman, Fidelity analysis of sampled imaging systems.// Optical Engineering – Vol.38 – No. 5 – 1999 – p.786-800.
6. F.O. Huck, R.A. Gartenberg, S.K. Park, Z. Rahman, Informatic-theoretic assessment of sampled imaging systems. // Optical Engineering – Vol.38 – No. 5 – 1999 – p.742-762.

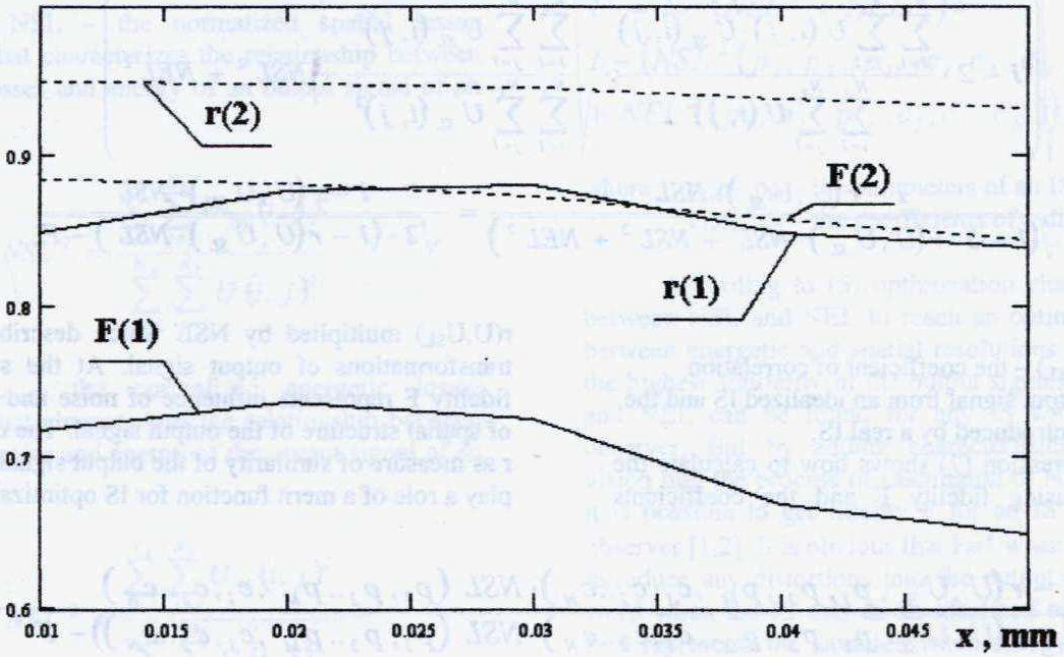


Fig.1. Relationship between the fidelity F, the correlation coefficient r and the dimension of a photo-sensitive element x (case 1 –four bars test-object with temperature difference ΔT = 0.05 K and angular period of a bar α = 1'; case 2 - ΔT = 0.1 K , α = 4', background temperature – T=293K).

ДВОТОЧКОВИЙ МЕТОД ПІДВИЩЕННЯ КОНТРАСТНОСТІ ЗОБРАЖЕНЬ

Роман Воробель, Олена Берегуляк

Фізико-механічний інститут НАН України, 79601, Львів, вул. Наукова, 5,
тел. (0322) 654803, vorobel@ah.ipm.lviv.ua

Two-point method for image contrast enhancement. Roman Vorobel, Olena Berehulyak The analysis of contrast enhancement for gray-scale monochrome images with the usage of local contrast enhancement under its linear description is realised. Two-point method of contrast enhancement of images based on definition of the sum of two neighbour elements' contrast relative to local background is proposed and proved. Example which illustrates the efficiency of proposed method by the way of generalised contrast enhancement is submitted.

1. ВСТУП

Одним з параметрів монохромного зображення, який суттєво впливає на його якість, є контраст. Тому, намагаючись поліпшити якість зображення, стараються перетворювати його так, щоб завдяки належному контрасту можна було добре розрізняти деталі об'єктів, що мають близькі за значенням яскравості. Серед багатьох класів методів поліпшення якості зображень, як в частотній області [1], так і в просторовій [2], виділяється метод перетворення локальних контрастів, започаткований у роботі [3]. Він був розвинутий далі у роботах [4-5], однак його суть при цьому не змінилась. Спільним недоліком методів цього класу є недостатня їх ефективнфсть. Тому за мету у даній роботі обрано підвищення ефективності методу перетворення локальних контрастів при покращанні зображень. Для її реалізації розглянемо відомий алгоритм [6] перетворення локальних контрастів та спосіб його модифікації.

2. ОДНОТОЧКОВИЙ МЕТОД ПІДСИЛЕННЯ ЛОКАЛЬНИХ КОНТРАСТІВ

Описаний у роботі [6] метод підсилення локальних контрастів побудований за триетапною схемою і використовує одноточкове визначення локального контрасту центрального елемента ковзного околу з координатами (i, j)

$$C(i,j) = \frac{\left|L(i,j) - \overline{L}(i,j)\right|}{LMAX}, \qquad (1)$$

де $\overline{L}(i,j)$ це усереднене значення яскравостей елементів локального околу, розміром $n \times m$:

$$\overline{L}(i,j) = \frac{1}{n \cdot m} \sum_{k=-\left[\frac{n}{2}\right]}^{\left[\frac{n}{2}\right]} \sum_{l=-\left[\frac{m}{2}\right]}^{\left[\frac{m}{2}\right]} L(i+k, j+l),$$

а $LMAX$ – максимально можлива яскравість елементів зображення L.

Другим етапом методу є нелінійне підсилення локального контрасту, яке полягає у реалізації перетворення виду

$$C^*(i,j) = F(C(i,j)), \qquad (2)$$

де за $F(C(i,j))$ обрано монотонні, опуклі функції типу $C^{\alpha}, 0 < \alpha < 1$, $\ln(1+\alpha)$, $th(\alpha)$ та ін.

Третім етапом є відновлення поліпшеного вхідного зображення, яке реалізовується через обчислення виразу

$$L^*(i,j) = \begin{cases} \overline{L}(i,j) - C^*(i,j) \cdot LMAX, \\ \qquad при\ L(i,j) \le \overline{L}(i,j), \\ \overline{L}(i,j) + C^*(i,j) \cdot LMAX, \\ \qquad при\ L(i,j) > \overline{L}(i,j). \end{cases} \qquad (3)$$

Однак описаний метод має недостатню ефективністю Для її збільшення нами запропонована модифікація цього методу, яка описана далі.

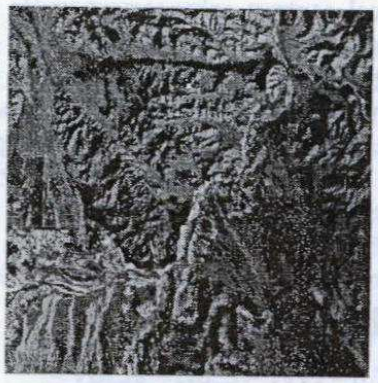
2. ДВОТОЧКОВИЙ МЕТОД ПІДСИЛЕННЯ ЛОКАЛЬНИХ КОНТРАСТІВ

Ідея модифікації методу базується на формуванні сумарного локального контрасту C_{Σ}, визначеного для двох сусідніх точок $L(i,j)$ і $L(i,j+1)$ [7]:

а) C_{gen}=0,3412

б) C_{gen}=0,0969

в) C_{gen}=0,3968



г) C_{gen}=0,1392

г) C_{gen}=0,4097

д) C_{gen}=0,1419

е) C_{gen}=0,4269

є) C_{gen}=0,1818

Рис. Ілюстрація ефективності двоточкового методу підвищення контрастності зображень

$$C_1(i,j)=\frac{\left|L(i,j)-\overline{L}(i,j)\right|}{LMAX},$$
$$C_2(i,j)=\frac{\left|L(i,j+1)-\overline{L}(i,j)\right|}{LMAX}, \quad (4)$$

звідки

$$C_\Sigma(i,j)=\begin{cases} C_1(i,j)+C_2(i,j), \\ \quad при \left|C_1(i,j)+C_2(i,j)\right|\le 1, \\ 1, при \left|C_1(i,j)+C_2(i,j)\right|>1, \end{cases} \quad (5)$$

чим завершується перший етап. На другому етапі проводиться аналогічне до перетворення (2) нелінійне підсилення сумарного локального контрасту:

$$C^*_\Sigma(i,j)=F(\left|C_\Sigma(i,j)\right|)$$

з використанням монотонних, однозначних і опуклих функцій. А на третьому етапі відновлюємо поліпшене вхідне зображення використовуючи при цьому вирази, які випливають з формул (4) і (5), а саме

$$L^*(i,j)=\begin{cases} 2\overline{L}(i,j)-L(i,j+1)+C^*_\Sigma(i,j)\cdot LMAX, \\ \quad при\, C_1(i,j)>0 \; i \; C_2(i,j)>0, \\ L(i,j+1)+C^*_\Sigma(i,j)\cdot LMAX, \\ \quad при\, C_1(i,j)>0 \; i \; C_2(i,j)\le 0, \\ L(i,j+1)-C^*_\Sigma(i,j)\cdot LMAX, \\ \quad при\, C_1(i,j)\le 0 \; i \; C_2(i,j)>0, \\ 2\overline{L}(i,j)-L(i,j+1)-C^*_\Sigma(i,j)\cdot LMAX, \\ \quad при\, C_1(i,j)\le 0 \; i \; C_2(i,j)\le 0. \end{cases} \quad (6)$$

Порівнюючи отриманий вираз (6) для відновлення вхідного зображення з відомим виразом (3) зазначимо, що вони подібні у двох випадках, а саме при $C_1(i,j)>0 \; i \; C_2(i,j)\le 0$ та $C_1(i,j)\le 0 \; i \; C_2(i,j)>0$. У всіх інших випадках запропонований метод має більше ніж у (3) підсилення локального контрасту на величину $\overline{L}(i,j)-L(i,j+1)$. Це дозволяє стверджувати, що запропонований метод більш ефективний, ніж описаний у роботі [6].

За кількісну оцінку ефективності методу обрано величину узагальненого контрасту C_{gen} зображення, при лінійному описі контрасту елементів [7]. Для перевірки ефективності методу проведено обробку зображень на його основі. Як приклад, на рисунку показано зображення "Лєна" (*а*) та "Гори" (*б*), що оброблялися за методом описаним у роботі [4] з використанням визначення за $\overline{L}(i,j)$ середньоарифметичного значення яскравостей ковзного околу – рис. (*в*) і (*г*); за методом [6] – рис. (*г*) і (*д*) та запропонованим методом – рис. (*е*) і (*є*). За функцію перетворення локального контрасту $F(C)$ вибиралася степенева функція C^α *з* $\alpha=0{,}75$, а усереднене значення $\overline{L}(i,j)$ визначалось у вікні 15×15 елементів. Отримані значення узагальненого контрасту зображень C_{gen} підтверджують його збільшення при обробленні запропонованим методом в порівнянні з відомими.

4. ВИСНОВКИ

Запропонований двоточковий метод підвищення контрастності зображень дозволяє отримати більше підсилення локальних контрастів через детальніше врахування околу центрального елемента з координатами (i,j) елементом з координатами $(i,j+1)$. Це сприяє точнішому і більш повному перетворенню яскравостей елементів зображення.

ЛІТЕРАТУРА

1. Hall E.L. *Computer Image Processing and Recognition*. New York: Academic Press.–1979.– 567p.
2. Прэтт У. *Цифровая обработка изображений*. М.:Мир, 1982.–790с.
3. Gordon R., Rangayyan R.M. Feature enhancement of film mammograms using fixed and adaptive neighborhoods. – *Applied Optics*. – 1984. – Vol.23. – № 4.– P. 560-564.
4. Beghdadi A., Negrate A.L. Contrast enhancement technique based on local detection of edges.– *Computer, Vision, Graphics and Image Processing*.– 1989. – V. 46.– №2. – P.162-174.
5. Dash L., Chatterji B.N. Adaptive contrast enhancement and de-enhancement.– *Pattern recognition*. – 1991. – V.24. – № 4. – P. 289-302.
6. Воробель Р. Нелінійне перетворення зображень з лінійним визначенням контрастности. Оброблення сигналів і зображень та розпізнавання образів. *Праці. Третя Всеукраїнська міжнародна конференція. 26-30 листопадп 1996 р. Україна*. Київ. – УАсОІРО. – 1996. – С. 194-196.
7. Воробель Р.А. Сприйняття сюжетних зображень та кількісна оцінка їх контрасту на основі лінійного опису визначення контрастності елементів. – *Доповіді НАН України*. –1998. – № 9. – С.103-108.

СИСТЕМА ІДЕНТИФІКАЦІЇ ОСОБИ ЗА СПЕКТРАЛЬНИМИ ОЗНАКАМИ ВІДБИТКІВ

Богдан Русин[1], Володимир Остап[2], Олег Остап[2]

[1] ФМІ НАН України, Львів 79601, вул Наукова 5а, т./ф.(0322)63-41-09, Rusyn@ipm.lviv.ua

[2] НУ "Львівська політехніка", Інститут Радіоелектроніки, Львів, 79046, вул. С.Бандери 12, т.(0322)93-16-91, Vova@ipm.lviv.ua, OstapOleg@org.lviv.net

Абстракт: Коротко описується система ідентифікації людини за відбитком пальця, блоки, методи обробки і розпізнавання. Під час розробки системи застосовано новий підхід до розпізнавання образів відбитків пальців людини, який кардинально відрізняється від загальновживаних в таких системах методів. Система відрізняється простотою і не має жодних вимог до кількості наявних особистих ознак на зображенні папілярного узору.

Ключові слова: відбитки пальців, розпізнавання, ідентифікація, попередня обробка.

ВСТУП

В сучасному суспільстві, в якому нікого вже не дивують нові пристрої, поява систем ідентифікації людини за відбитками пальців, всетаки, привернула увагу навіть найпрогресивніших людей. Ще недавно відбитки пальців асоціювалися виключно з криміналістикою, а вже тепер область їх застосування поширюється і на повсякденне життя. Уже нікого не дивуватимуть мобільні телефонами, портативні комп'ютери, кредитні карточки, електронні замки із вбудованими системами ідентифікації власника за відбитком пальця. Відбиток пальця аналогічний унікальному коду, який не потрібно запам'ятовувати.

Дослідження аналогічних не біометричних систем захисту показують, що найбільш вразливим місцем в них є код, який повинен бути унікальним і заодно його повинен запам'ятати користувач. Серед інших біометричних систем (ідентифікація за голосом, рогівкою ока, формою долоні, формою вуха, лицем, і т.д.) ідентифікація за відбитком пальця зайняла провідне місце завдяки простоті апаратного виконання і незмінності папілярного візерунку протягом всього життя людини [1].

Системи біометричної ідентифікації можна розділити на 4 групи в порядку спадання їх поширеності:

1) Системи з алгоритмами розпізнавання за особистими ознаками відбитків [2].

2) Системи базовані на ознаках зформованих набором фільтрів Габора [3].

3) Системи розпізнавання з застосуванням математичного апарату: розкладання зображень в базисах ортогональних функцій, кореляція зображень, нейронні мережі [4, 5].

Системи останньої групи застосовуються нечасто, але все більше дослідників приділяють їм увагу. Системи з першої групи вимагають наявності достатньої кількості особистих ознак (наприклад 20 [2]), системи ж другої групи не забезпечують достатньої точності ідентифікації. Практичні дослідження візерунків відбитків показали, що не всі відбитки мають достатню кількість особистих ознак, навіть при повному скануванні усього візерунку і, нажаль, такі ознаки не повністю описують всю наявну для розпізнавання інформацію (напрям і форма папілярних ліній), а лише описують розташування певних особливостей на ньому (закінчення, розвилка [1,2]). Методи з другої групи не описують прямо особисті ознаки, але все ж таки описують аномальну поведінку папілярного потоку ліній. Вище згадані недоліки усуваються в системах третьої групи, до якої відноситься розроблена.

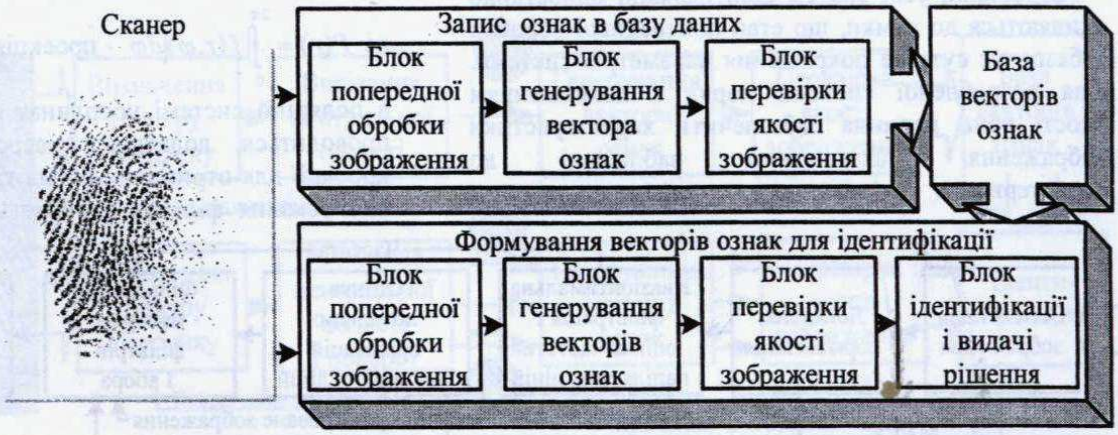


Рис.1. Загальна блок-схема системи ідентифікації.

Роботу системи можна бути представити блок-схемою зображеною на рис.1.

Система включає 4 блоки: блок вводу зображення (сканер), блок запису векторів ознак в базу даних, блок ідентифікації і базу даних (БД) з векторами ознак (ВО) відбитків пальців. Система працює в двох режимах – це ввід відбитків пальців у базу даних і режим ідентифікації. В обох режимах першим етапом передбачається попередня обробка, з метою отримання зображень з найближчими до оптимальних для розпізнавання параметрами. Другим етапом проводиться генерування ВО. За принципом роботи ці етапи однакові для двох режимів, але під час ідентифікації формується не один ВО, а цілий набір векторів ознак (НВО) для різних кутів повороту зображення. Це викликано необхідністю досягнення інваріантності системи до повороту з неінваріантним методом розпізнавання. ВО, які використовуються в цій системі дозволяють також визначити якість зображення і проводити грубе відкидання неякісних відбитків під час вводу і ідентифікації, що виконується у наступному етапі. Великого значення процес відкидання неякісних відбитків набуває для введення їх у БД, оскільки у цій системі виконується порівняння об'єктів (зісканованих зображень) із еталоном (зображення із БД), а вимоги завжди жорсткіші до якості еталонів.

Ідентифікація особи за відбитками пальців проводиться за НВО W порівнюючи їх з усіма ВО з БД. Рішення про ідентифікацію приймається коли коефіцієнт подібності для пари векторів: ВО з БД та одного ВО з НВО W перевищує певну межу K. Відмінність між автоматизованими дактилоскопічними інформаційно-пошуковими системами (АДІС) і біометричними системами ідентифікації закладена в формі видачі результату. Так АДІС видають як результат – список претендентів і степінь подібності до вхідного, а у системах ідентифікації необхідна чітка ідентифікації з єдиною відповіддю – є в БД ідентичний відбиток чи ні.

1. ПОПЕРЕДНЯ ОБРОБКА ВХІДНОГО ЗОБРАЖЕННЯ

Розробники усіх систем ідентифікації одностайно схиляються до думки, що етап попередньої обробки забезпечує суттєве покращення параметрів системи. Для розробленої системи поряд з покращенням якості вона повинна забезпечити характеристики зображення максимально наближені до характеристик ідеального дактилоскопічного зображення [6]. Блок-схема представлена на рис.2 відображає блоки попередньої обробки і їх взаємозв'язки.

Сегментація зображення. Суть процесу сегментації зображення зводиться до формування так-званої маски зображення, яка виділяє інформативну область зображення E, де знаходиться відбиток. Метою даного етапу є усунення фонового наповнення зображення і зменшення затрат часу під час наступної обробки, завдяки локалізації папілярного узору, і проведення наступних етапів в межах інформативної області E. Детально метод сегментації описаний в [6] .

Локальна нормалізація. Дуже рідко зустрічаються відбитки без спотворень типу "невідповідний контакт" [6]. Ця ситуація вимагає введення даного етапу. Він дозволяє повністю усунути вище згадані спотворення спрощеної форми (опік) та частково спотворення складної форми (пересихання шкіри, нерівномірність зволоження, шкірні захворювання) [6]. Після проходження зображення через даний етап вихідне зображення матиме однакові локальні значення математичного очікування і середньоквадратичного відхилення яскравості зображення в інформативній області E. Детальний опис приведено в [6].

Квазіоптимальна фільтрація і оцінка періоду папілярних ліній. В попередніх роботах нами було запропоновано використовувати адаптивну смугову фільтрацію для усунення ВЧ і НЧ шумів, як в роботі [7]. Така фільтрація дозволяє значно покращити оцінку локальної орієнтації папілярних ліній, але кращі результати отримуються при застосуванні комбінації адаптивної смугової і квазіоптимальної фільтрації. В цьому випадку формується квазіоптимальний двомірний фільтр з адаптивно обмеженою смугою пропускання.

Нехай $f(\omega_x,\omega_y)$ - спектр вхідного зображення $g(x,y)$, а $f(r,\varphi)$ - тойже спектр, але представлений в сферичній системі координат, тоді квазіоптимальний фільтр дактилоскопічного зображення з коловою анізотропією спектра матиме вигляд:

$$F(x,y)=P(r'),\quad r'=\sqrt{x^2+y^2}, \tag{1}$$

де $P(r)=\int_0^{2\pi}\left|f(r,\varphi)\right|d\varphi$ - проекція спектра на вісь OR в полярній системі координат (Рис.3). На практиці проводиться, додатково, усереднююча фільтрація проекції для отримання більш гладкої кривої. Смуга пропускання фільтра вибирається аналогічно смузі

Рис.2. Блок-схема попередньої обробки зображень.

пропускання адаптивного смугового фільтра описаного в [7].

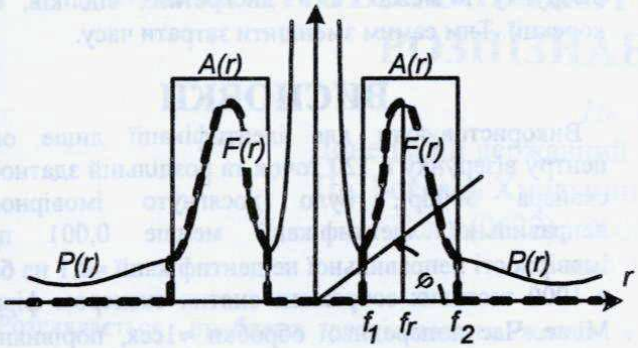


Рис.3. Січення амплітудно частотих характеристик на вісь OR адаптивного смугового $A(r)$ і квазіоптимального $F(r)$ фільтрів з проекцією спектра зображення $P(r)$ та смугою пропускання $f_1 - f_2$.

Під час оцінки смуги пропускання фільтра визначається пік проекції $P(r)$ в області середньочастотних складових спектра, який в свою чергу відповідає середній частоті папілярних ліній f_r. В даному випадку частота папілярних ліній виражає обернену величину до середнього періоду папілярних ліній [1].

Оцінка локальних орієнтацій папілярних ліній. Цей етап обробки не вносить жодних змін в саме зображення, а оцінює локальну орієнтацію папілярних ліній зображення в межах інформативної області *E*. На цьому етапі застосовується покращений метод оцінки локальних орієнтацій базований на градієнтах яскравості зображення, який детально описаний в [8].

Фільтрація зображення набором фільтрів Габора. Цей етап є найвагомішим в обробці, але не може бути якісно проведений без належної оцінки локальної орієнтації папілярних ліній і їх середньої частоти. Так неточності оцінки наведених величин призведуть не до покращення якості, а до повного спотворення візерунку. Нами використано набір з 16 фільтрів з кутами повороту імпульсних характеристик 0, 12, ..., 168 градусів. Детальний опис алгоритму фільтрації і фільтра Габора можна знайти у [8].

Перетворення гістограми. Після двох типів фільтрації (квазіоптимальної частотної та фільтрації фільтром Габора), гістограма зображення набуває характеру Гаусівської кривої, що кардинально відрізняється від гістограми ідеального дактилоскопічного зображення, яке має гістограму у вигляді параболи [6]. Цей етап дає не тільки візуальне покращення зображення але і забезпечує необхідні параметри вихідного зображення для генерування ознак та розпізнавання. Опис перетворення гістограми наведений в [6].

Після попередньої обробки зображення матиме такі, важливі для етапу розпізнавання, властивості:

1) Різниця локальної потужності яскравості зображення у двох довільних однакових за розміром областях зображення в інформативній області E прямує до нуля.

2) Основні спектральні складові зображення, що відповідають першим гармонікам спотворених гармонічних хвиль, зосередженні в СЧ смузі частот і утворюють колову анізотропію спектра.

2. ГЕНЕРУВАННЯ ВЕКТОРІВ ОЗНАК ТА ІДЕНТИФІКАЦІЯ

Використання кореляційного методу для порівняння двох зображень за спектральними ознаками не передбачає кореляції векторів, як при класичній інтерпритації таких методів, а кореляцію двох зображень в частотній області з використанням, відновлених з векторів, двох спектрів [9]. Такий метод розпізнавання зберігає основні переваги кореляційних методів, але враховує колову анізотропію спектра, що значно зменшує об'єм векторів ознак. Формування вектора в системі ідентифікації з усього зображення відбитку є недоцільним, а в більшості випадків і небажано, оскільки значно погіршує якість розпізнавання із-за наявних нелінійних геометричних спотворень візерунку, які змінюються від зміни умов формування відбитку. Дослідити і компенсувати такі спотворення неможливо із-за їх нестаціонарності. У

Рис.4. Розгорнута блок-схема блоків генерування ознак й ідентифікації.

випадку нашої системи, пропонуємо використовувати для ідентифікації особи лише центральну частину папілярного візерунку - навколо центру узору.

Загальна структура блоків генерування векторів ознак і блоку розпізнавання представлена на рис.4.

Використання для розпізнавання лише центральної частини узору зумовлене наступними факторами:

1) Центральна область найменш піддається геометричним спотвореням типу "губка".
2) Центральна частина узору є найбільш інформативною з точки зору унікальності.
3) Можлива чітка прив'язка до центру узору.
4) Центральна область відбитка має найкращий контакт із сенсором сканера.

Визначення центру візерунку. В цьому блоці застосований відомий алгоритм визначення центру візерунку за напрямками папілярних ліній і детально описаний в багатьох публікаціях [10].

Вирізання центральної області. Вирізається круговий окіл для того, щоб під час повороту, область зображення, яка корелюється, залишалася тією ж (Рис.4). В режимі вводу відбитка в БД окіл має діаметер 120 дискретних відліків, а в режимі ідентифікації 100.

Генерування векторів ознак. Ці блоки детально описані в попередній публікації [9], різниця полягає лише у використанні для генерування ознак центральної області візерунку, а не всього зображення.

Перевірка якості зображення. Основна мета цього етапу відкидання неякісних і неповних (неповністю відбита центральна область) зображень під час введеня їх в БД. В режимі ідентифікації цей етап допоміжний і не несе вагомого впливу, так як в цьому режимі межа відкидання за якістю значно менша. Для перевірки встановлюється межа потужності яскравості центрального околу, за умови перевищення якої вхідне зображення вважається придатним для розпізнавання (5E4 для вводу в БД і 2E4 для ідентифікації).

Ідентифікація. Під час ідентифікації корелюються зображення в частотній області за допомогою спектрів відновлених зі спектральних ознак. Вектори ознак W порівнюються з усіма векторами ознак V, записаними в БД [9]. Якщо нормоване значення піку взаємокореляційної функції перевищує порогове $K = 0.78$ то приймається рішення про ідентифікацію особи. У випадку якщо нормоване значення піку менше K, але більше $Q = 0.45$ то проводиться корекція координат центру візерунку згідно координат піка кореляційної функції і повороту зображень, після чого повторно порівнюються вектори ознак і приймається рішення про ідентифікацію, або неідентифікацію. Введення корекції координат центру дозволяє коректувати похибку алгоритму визначення центру візерунку, який, для певних типів візерунку, має значну похибку (до 30 дискретних відліків). Використання двох різних околів для вводу відбитка та ідентифікації дозволяє усунути чутливість системи до неточності визначення координат центру візерунку в межах ±10 дискретних відліків, без корекції. Тим самим зменшити затрати часу.

ВИСНОВКИ

Використовуючи для ідентифікації лише окіл центру візерунку в 120 точок та роздільній здатності сканера 500dpi, було досягнуто імовірності неправильної ідентифікації менше 0,001 при імовірності неправильної неідентифікації ≈0.1 на базі в 1000 тестових зображень знятих сканером фірми Mirae. Час попередньої обробки ≈1сек, порівняння двох зображень ≈0.2сек на комп'ютері з процесором PentiumIII-600МГц. Використання такого алгоритму дозволило максимально зберегти переваги кореляційних методів розпізнавання з одночасним зменшенням впливу геометричних спотворень і зменшенням затрат часу. Наступним етапом роботи буде поєднання даного методу з менш точними, але швидшими, для попереднього відсіюванням кандидатів з БД.

ЛІТЕРАТУРА

1. *Дактилоскопическая экспертиза: современное состояние и перспективы развития.* –Красноярск, 1990.
2. A.Jain, L.Hong, R.Bolle. *On-line fingerprint verification.* // IEEE Trans. Pattern Anal. And Machine Intell., 1997. №19(4), pp.302-304.
3. A.R.Jain, S.Prabhakar, L.Hong, S.Pankanti. Filterbank-based fingerprint matching. // IEEE Trans. Image Processing, 2000. №9(5), pp.846-859.
4. K.H.Fielding, J.L.Horner, C.K.Makekau. *Optical fingerprint identification by binary joint transform correlator.* // Optical Engineering, 1991. №30(12), pp.1958-1961.
5. P.Baldi, Y.Chauvin. Neural networks for fingerprint recognition. // Neural Computation, 1993. №5(3), pp.402-418.
6. Б.П.Русин, І.Н.Прудиус, В.П.Остап. *Спотворення й алгоритм попередньої обробки дактилоскопічних зображень* // Відбір і обробка інформації. 2002. Львів: ФМІ НАН України, №16, С.87-91.
7. Русин Б.П., Остап В.П., *Попередня фільтрація при розпізнаванні зображень.* // Комп'ютерні технології *у друкарстві*, 2000. Львів: Академія друкарства, №4, С.295-300.
8. Остап В.П., *Обробка дактилоскопічних зображень спрямованими фільтрами* // КМН-2001, 2001. Львів:ФМІ НАН України, С.205-210.
9. Остап В.П. *Кореляційний метод порівняння дактилоскопічних зображень за спектральними ознаками.* // // Вісник НУ "Львівська політехніка". 2002. №440, С. 188-195.
10. Р.Я.Косаревич, Б.П.Русин. *Детектування центра папілярного візерунка на зображенні.* // Відбір і обробка інформації. 2002. Львів: ФМІ НАН України, №16, С.92-95.

МОЖЛИВІСТЬ ВИКОРИСТАННЯ КОЛЬОРУ ПРИ РОЗПІЗНАВАННІ ЖЕСТІВ

Нюнькін К.М.
Донецький державний інститут штучного інтелекту
Пр. Богдана Хмельницького, 84, м. Донецьк, 83050
Тел.: (0622) 926082, факс: (0622) 926082
E-mail: kmn@iai.donetsk.ua

Розглядається проблема розпізнавання жестів у послідовності кольорових зображень. Зокрема, досліджується можливість використання кольору для визначення на зображеннях ділянок, що відповідають рукам. Показано обмеження використання для цих цілей кольору шкіри людини. Рекомендовано застосовувати для подачі команд жестами предмети насичених кольорів.

ВСТУП

Головна мета досліджень в галузі розпізнавання жестів полягає в тому, щоб створити систему, що може ідентифікувати певні людські жести і використовувати їх для передачі інформації або для керування різними пристроями.

Незважаючи на свою інтуїтивну зрозумілість, термін "розпізнавання жестів" різними дослідниками трактується по-різному. Дуже часто розпізнавання жестів ведеться з використанням спеціальних пристроїв, таких як рукавички з 10 - 20 датчиками (SensorGlove, PowerGlove, Cyberglove) [1, 2, 3] чи навіть за допомогою джойстика [4].

До області розпізнавання жестів відносять також стає останнім часом популярною задачу введення в комп'ютер з метою одночасного чи подальшого відображення формул, що малюються на звичайній дошці, схем, діаграм і т.п. [5]. Більш правильним представляється трактування поняття "розпізнавання жестів", яке полягає в тім, що приймання жестів людини здійснюється за допомогою відеокамери (аналогічно людському зору) і не вимагає застосування спеціальних пристроїв, що користувач повинен був би надягати або якими йому необхідно було б керувати.

Таким чином, при розпізнаванні жестів необхідно виявляти на зображенні руки людини і прослідковувати зміну їхнього положення чи форми з часом. Для цього звичайно використовують або детектування руху [6, 7, 8], або виявлення областей, що мають колір шкіри, [9, 10] або обидва ці підходи використовують спільно [11].

Як правило жести, що виявляються, мають досить великий масштаб. Це можуть бути вказуючи жести, у яких задіяна вся рука, або рухи рук, подібні до крил, що махають [6]. У таких ситуаціях звичайно разом із кольором використовується детектування областей руху.

При розпізнаванні дрібномасштабних жестів, у яких задіяні окремі пальці, кисть руки займає істотну частину кадру [7, 8]. Фон в цих випадках однорідний і контрастує по кольору зі шкірою людини.

1 ВИЯВЛЕННЯ НА ЗОБРАЖЕННІ ОБЛИЧЧЯ З ВИКОРИСТАННЯМ КОЛЬОРУ

Розглянемо критерії виділення на зображенні областей, що мають колір людської шкіри. Ця задача вирішується не тільки при розпізнаванні жестів, але і при виявленні обличчя людини. У той же час, дослідження з виявлення на кольорових зображеннях обличчя ведуться більш інтенсивно.

При виявленні на зображенні галузей, що мають колір шкіри, поряд зі звичайним RGB представленням кольору, тобто інтенсивностями червоної, зелений і синьої складової кольору, використовується представлення, засноване на кольоровості - HSL представлення (Hue - колір або відтінок, Saturation - насиченість, Luminosity - яскравість):

$$H = \frac{arctan(y/x)}{2\pi}, \quad S = \sqrt{x^2 + y^2}, \quad L = \frac{R+G+B}{3},$$

$$x = R - 0.5 \cdot (G + B), \quad y = \frac{\sqrt{3}}{2}(G-B).$$

Li & Kobatake [12] використовували колірний простір YIQ, що відповідає стандарту NTSC американського телебачення, для локалізації обличчя і його деталей. Для локалізації обличчя автори використовували закономірність, що "для жовтої раси в представленні YIQ обличчя має значення Q більше, а значення I менше, ніж темне оточення".

Як недоліки описаного підходу сегментації обличчя людини можна відзначити наступні моменти:

- У даному підході використані співвідношення між кольорами визначених елементів обличчя для конкретної "жовтої" раси. Для людей інших рас багато чого не має бути застосовано.
- Не враховується можлива наявність бороди, вусів, окулярів, нафарбованих губ і т.п.
- Фон і волосся покладаються темними, що не завжди справедливо.

Крім того, схоже, автори не праві, коли

стверджують, що "обличчя має значення Q більше, а значення I менше, ніж темне оточення". Насправді, як показують наші експерименти, закономірності виходять зворотні. Великі значення Q і маленькі I відповідають синім ділянкам, у той час як маленькі значення Q і великі I - жовтим.

Tock & Craw [13] для виявлення обличчя водія в автомобілі, що займає істотну частину кадру, використовували колірний простір, аналогічний HSL. Формули для переведення RGB->HSL не приводяться, але вони відрізняються від загальноприйнятих, оскільки для кольору обличчя наводиться діапазон $150 \le H \le 200$ (звичайно H в області обличчя ~25-40), і, відповідно до приведених картинок, "колір" відраховується від зеленого, а не від червоного. Поряд із указаним діапазоном H, для області обличчя приймається наступний діапазон S: $128 \le S \le 224$, тобто зображення повинні мати добру колірну насиченість. За звичай значення S дуже рідко перевищують 100, а типові значення для області обличчя лежать у діапазоні 30-90.

Sirohey & Rosenfeld [14] для виявлення обличчя, що займає значну частину кольорового зображення з однорідним фоном, використовували дві нормалізовані компоненти RGB представлення:

r=R/(R+G+B)*100 і g=G/(R+G+B)*100.

Автори провели дослідження, у яких на кольорових зображеннях вручну задавалася область обличчя і розраховувалися гістограми розподілу r і g. Отримана двовимірна гістограма апроксимувалася гаусіаном, у результаті чого для бази даних Aberdeen (UK), що містить 72 фотографії, було отримано, що центр розподілу відповідає точці μ=[42.91,32.28] з дисперсією σ=[19.28,-5.45;-5.45,8.55].

Аналогічні дослідження були проведені для власної бази даних з 21 фотографії з не зовсім однорідним фоном. Для цієї бази даних були отримані трохи інші параметри гаусіана: μ=[52.66,29.99], σ=[68.94,-30.89;-30.89,18.11].

Автори відзначають, що навіть незважаючи на нормировку, колір обличчя на фотографії залежить від умов освітлення, особливо від кольору джерела світла. У той же час вони вказують, що, якщо для даних умов освітлення визначити параметри гаусіана, те метод працює досить стійко.

Варто помітити, що параметри гаусіана залежать від умов освітлення, використовуваних камер і їхніх настроювань. Тому визначити межі змін колірних компонентів (навіть нормалізованих) раз і назавжди не можна. Час від часу потрібна їхня корекція.

У технічному звіті лабораторії AT&T [15] для локалізації обличчя використовується три канали: 1) колір, 2) рух, 3) аналіз форми. Основна мета - скорочення часу при роботі з послідовністю кадрів. Звідси і використання каналу кольору як найшвидшого. Однак він також вважається і самим ненадійним. Тому його результати періодично підтверджуються аналізом руху.

Для колірного аналізу використовуються нормалізовані значення rgb, що дозволяє зменшити залежність алгоритму від яскравості. При цьому темні пікселі, тобто ті, у яких сума (R+G+B)<30, обнуляються. Точні критерії вибору області обличчя не приводяться. Відповідно до малюнка розподілу відносних кольорів зображення в площині rg, на якому нахиленим еліпсом, що може бути уписаний у прямокутник r=[125,...,175], g=[50,...,90], обведена область, що відповідає "кольору шкіри".

Обличчя в даній роботі є домінуючим об'єктом на зображенні. Тому, якщо пікселів з кольором обличчя в одному зв'язному компоненту стане занадто мало чи занадто багато, такий аналіз повинен бути припинений.

Звернемо увагу на те, що автори не дуже довіряють колірної сегментації, постійно перевіряють її результати і мають критерії її припинення. Також варто врахувати, що колірна сегментація вимагає виконання попереднього калібрування колірного простору.

В авторефераті Талеба з ИТК НАН Бєларусі [16] при ідентифікації ділянок шкіри людини на кольорових зображеннях також використовувався колірний простір HSL разом з деякими комплексами, складеними з компонентів R, G і B.

Використаний ним набір ознак, з одного боку, виглядає трохи перевантаженим, а з іншого боку - не містить насиченість S.

Аналізуючи роботи з виявлення обличчя людини на основі інформації про колір, не можна не помітити наступні моменти:

1. Колір шкіри в різних людей (навіть однієї раси) варіюється в досить широких межах.
2. Колір шкіри залежить від умов освітлення.
3. Перед виконанням виявлення обличчя цим методом бажано виконати калібрування колірного простору.
4. Найчастіше при виявленні обличчя канал кольору використовується разом з іншими: рухом, градієнтом яскравості, формою, аналізом оточення.

Імовірно, зазначені обмеження і послужили причиною того, що такі методи не знайшли широкого застосування.

2 ВИЯВЛЕННЯ НА ЗОБРАЖЕННІ РУК ЛЮДИНИ

Для дослідження колірних характеристик областей обличчя і рук людини була розроблена спеціальна комп'ютерна програма, що дозволяє вводити кольорові зображення, одержувані підключеної до комп'ютера веб-камери, а також із вже існуючих файлів. Програма розраховувала компоненти різних представлень кольору (RGB, HSL, YIQ) і дозволяла виділяти на зображенні ділянки, що відповідають різним критеріям, складеним з цих компонентів.

Експерименти, проведені з використанням указаної програми, показали, що

- Значення кольору (H) шкіри людини можуть лежати в діапазоні від 4 до 40.
- Значення насиченості (S) шкіри людини можуть лежати в діапазоні від 30 до 90.
- Колір шкіри у різних людей варіюється в досить широких межах.
- Колір шкіри навіть однієї людини залежить від умов освітлення.
- Одержуваний на зображенні колір шкіри людини залежить від настроювань камери, що найчастіше виконуються автоматично і не контролюються.
- Колір, близький до кольору шкіри людини, можуть мати деталі одягу, що оточують предмети і стіни приміщення (див. таблицю 1).

Таблиця 1 - Об'єкти, що мають колір, близький до кольору шкіри людини (H=4-40)

Об'єкт	Значення H	Об'єкт	Значення H
Стіни	17-22	Дошки	7-27
Штори	17-19	Бетонні доріжки	20-30
Цегельний будинок	5-18	Камені	17-22
Автобус	14-20	Пісок	20-25
Підфарники	9-20	Листи	31-45

Таким чином, використання кольору шкіри людини при створенні системи розпізнавання жестів для роботи в реальних умовах (неконтрольовані умови освітлення, різні люди оператори, довільний навколишній фон) виглядає дуже проблематичним. Урахування руху як додаткового критерію при виявленні рук людини мало що змінює, оскільки в реальних умовах у кадрі можуть рухатися не тільки руки оператора, але і навколишні об'єкти.

У той же час, у проведених експериментах, у тому числі і поза приміщенням, було відзначено вкрай мало об'єктів, що мають насичені кольори (зі значеннями S, що перевищують середину можливого діапазону, тобто 128). Це були деталі одягу і яскраві квіти, у яких значення S перевищували 160. А от колір зелені в переважній більшості випадків не дуже насичений: значення S найчастіше не перевершує 70, хоча у яскраво освітленої сонцем свіжої зелені може досягати 150 (з 255 можливих). У таблиці 2 наведені об'єкти, які мають високі значення насиченості, що виявляються поза приміщенням у ясний сонячний день. Слід зазначити, що такі об'єкти були виявлені тільки на 19 кадрах з 71, хоча оператор знімав саме ті об'єкти, що на його думку мали яскраві насичені кольори.

Таблиця 2 - Об'єкти, що мають високі значення насиченості (S>128)

Об'єкт	Значення S	Значення H
Червона машина	150-190	250-256
Жовтий автобус	130-160	17-24
Свіже листя на деревах	130-150	49-57
Червоні троянди	160-210	245-256
Зелена вода в басейні	130-135	42-46
Рожева блузка	140-150	237-245
Малинова сукня	130-160	235-247
Червона блузка	130-180	250-255

На мал. 1 наведені гістограми розподілу насиченості на зображеннях, знятих поза приміщенням у ясний сонячний день. Можна помітити, що кількість точок, що мають дану насиченість, зменшується з її ростом за показовим законом.

Малюнок 1 - Гістограми розподілу насиченості:
а) усі зображення, б) зображення, що не містять насичених кольорів, в) зображення з об'єктами, що мають насичені кольори.

Наведені вище результати експериментів говорять про те, що при виявленні рук оператора можна використовувати насичені кольори, для чого варто застосовувати надягнуті на руки рукавички визначеного насиченого кольору. Поріг насиченості при цьому, імовірно, буде залежати від яскравості освітлення.

ВИСНОВКИ

Використання кольору шкіри людини при виявленні на зображенні обличчя і рук оператора має певні істотні недоліки, головними з який можна вважати різниці кольору цих об'єктів у різних людей, залежність кольору від умов освітлення, наявність великої кількості інших об'єктів, що мають схожі кольори. І якщо перші два можуть бути усунуті калібруванням колірного простору, проведеної перед кожним сеансом роботи, то останній у випадку неможливості усунення з кадру сторонніх об'єктів, близьких по кольору до шкіри людини, перетворюється в серйозну проблему. Та й постійне виконання калібрування теж не завжди можливо. У результаті використання кольору при виявленні на зображенні обличчя і рук оператора має обмежену область застосування і найчастіше використовується разом з іншими критеріями: рухом, градієнтом яскравості, формою, аналізом оточення.

З іншого боку, проведені експерименти по дослідженню колірних характеристик різних об'єктів показали, що вкрай мало об'єктів мають насичені кольори. При виявленні рук оператора можна використовувати цей факт, для чого варто застосовувати надягнуті на руки рукавички визначеного насиченого кольору. Для рішення питання про поріг насиченості і його можливій залежності від умов освітлення необхідне проведення додаткових досліджень.

ЛІТЕРАТУРА

1. Murakami K., Taguchi H. Gesture recognition using recurrent neural networks. // Journal of the ACM, 1(1), January 1991. -p.237-242.
2. Hofmann F., Henz Ue. The TU-Berlin SensorGlove. - Technical Description. 1995. - http://pdv.cs.tu-berlin.de/forschung/SensorGlove2_engl.html.
3. Fels S., Hinton G. Glove-TalkII: An Adaptive Gesture-to-Formant Interface. // Proc. of the 1995 Conf. on Human Factors in Computing Systems, May 1995.
4. Keates S, Clarkson PJ, Robinson P. Gesture input and motion-impaired users. // Encyclopedia of Library and Information Science, Marcel Dekker, Inc., 2001, Vol.70. -p.128-150.
5. Damm, C.H., Hansen, K.M., Thomsen, M. Tool Support for Object-Oriented Cooperative Design: Gesture Based Modeling on an Electronic Whiteboard. // Proc. of CHI'2000. ACM Press, 2000.
6. Cutler R., Turk M. View-based Interpretation of Real-time Optical Flow for Gesture Recognition. // IEEE Int.Conf.on Automatic Face and Gesture Recognition, April 1998, Nara, Japan.
7. Laptev I., Lindeberg T. Tracking of multi-state hand models using particle filtering and a hierarchy of multi-scale image features.- Technical report CVAP245, ISRN KTH NA/P--00/12--SE. Department of Numerical Analysis and Computer Science, KTH (Royal Institute of Technology), Stockholm, Sweden, September 2000.
8. Gupta D. Computer Gesture Recognition: Using the Constellation Method. // Caltech Undergraduate Research Journal, 2001, vol.1, N1. - pp.26-31.
9. Marcel S., Bernier O., Viallet J.-E., Collobert D. Hand Gesture Recognition Using Input-Output Hidden Markov Models. // Proc. of the Fourth IEEE Int. Conf. on Automatic Face and Gesture Recognition 2000.
10. Bretzner L., LaptevI., Lindeberg T. Hand Gesture Recognition using Multi-Scale Colour Features, Hierarchical Models and Particle Filtering. // Proc. of the 5th Int. Conf. on Automatic Face and Gesture Recognition, 2002.
11. McKenna S., Gong S. Gesture Recognition for Visually Mediated Interaction using Probabilistic Event Trajectories. // Proc. of BMVC'98, 14-17 September 1998, Southampton, England.
12. Li Y., Kobatake H. Extraction of Facial Sketch Images and Expression Transformation Based on FACS. // Proc. of IEEE Int. Conf. on Image Processing. - Washington D.C., 1995, v3.
13. Tock D., Craw I. Tracking and measuring drivers eyes. // Proc. of BMVC-95, -p.383-392.
14. Sirohey S., Rosenfeld A. Eye Detection. - Technical report CS-TR-3971 (CAR-TR-896) Center for Automation Research University of Mryland, 1998. -49p.
15. Graf H.P., Cosatto E., Gibbon D., Kocheisen M., Petajan E. Multi-Modal System for Locating Heads and Faces. - AT&T Lab Technical Report 95.5.1, 1996.
16. Талеб М.А. Комбинированные алгоритмы сегментации цветных изображений. - Автореферат диссертации на соискание ученой степени кандидата технических наук. Минск, Ин-т технической кибернетики НАН Беларуси. 2002р. -20с.

ІДЕНТИФІКАЦІЯ ПІДПИСІВ ПО ХАРАКТЕРНИМ ОЗНАКАМ ПЕРЕТВОРЕННЯ РАДОНУ

В.В.Трипольська
Дніпропетровський національний університет
49050, Дніпропетровськ, пров. Науковий , корпус 12
E-mail: trip@ukr.net

АНОТАЦІЯ

Представлено метод нейромережової ідентифікації підписів як бінарних зображень у просторі інформативних ознак перетворення Радону. Використовуються моменти інваріантні до обертання, зсувів та масштабних перетворень для побудови вектора особливостей. Добуття інформативних ознак одержується в просторі вейвлет декомпозиції, псевдофазовой системі координат, що забезпечує з однієї сторони надійність ідентифікації, з іншого боку – зменшується розмірність простору ознак.
Запропонований метод здобуття особливостей і ідентифікації підписів перевірений на нейроній мережі зворотного поширення помилок. Тестування проводилося на різних підписах, і отримані результати підтвердили надійність методу.

Ключові слова: ідентифікація підписів, перетворення Радону, вейвлет декомпозиція, інваріантні моменти, псевдофазовая система координат, нейронна мережа.

1.ВСТУП

Ідентифікація підпису як частина проблеми розпізнавання рукописних символів займає важливе місце й усе частіше привертає увагу наукового світу. У цій області було досліджено багато різних підходів, але в цілому проблема ще далека від остаточного рішення [1]. Головні труднощі полягають у постійній мінливості підпису кожного індивіда, що викликає проблеми зі створенням надійного еталона.

У цьому зв'язку, дуже актуальною є розробка інтегральних методів виділення стійких ознак підписів при розгляді кожного підпису як бінарного зображення. Акцент саме на виділенні інтегральних інформативних ознак зв'язаний з фізіологічним сприйняттям підпису як єдиного цілого, принаймні в первісний момент процедури ідентифікації.

Для виділення інтегральних ознак при верифікації підписів традиційно використовуються методи спектрального аналізу (Фур'є, Адамара і т.д.). Однак спектральний підхід дуже часто не дає бажаних результатів, тому що підпис, як бінарне зображення, характеризується широким енергетичним спектром, у результаті чого, відмінність спектра підробки від спектра оригіналу може бути незначною. У цьому зв'язку нами було запропоновано використовувати перетворення Радону для переходу від бінарного зображення до напівтонового з 256 градаціями сірого[2].

Методи виділення інформативних ознак включають статичні підходи, засновані на визначенні моментів[3, 4], моделі авторегрессій, вейвлет перетворення [5]. Клас моментів інваріантних до повороту, зсуву і масштабних перетворень застосовується як вектор особливостей підпису, тому що зображення може бути зміщене в межах заданої рамки на документі. Використання вейвлет декомпозиції дає можливість виділення різних параметрів і збереження самого зображення без збільшення обсягу інформації. Набір отриманих параметрів є вхідними даними для класифікатора. Вектором вхідних особливостей є інваріантні моменти й коефіцієнти вейвлет декомпозиції після перетворення Радону.

Нейромежевої підхід до ідентифікації підпису як ідентифікації бінарного зображення дає недостатню надійність як через мінливість самих підписів так і варіантів їхнього місцезнаходження в межах заданої рамки документа.

Процес визначення приналежності зразка до одного з класів може бути визначений за допомогою багатошарового персептрона, байсевського чи класифікатора нейронної мережі [6]. Особливістю даного методу є те, що відомо параметри тільки одного класу, істинного підпису. Тому як класифікатор використовується нейрона мережа.

Тестування методу проводилося на 50 наборах підписів, що складаються з еталона, 2 істинних підписів і 2 підробок.

2.ПЕРЕТВОРЕННЯ РАДОНУ

При автоматичній класифікації підписів однією з труднощів є чутливість до варіантів місця розташування підпису в межах заданої рамки (на документі). У цьому зв'язку бажано перейти від бінарного зображення до яркостного. Для цієї мети використовується перетворення Радону, тому що має

ряд переваг, наприклад, у порівнянні з енергетичним спектром Фур'є. [7]

При розгляді бінарного зображення, підданого перетворенню Радону як напівтонового зображення (рис.1) виділяються «блискучі» крапки, що можуть використовуватися як інформативні ознаки. На нейронну мережу подавався характеристичний вектор, що містить повне зображення. Але через проблему розмірності, кількість вхідних параметрів необхідно зменшити. Одним з можливих рішень цієї проблеми є використання коефіцієнтів вейвлет декомпозиції як інформативних ознак нейронної мережі.

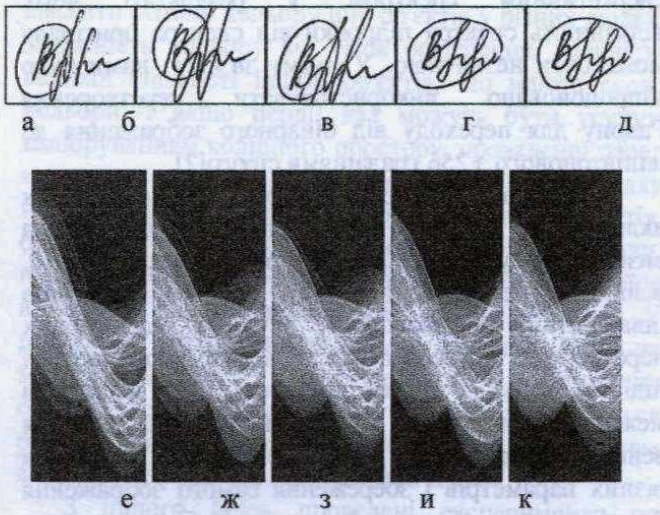


Рис. 1 Вихідні підписи і відповідне їм перетворення Радону.
а – еталонний підпис, б, в – справжні підписи, г, д – підробки; е,ж, з, и,к – відповідні перетворення Радону.

3.ВИТЯГ ІНФОРМАТИВНИХ ПРИЗНАКІВ

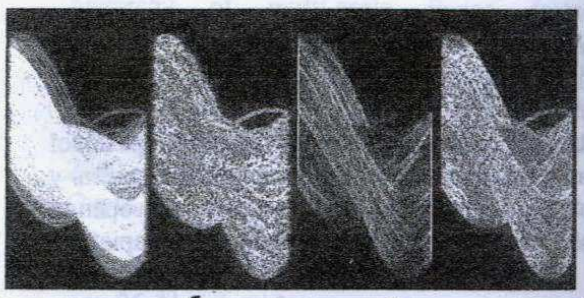
3.1.ВЕЙВЛЕТ

Оскільки растрове зображення є двовимірною матрицею, то до нього можуть бути застосовані двовимірні дискретні вейвлет-перетворення. Вейвлет-перетворення визначається формулою:

$$W(\alpha,\beta)=\sum_{-\infty}^{\infty}\frac{1}{\alpha}\overline{\psi\left(\frac{x-\beta}{\alpha}\right)}f(x) \qquad (1)$$

Підвищення точності класифікації підписів і зменшення розмірів параметричного вектора досягається застосуванням декомпозиції зображення за допомогою вейвлет перетворення. Застосування різних фільтрів (базисів) дає результати, що відрізняються по своїм характеристикам. Аналізуються деталі зображення отримані на перших рівнях розкладання, які характеризують особливості, що притаманні данному підпису. Використання декількох рівнів для аналізу обумовлене тим, що на більш низькому рівні розкладання залишаються загальні властивості підпису і деталі, що зустрічаються тільки в даному екземплярі. Ці параметри також є частиною сукупності інформативних ознак, що характеризують підпис. [8]

Справжній підпис



Підробка

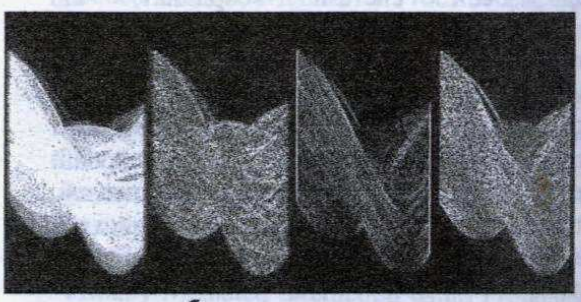


Рис.4. Коефіцієнти після вейвлет декомпозиції
а- вихідне зображення, б – горизонтальні, в – прямовисні, г – діагональні

3.2.ІНВАРІАНТНІ МОМЕНТИ

Через мінливість місця розташування підпису, необхідно ввести додаткові ознаки, що не будуть змінюватися при повороті, масштабуванні й обертанні зображення. Одним з підходів є обчислення різних інваріантних моментів. Моменти низьких порядків відповідають за «закономірності», а більш високі моменти визначають особливості зображення.

У роботі [9] запропоновані сім нелінійних комбінацій, побудованих на моментах, що інваріантні до зсуву, обертанню і масштабированию. Ці сім, так називані, інваріантні моменти, визначені через центральні моменти і використовуються для класифікації образів.

На нейронну мережу подавався вектор параметрів, що містить значення моментів, до 4-го порядку включно, і інваріантні моменти. Проведені експерименти показали, що даний вектор інформативних ознак не дає достатньої надійності ідентифікації підписів, у зв'язку з чим виникає

необхідність збільшення простору інформативних ознак.

Ця задача була вирішена за допомогою застосування інваріантних моментів Цернике. У роботі [9] застосовуються ортогональні моменти, засновані на теорії ортогональних поліномів, що ефективні для вирішення задач ідентифікації зображень. Клас ортогональних моментів, що володіють стійкістю до обертання – моменти Цернике. Поворот зображення не змінює величину моментів Цернике, тому вони використовуються як один з параметрів при визначенні вектора особливостей зображення. Для одержання інваріантів зсуву і масштабування, зображення спочатку нормується, використовуючи для цього центральні моменти. Моменти Цернике, інваріантні до обертання одержуються з нормованих моментів. У роботі [10] наведено зв'язок між інваріантними моментами, інваріантами Цернике і центральними моментами.

На нейронну мережу подавалися різні комбінацій інваріантних моментів. Найкращі результати отримані при використанні семи інваріантних моментів, запропонованих Ху й 11 моментів Цернике.

3.3.ПСЕВДОФАЗОВА СИСТЕМА КООРДИНАТ

Одним з методів рішення проблеми компресії є підхід до витягу інформативних ознак перетворення Радону в псевдофазовой системі координат. Базисна ідея підходу полягає в тім, що при переході в нову систему координат, можна виділити невелику ділянку, в якій і будуть зосереджені основні локальні зміни підписів, тоді як у більшій частині нової площини будуть зосереджені загальні (практично незмінні) ознаки.

Якщо ввести позначення

$$R_\Theta(x')=R \qquad (1^0 \le \Theta \le 180^0), \qquad (2)$$

те відображення залежності (2) у «фазовій» системі координат [11] зводиться до відображення в площин $(R - |\dot{R}|)$, де

$$|\dot{R}| = \left[\left(\frac{\partial R_\Theta(x')}{\partial \Theta}\right)^2 + \left(\frac{\partial R_\Theta(x')}{\partial x'}\right)^2\right]^{1/2} \qquad (3)$$

відкіля випливає, що при такому відображенні інформація про координати вихідного зображення $R_\Theta(x')$ відкидається (аналогія: механіка Ньютона і механіка Лагранжа), що, у принципі, і дозволяє сконцентрувати локальні розходження аналізованих підписів у невеликій частині площини $(R - |\dot{R}|)$.

Однак такий підхід, з огляду на специфіку перетворення Радону для підписів (наявність «блискучих» крапок), із практичної точки зору незручний, тому що в цих крапках похідна має розрив, що приводить до зростання впливу випадкових (шумових) компонент. Тому нами була використана «псевдофазова» система координат [12], що зводиться до відображення в площині

$$R_{\Theta_k}(x'_l) - R_{\Theta_{k+1}}(x'_{l+1}) \qquad (4)$$

$$1^0 \le k \le 180^0; \quad 1 \le l < L-1,$$

де L – розмір рамки вихідного підпису на осі «x».

Як випливає з розгляду псевдофазових характеристик на рис.2, основні локальні зміни групуються у відносно невеликій лівій верхній частині аналізованої площини, що і відкриває можливість істотного скорочення розмірності ознакового простору з одночасним забезпеченням високої чутливості до виявлення підроблених підписів.

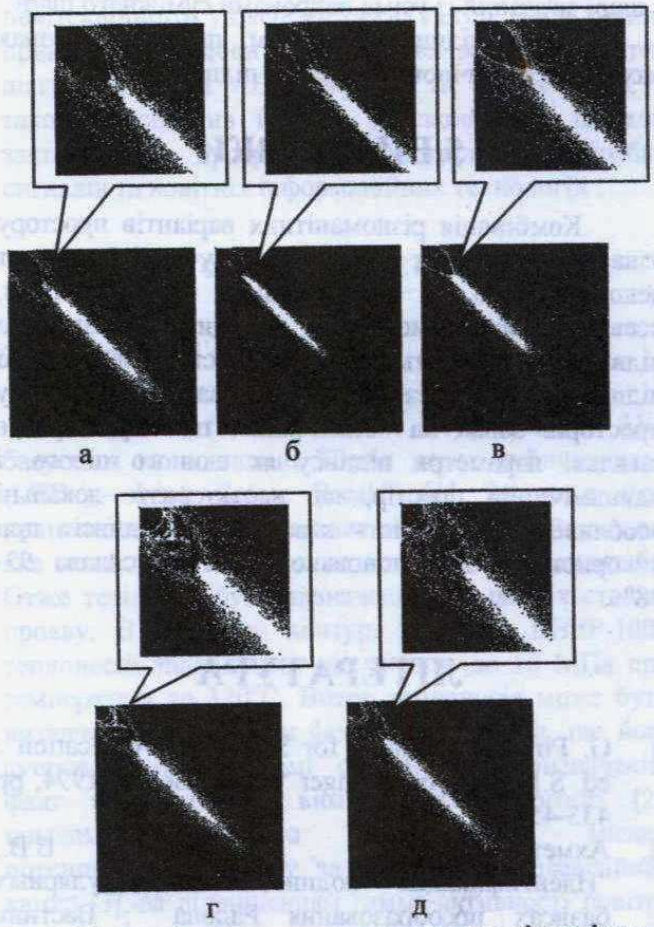


Рис.2. Псевдофазова характеристика, що відповідає перетворенню Радону

а – еталонний підпис, б, в – справжні підписи, г, д – підробки.

4.ПОБУДОВА «КАРТИ ПІДПИСУ»

Для зменшення кількості інформативних ознак можна використовувати карту зображення. Вивчення характеристик зображення показало, що просторові частоти, одержані з блоків зображення, адекватно розділяються в класи. Зображення розбивається на непересічні блоки розміром 8 x 8 пикселов. Обчислюється спектр двовимірного перетворення Фур'є до блоку зображення. Процес класифікації кожного блоку зображення зроблений за допомогою нейронной мережі, що використовує алгоритм зворотного поширення помилки. Блок зображення, отриманого після перетворення Фур'є, повинний бути віднесен до одного з 8 класів напрямків ліній: 0.0 °, 22.5 °, 45.0 °, 67.5 °, 90.0 °, 112.5 °, 135.0 ° и 157.5 °. Ці класи є виходами нейронної мережі. Навчання мережі виконувалося таким чином, щоб активним на виході був тільки один нейрон. Для розглянутого застосування, рівень точності 8 класів є достатнім тому, що представлено основні напрямки ліній. Вибір кількості схованих нейронів був зроблений протягом навчальної стадії і змінювався від 5 до 20. Нейрони схованого шару мають сигмовидну функцію передачі. Кожен нейрон у шарі зв'язаний із усіма нейронами суміжного шару.

Карта підпису є новим простором ознак вихідного статичного бінарного підпису. [7]

5.ВИСНОВКИ

Комбінація різноманітних варіантів простору ознак перетворення Радону в сполученні з вейвлет декомпозицією, інваріантними моментами, псевдофазовою системою координат дозволило підвищити точність та вірогідність класифікації підписів нейромережним методом за рахунок поділу просторів ознак на «статичний», що характеризує загальні параметри підпису як єдиного цілого, і «динамічний» простір, що характеризує локальні особливості. Надійність класифікації підписів при використанні запропонованого методу склала 93-98%.

ЛІТЕРАТУРА

1. G. Pirlo. Algorithms for Signature Verification / ed. S.Impedovo, Springer Verlag, Berlin, 1994, pp 433-454.
2. Ахметшин А.М. , Трипольская В.В., "Идентификация подписей в сингулярных базисах преобразования Радона ", Вестник херсонского государственного технического университета №12, 22-27с, 2001г.
3. S.Belkasim, M.Shridhar, M.Ahmadi, "Pattern Recognition with Moment Invariants: A Comparative Study and New Results", Pattern Recognition, Vol.24, No. 12, pp. 1117-1138, 1991.
4. M.K.Hu, "Visual Pattern Recognition By Moment Invariants", IEEE Trans. on Information Theory, Vol.IT-8, pp.179-187, Feb.1962
5. Q.M Tieng, W.W. Boles, "Recognition of 2D Object Contours Using the Wavelet Transform Zero-Crossing Representation", IEEE Trans. on PAMI, Vol.19, No.8, pp. 910-916, 1997.
6. Handbok of Character Recognition and Document Image Analysis. Editors H.Bunke, P.S.P. Wang, London, World Scientific, 1997
7. Трипольская В.В., Ахметшин А.М. «Нейросетевая идентификация подписей в пространстве признаков преобразования Радона », Проблемы программирования. №1-2, 2002г, с.576-582
8. Ахметшин А.М., Трипольская В.В. «Повышение точности нейросетевой классификации подписей в пространстве признаков вейвлет декомпозиции преобразования Радона», Міжнародна конференція з індуктивного моделювання 2002, Львов, т.3 , с. 196-202
9. M.R.Teaque, "Image analysis via the general theory of moments",J.Opt.Soc.Am., Vol.70, No.8, August 1980
10. A.Khotanzad, Y.Poggios, "Invariant Image Recognition by Zernike Moments", IEEE Trans. on PAMI, Vol. PAMI-12, 1990, pp. 489-497
11. В.И. Арнольд. Математические методы классической механики. М., Наука, 1989. – 472с.
12. Ахметшин А.М., Иванова Н.А. «Качественная интерпритация результатов метода ультрозвуковой спектральной дефектоскопии по интегральным показателям в фазовой системе координат» - Дефектоскопия, 1996, №12, с.36-45

ПРИНЦИПИ ПОБУДОВИ ДІАГНОСТИЧНИХ СИСТЕМ ДЛЯ АЕС НА ОСНОВІ НОВИХ ІНФОРМАЦІЙНИХ ТЕХНОЛОГІЙ

В.О. Болтенков, М.В. Максимов, О.В. Маслов

Одеський національний політехнічний університет
65044 Одеса, просп. Шевченка, 1
Тел.: +380 482 288-440 Факс: +380 482 344-308
Електронна пошта: Galchenkov@od.goldentele.com

Abstract

The advanced diagnostic systems for NPP are discussed: leak diagnostic system and fuel diagnostic system. Leak diagnostic is based on two different principles — measuring activity of air surrounding place of leak (early leak detection) and detection of leak produced sound with distributed self-organizing net of microphone sensors (localization of leak place). Fuel burnup diagnostic is made by means of gamma-specrtometry with application of CdZnTe detectors. New technology to estimate fuel burnup in real time is proposed. The prospect of NPP diagnostic systems perfection is discussed.

1. Вступ

Ядерна енергетика є домінуючою енергогенеруючою технологією в Україні. Якщо у світовому енергетичному балансі доля енергії, що видобута АЕС, дорівнює в середньому 17%, в Україні цей показник становить біля 50%. У зв'язку з цим експлуатація АЕС потребує постійного наукового супроводу. Сьогодні в країні експлуатується 13 енергоблоків типу ВВЕР. Усі вони потребують модернізації, дообладнання новими системами та усунення слабких місць, що виявили себе в процесі експлуатації. Проектними рішеннями на енергоблоках не були передбачені системи оперативної діагностики деяких видів пошкоджень, а саме течій в обладнанні АЕС. В той же час експлуатація демонструє, що саме течії є однією з причин позапланових зупинок блоків.

Збитки економіки держави від вимушених зупинок енергоблоків АЕС та їх роботи на понижених потужностях за компетентними оцінками в декілька разів перевищують витрати на реконструкцію та модернізацію енергоблоків, що дозволяє довести рівень їх експлуатації до рівня західних АЕС. Такі оцінки підтверджують економічну доцільність реконструкції та модернізації українських енергоблоків.

Іншою важливою проблемою для АЕС України є контроль ядерного палива (ЯП) на всьому протязі ядерного циклу. Це питання стає особливо актуальним в умовах організації самостійного збереження відпрацьованого палива в сухих сховищах на території України. У зв'язку с цим потрібні системи оперативного контролю та діагностики ЯП на протязі всього його життєвого циклу.

Керуючись викладеним вище, можна зробити висновок про велику актуальність розробки та впровадження на АЕС систем діагностики течій в обладнанні та оперативної діагностики ядерного палива. У статті викладено наукові результати, що отримані на протязі останніх років у лабораторії "Атомспецавтоматика" Одеського національного політехнічного університету в галузі дослідження принципів побудови та практичної реалізації систем діагностики для АЕС України та інших країн, а також перспектив їхнього удосконалення шляхом застосування прогресивних алгоритмів обробки сигналів та новітніх інформаційних технологій .

2. Система діагностики течій.

2.1. Загальні положення

Діагностика течій в обладнанні АЕС базується на концепції "Течія перед руйнуванням" (LBB — Leak Before Break") [1]. Згідно з цією концепцією течія спричиняє подальше руйнування, яке може мати наслідки катастрофічного масштабу. Отже течія має бути визначеною на ранніх стадіях прояву. В першому контурі реактору ВВЕР-1000 теплоносій знаходиться під тиском до 16 МПа при температурі до 320°С. Виток теплоносія може бути визначений за різними фізичними полями, що його супроводжують. Відомі системи, що визначають факт течії шляхом вимірювання вологості [2], шляхом візуального нагляду за місцем потенціального витоку за допомогою телевізійних камер [3], за підвищенням гамма-активності повітря в зоні контролю за рахунок виділення продуктів активації теплоносія [4], та також системи, що фіксують акустичні сигнали, якими супроводжується виток. Системи діагностики за акустичними сигналами поділяються на два класи: контактні, системи, тобто такі, де акустичні сенсори виявляють

широкосмугові акустичні коливання в діапазоні частот (100…400) кГц, що виникають при течії та поширюються в матеріалі трубопроводу [5], та безконтактні, що визначають акустичні коливання, які поширюються в повітрі навколо течії [6].

Вологомірні системи мають досить високу чутливість (дають змогу виявити течію з витратою близько 70 літр/годину), але не дають змоги визначити координати місця течії. Окрім того обладнання для відбору проб повітря для систем такого класу є дуже коштовним і складним у монтажі. Системи візуального нагляду фактично потребують постійної участі оператора, оскільки на сьогодні вони не мають реальних перспектив автоматизації. Системи контролю активності високочутливі (їхня чутливість складає (3…60) літр/годину), але теж не дають змоги оцінити місце течії.

Щодо акустичних систем діагностики течій, слід зазначити таке. Контактні системи мають середню чутливість — (60…300) літр/годину — але для визначення місця течії з потрібною похибкою потребують розміщення акустичних сенсорів практично на кожному з трубопроводів, що контролюються. Отже для одного об'єкту контролю число сенсорів (і відповідно ліній зв'язку) може сягати десятків-сотен. При високій коштовності системи (ціна її при такій конфігурації може становити мільйони євро [1]) вона повністю втрачає свою експлуатаційну надійність, а під час планового ремонту руйнується практично без змоги відновлювання. Отже найбільш перспективними вдаються акустичні системи безконтактного типу, тим більше, що система такого класу була успішно впроваджена на реакторі РБМК на Ленінградській АЕС [6].

Конкретна система діагностики течій розроблялась для верхнього блоку реактора ВВЕР-1000. Верхній блок налічує 92 патрубки із фланцевими з'єднаннями, в яких течії теплоносія спостерігаються досить часто [4] і, як правило, мають мають досить тяжкі наслідки. Згідно з міжнародними вимогами до діагностичних комплексів АЕС система повинна бути комплексною, а саме буди побудованою не менш як на двох фізичних принципах. Спираючись на приведений вище аналіз, було обрано структуру діагностичного комплексу такою — виявляються та реєструються активність повітря в коробах вентиляційної системи верхнього блоку реактору й акустичні сигнали в повітрі, що оточує зону контролю.

2.2. Підсистема радіаційного контролю

Підсистема радіаційного контролю (ПРК) має призначенням раннє виявлення течії. Вона складається з 6 CdZnTe гамма-детекторів, що розміщуються в шести коробах вентиляційної системи повітряного охолодження приводів керування реактором. Кожний детектор являє собою інтелектуальний сенсор, який окрім власне детектора включає підсилювач заряду та цифровий багатоканальний аналізатор, реалізований за допомогою швидкого АЦП та 32-розрядного сигнального процесора. У випадку течії теплоносія спостерігається стабільний пік гамма-активності продуктів активації у районі 511 кеВ. Обробка спектрів гамма-випромінювання, зареєстрованого гамма-детектором детально викладена в [7].

2.3. Підсистема акустичного контролю

Для обгрунтування принципів побудови підсистеми акустичного контролю (ПАК) та технології обробки інформації були проведені попередні дослідження на спеціальних стендах. Стенди використовувалися для імітації течій при робочих параметрах теплоносія крізь дефектні деталі реального обладнання АЕС. Метою експерименту була оцінка спектрально-кореляційних параметрів акустичних сигналів, що утворюються в повітряному середовищі навколо витоку, та перевірка принципів оцінювання координат витоку. Для реєстрації акустичного поля течії використовувалися широкосмугові мікрофонні сенсори. Дослідження показали, що сигнали від течій є досить широкосмугові, шумоподібні і лежать у діапазоні (10…40000) Гц. Конкретні особливості спектру сигналів (нерівномірність, уні- або мультимодальність та ін.) сильно залежать від характеру течії, яка у свою чергу визначається розміром та типом дефекту. Попередні дослідження переконали в тому, що для виявлення течії та оцінювання її координат потрібно використовувати адаптивні алгоритми.

Основною завадою при виявленні акустичних сигналів, що їх породжує течія, є технологічні шуми в приміщеннях АЕС, а саме в шахті реактору. Дослідження шумів показало, що це є типові сигнали від роботи різних механізмів і їх спектр має типовий характер — набір гармонійних компонент на фоні злитого спектра, можна вважати, що за граничною частотою 2 кГц технологічні шуми практично відсутні.

Конфігурація ПАК була вибрана такою. Акустичні сенсори в кількості 24 одиниці розміщуються на стінках циліндричної шахти реактора чотирма горизонтальними кільцями по 6 сенсорів у кожному з вертикальним рознесенням кілець приблизно на 0.8 м. Отже чотири кільця мікрофонних сенсорів повністю обіймають зону контролю, а на горизонтальному рівні знаходяться приблизно навпроти можливих течій.

Оскільки вся зона контролю повністю заповнена патрубками, умови розповсюдження сигналів від місця течії до сенсора досить складні: мають місце віддзеркалення як від патрубків, так і від бетонних стінок шахти. Це також потребує застосування спеціальних алгоритмів обробки сигналів.

2.4. Алгоритми функціонування системи діагностики течій

До алгоритму діагностики пред'являються такі вимоги:

— визначення факту течії на якомога ранній ступіні її прояву,
— локалізація течії з максимально можливою точністю,
— оцінка масової або об'ємної витрати теплоносія крізь течію,
— прогноз розвитку процесу.

Алгоритм побудовано двохетапним. На першому етапі визначається факт течії за допомогою ПРК, при цьому ПАК функціонує в черговому режимі: усі акустичні сенсори працюють, але їхня інформація консервується в базі пасивного моніторингу і не приймає участі в процесі обробки. Якщо в двох із шести каналів радіаційного контролю рівень пику в районі 511 кеВ перевищує порогове значення, факт течії вважається установленим, і аларм-сигнал поступає на термінал оператора системи і на запуск ПАК для локалізації течії.

Для локалізації течії побудовано адаптивний алгоритм, що самоорганізується. Його основні кроки:

1) З 24 акустичних сенсорів вибираються 4, на яких рівень звукового тиску найбільший. Ці сенсори є найближчими до місця течії. Далі в обробці беруть участь тільки їхні сигнали.

2) Оскільки акустичні сигнали в повітрі мають сильне частотно-залежне ослаблення, а шлях їх розповсюдження до сенсора може бути складним за рахунок багатократних віддзеркалень, неможливо прогнозувати в якому діапазоні частот організовувати локалізацію. Для визначення робочого діапазону для кожної пари з 4-х сенсорів обчислюється функція частотної когерентності. В якості робочого діапазону вибирається та 1/3 октавна смуга, де когерентність максимальна.

3)Для кожної пари сенсорів обчислюється кореляційна функція, відшукується її максимум і таким чином оцінюється різниця часів приходу (РЧП) (TDOA — time difference of arrival) згідно з принципом узагальненого корелятора Кнеппа-Картера.

4) Для кожної трійки (дві пари) сенсорів на основі оцінених РЧП складається система з двох гіперболічних рівнянь, яка чисельно вирішується й знаходяться оцінки координат течії та дисперсія оцінки.

5) З шести отриманих таким чином оцінок координат формується середньо-взважена оцінка, яка передається на термінал оператора системи.

При використанні двох комп'ютерів (для збору даних та для обчислення згідно з алгоритмами) усі процедури легко реалізуються в реальному часі.

Викладена технологія дозволяє оцінити координату течії з потенційною точністю (0,3…0,6) м. Однак попередні іспити довели, що реальна похибка може бути на (50…60)% більшою. Це пов'язане з втратою просторової когерентності акустичного сигналу за рахунок віддзеркалень з одного боку ті сильно частотно-залежного ослаблення з іншого боку.

З метою поліпшення показників точності оцінювання координат течії зараз вивчається кілька алгоритмів підвищення просторової когерентності акустичних сигналів в умовах замкнутого приміщення. Це такі алгоритми як кепстральна деконволюція, яка дає змогу послабити в результуючому сигналі, що приймається сенсором багаторазових дзеркальних реплік, а також технологія вейвлет-аналізу за схемою вейвлет-трансформація — кліппування — зворотна вейвлет-трансформація. З точки зору підвищення імовірності правильного виявлення сигналу від течії досліджується застосування методів визначення розладки за алгоритмом кумулятивних сум.

3. Система діагностики ядерного палива

3.1. Технологія оцінювання стану ЯП

Діагностика ЯП є важливою умовою забезпечення ядерної безпеки при експлуатації АЕС. Організація сухих сховищ відпрацьованого ядерного палива на українських АЕС гостро потребує створення засобів експресного інструментального контролю ЯП безпосередньо перед його перевантаженням або відвантаженням на сухе збереження. Основним параметром, що характеризує стан ЯП і критерієм можливості, його переміщення на сухе збереження, є глибина його вигоряння. Для виміру вигоряння ЯП авторами розроблена система, заснована на реєстрації власного гамма-випромінювання тепловиділяючої зборки, що вивантажується з активної зони реактора, (ТВЗ). Система побудована за принципом об'єднання в часі двох технологічних операцій — перевантаження палива і вимірювання його власного гамма-випромінювання. Такий принцип дозволяє істотно зменшити часові витрати й організувати функціонування системи в реальному масштабі часу. Реальний масштаб часу вимагає створення нових інформаційних технологій оцінки глибини вигоряння, що забезпечують мінімальні обчислювальні витрати на обробку зареєстрованого гамма-випромінювання і високу вірогідність контролю. Пропонована технологія базується на аналізі гамма-спектрів випромінювання продуктів поділу (ПП) ЯП, що містяться у відпрацьованій ТВЗ.

Зареєстроване гамма-випромінювання із системи CdZnTe детекторів, перетворене в цифрову форму, надходить на блок формування гамма-спектрів і їхньої обробки. Обробка спектрів є досить добре відпрацьованою процедурою в гамма-спектрометрії. Результатом її є виділені піки повного

поглинання (ППП) ізотопів - ПП і їхні параметри. На основі обробки численних реальних даних, зареєстрованих у виробничих умовах, авторами встановлено ряд емпіричних залежностей, що зв'язують вигоряння ЯП з обмірюваною активністю окремих ПП [8]. Далі алгоритм оцінювання вигоряння є розгалужений, причому критерієм вибору робочої галузі є якість виділених піків, що спостерігаються в гамма-спектрі, а саме їхня статистична переконливість. При відсутності в гамма-спектрі добре ідентифікуємих ППП здійснюється аналіз наявності піка, пов'язаного з підсумовуванням гамма-випромінювання ПП ^{134}Cs ^{95}Zr, ^{95}Nb. У випадку позитивних результатів (тобто якщо наявність ^{95}Zr, ^{95}Nb підтверджується) робиться висновок про недостатню витримку відпрацьованої ТВЗ і неприпустимості її відвантаження в сухе сховище. Визначається час витримки відпрацьованої ТВЗ із використанням відношення обмірюваної інтенсивності γ-випромінювання ^{137}Cs до інтегральної інтенсивності γ-випромінювання. Якщо отримана оцінка значення часу витримки складає менш 3 років, відвантаження відпрацьованої ТВЗ у сховище також забороняється. Потім проводиться аналіз обмірюваних спектрів на предмет наявності чітких ППП довго живучих ізотопів. При наявності тільки чіткого піка ^{137}Cs оцінюється час витримки ЯП, і здійснюється розрахунок його вигоряння з використанням відносини обмірюваної інтенсивності γ-випромінювання ^{137}Cs до інтегральної інтенсивності γ-випромінювання. При наявності добре ідентифікуємих піків ^{137}Cs, ^{134}Cs, ^{106}Ru, ^{154}Eu вигоряння розраховується з урахуванням часу витримки і початкового збагачення. Цей алгоритм є основним при визначенні вигоряння відпрацьованих ТВЗ, що відвантажуються на сухе збереження.

3.2. Перспективні напрямки розвитку діагностичних технологій для ЯП

Було побудовано модель утворення поля гамма-випромінювання ТВЗ з урахуванням того, що вона складається з 312 тепловиділяючих елементів (твелів) [9]. Оцінка розподілу гамма-активності ПП між окремими твелами також являє великий практичний інтерес. Зокрема, знання вигоряння ЯП важливе при компоновці активної зони реактора. Нульова активність деяких ізотопів-ПП в окремих твелах може бути ознакою того, що герметичність твела порушена. Отже можна пропонувати новий спосіб контролю герметичності оболонки (КГО) твела. Відновлення активності твелів за вимірюванням загального поля гамма-випромінювання ТВЗ є зворотною задачею гамма-спектрометрії. Її вирішення можливе як методами комп'ютерної томографії [10], так і шляхом вирішення перевизначеної системи лінійних рівнянь методом Мура-Пенроуза. В даний час ці два методи вирішення зворотної задачі досліджуються на обумовленість.

4. Висновки

Запропоновані в доповіді нові технології сумісної обробки сигналів різної фізичної природи дозволили побудувати нові високоефективні системи діагностики течій та палива для АЕС і таким чином підвищити економічність та безпеку їх експлуатації.

Література

1. Гетман А.Ф. Концепция безопасности "Течь перед разрушением" для сосудов и трубопроводов АЭС. М.: Энергоиздат, 1999. — 262 с.
2. Петровский Б.С., Аксенов В.И., Давиденко Н.Н., Карякин Ю.Е. Компьютеризированная система раннего обнаружения протечек теплоносителя на верхнем блоке реактора ВВЭР-1000.// Вторая Международная конференция "Безопасность, эффективность и экономика ядерной энергетики". Ч.1. М., 2001. С. 170-172.
3. Усанов А.И. Состояние и проблемы внедрения систем оперативной диагностики реакторных установок ВВЭР.// Труды научно-технической конференции концерна "Росэнергоатом". М., 2000. С. 64-71.
4. Болтенков В.А., Верпета В.И.,Калашников А.Н., Максимов М.В., Маслов О.В. Диагностика протечек теплоносителя на верхнем блоке ВВЭР-1000: проблема и пути решения. Атомна енергетика та промисловість України, №2(2),1999. С.51-53.
5. Kunze U. Acoustic leak monitoring in VVER plants: ten year of experience, Nuclear Europe Worldscan, №7-8 (1999). P.114-116.
6. H. Mochizuki, H. Takeda, Y. Kasai et al. Development of leak detection system for piping using high-temperature resistant microphones. In *Proceedings of 8th International Conference on Nuclear Engineering (ICONE'8)*, Baltimore, 2-6 April 2000, ASME, Baltimore, USA 2000.P.8526-8551.
7. 7. Маслов О.В., Сухарев К.В., Кальнев Л.Л., Галченков О.Н. Цифровой спектрометр гамма-излучения.// 4-е международное совещание «Проблемы прикладной спектрометрии и радиометрии - 2000». Одесса, 2000. С. 40.
8. Билей Д.В, Маслов О.В., Фридман Н.А., Максимов М.В. Методика, алгоритм и критерии определения выгорания ОЯТ в процессе перегрузки // Ядерная и радиационная безопасность — 2001. — Т. 4, вып. 3. — С.38 - 46.
9. Олейник С.Г., Болтенков В.А. Имитационное моделирование системы определения глубины выгорания облученного ядерного топлива // Изв. вузов. Ядерная энергетика, №1, 2002. С.39-52.
10. S. Jacobsson. Theoretical Investigations of Tomographic Methods used for Determination of the Integrity of Spent BWR Nuclear Fuel. — Uppsala University, Internal Report ISV-6/97, 1996. — P. 59.

АВТОМАТИЗОВАНА СИСТЕМА ІДЕНТИФІКАЦІЇ ТА КЛАСИФІКАЦІЇ НОМЕРНИХ ЗНАКІВ

С. Г. Антощук, В. М. Крилов, В. О. Давидов
Одеський національний політехнічний університет
м. Одеса, пр. Шевченко 1. Факс (0482)344-308
E-mail: tom2k@mail.od.ua

The structure of automatic system for vehicle numbers identification and classification are designed. The system is consisting of previous processing and classification modules. The hard wear of system is consisting of database, TV- sensor, monitor, video plate, image catch module.

The high noise level is main factor, which districts installation of real time signal or image recognition automatic systems. The previous processing module has procedures: vehicle number search from vehicle image; filtering; signal/noise ratio improving (denoising) by Haar domain and threshold processing; threshold binarization; correlation symbols segmentation; detected number image normalization. The noise influence decreases by linear filtering also.

The classification module is neuronal net. The neuronal net choice is discussed. Automatic system has high noisy stability, shift, scale, and rotation invariants.

The right recognition probability for symbol is 0,96. Also the possible ways indicates of their decision.

ВСТУП

Автоматизовані системи (АС) ідентифікації та класифікації номерних знаків автотранспортних засобів можуть знайти застосування при контролі транспортних потоків на контрольно-пропускних пунктах ДАІ, при необхідності реєстрації транспортних засобів на автомобільних стоянках і гаражах, а також при аналізі аварійних ситуацій: проїзд на червоне світло світлофора, перевищення швидкості і т.п..

Існуючі аналоги не дозволяють цілком автоматизувати розпізнавання й ідентифікацію транспортних засобів, мають високу вартість і непридатні для масового використання.

Проведений аналіз виявляє ряд основних вимог до проектованої АС: інваріантість до зміни в деяких межах кута повороту між віссю камери і вектором напрямку руху автомобіля як у вертикальній, так і в горизонтальній площинах; інваріантість до зміни масштабу номеру, тобто відстань між камерою й автомобілем у момент захоплення зображення не повинна бути строго фіксованою; інваріантість до перекручування номерного знака, тому що при установці номерного знака на автомобілі не у строго вертикальній площині і спостереженні його збоку знак являє собою не прямокутник, а паралелограм, у цьому випадку символи на зображенні спотворюються, система повинна зберігати працездатність в умовах сильних перешкод, наприклад, дощ, сніг, бруд на номерному знаку, освітлення в залежності від часу доби можуть привести до появи адитивних, мультиплікативних, імпульсних і флюктуаційних перешкод; номерні знаки повинні відповідати державному стандарту [1].

СТРУКТУРА СИСТЕМИ

Систему розпізнавання автомобільних номерів можна розглядати як програмно-технічний комплекс, який вміщує апаратну частину та програмне забезпечення (ПЗ). Апаратна частина АС уявляє собою комплекс стандартних комплектуючих. Відеокамера, підключається до персонального комп'ютера за допомогою відеоплати, що має відповідний вхід. Зображення, що фіксується камерою, виводиться на монітор.

ПЗ даної системи можна розділити на чотири основні частини:

— Модуль попередньої обробки — призначений для пошуку місця розташування області номерного знака (ОНЗ), виділення ОНЗ, сегментації і підготовки даних для розпізнавання.

— Модуль класифікації — безпосередньо робить розпізнавання сегментованого зображення номерного знака за допомогою нейроної мережі (НМ).

— Модуль бази даних — дозволяє зберігати інформацію, що надходить, з можливістю її наступного аналізу.

МОДУЛЬ ПОПЕРЕДНЬОЇ ОБРОБКИ

Модуль попередньої обробки складається з наступних процедур:

— *Пошук місця розташування ОНЗ.* Задачею процедури є виділення ОНЗ, що значно зменшує об'єм графічної інформації. Це вирішується в області двовимірного хвильового перетворення Хаара. У ході перетворення вихідне зображення розбивається на клітки 2x2, для яких виконується:

$$\begin{bmatrix} a_0 & a_1 \\ a_2 & a_3 \end{bmatrix} \rightarrow \begin{bmatrix} \frac{a_0+a_1+a_2+a_3}{2} & \frac{a_0+a_1-a_2-a_3}{2} \\ \frac{a_0-a_1+a_2-a_3}{2} & \frac{a_0-a_1-a_2+a_3}{2} \end{bmatrix} = \begin{bmatrix} b_0 & b_1 \\ b_2 & b_3 \end{bmatrix}$$

Елементи *b0, b1, b2 і b3* формують чотири матриці: *b0* являє собою середнє арифметичне у вікні 2x2, у *b1* підкреслюються горизонтальні контурні перепади яскравості; у *b2* - вертикальні, а в *b3* - діагональні. Розмір кожної матриці в два рази менше за розмір вихідного зображення. З точки зору вирішення поставленої задачі найбільший інтерес уявляє матриця *b2*, яка в подальшому оброблюється.

— *Бінаризація.* Після хвильового перетворення виходить зображення з підкресленими контурними перепадами, які орієнтовані вертикально. В ОНЗ кількість таких перепадів більш ніж у інших частинах зображення, що приводить до появи області прямокутної форми, в який спостерігаються різкі флуктуації інтенсивностей пікселей.

Бінаризація (мал.1) здійснюється пороговим методом, при якому запропоновано оцінювати дисперсію *D(x,y)* сигналу в деякій області щодо поточного пикселя зображення *I(x,y)*. Оцінка дисперсії *D(x,y)* порівнюється з деяким порогом *P*:

$$I_{вих}(x,y) = \begin{cases} 1, D(x,y) \geq P \\ 0, D(x,y) < P \end{cases},$$ де: $I_{вих}(x,y)$ — інтенсивність пікселя у бінаризованому зображенні.

Мал.1. Бінаризоване зображення

— Фільтрація. Задачею фільтрації є знищення як можна більшого числа об'єктів на бінаризованому зображенні, які не входять до ОНЗ.

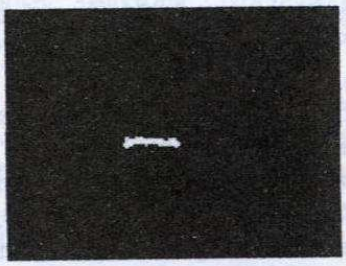
Процедура фільтрації складається з двох частин:

а) стандартна медіанна фільтрація у вікні 3x3 пікселя:

$$B = \begin{cases} 1, \sum_{i=1}^{N_s} I_i > 4 \\ 0, \quad інакше \end{cases}$$

де I_i, B — значення поточного пікселя бінаризованого зображення до та після медіанної обробки.

б) простежування злитних груп, визначення їх параметрів і аналіз на відповідність щодо віднесення до ОНЗ (мал. 2).



Мал.2. Зображення після фільтрації

Для кожної простеженої злитної групи визначаються висота, ширина, площа і відношення висоти до ширини. Об'єкти, які не задовольняють критеріям ОНЗ знищуються.

— *Виділення номерного знака.* Для виділення ОНЗ використовується метод кореляційно-екстремального аналізу [2]. Суть методу полягає в плоскопараллельному зсуві маски М по зображенню В з метою розрахунку кореляційної функції:

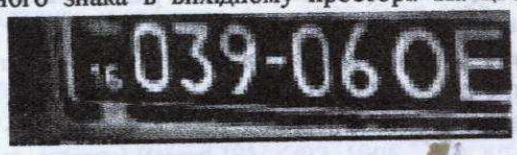
$$K(x,y) = \sum_{i=1}^{L_x}\sum_{j=1}^{L_y} B(x+i, y+j) \cdot M_{ij}.$$

де *(x,y)* — координати поточого пікселя зображення; L_x — розмір маски по горизонталі; L_y — розмір маски по вертикалі; *(i,j)* — координати пікселя у середині маски.

Взаємне положення маски та ОНЗ, при якому значення кореляційної функції максимально, відповідає реальному положенню ОНЗ.

— *Визначення типу зображення:* умовно вважаємо позитивом світлі символи на темному фоні. Дана процедура передбачає гистограмний аналіз зображення:

— *Поліпшення якості та бінарізація* області номерного знака в вихідному просторі. На цьому



Мал.3. Зображення ОНЗ

етапі виконуєть смугова фільтрація, контрастування та порогова бінарізація. Типове зображення ОНЗ після смугою фільтрації представлене на мал.3.

— *Сегментація*. Для сегментації зображення ОНЗ на окремі символи застосувався раніш згаданий метод простежування злитних груп.

— *Трансформація сегментів*. Здійснюється трансформація виділених сегментів до єдиних розмірів, що забезпечує вимоги інваріантісті системи до масштабу.

МОДУЛЬ КЛАСИФІКАЦІЇ

Найбільш перспективним методом класифікації зображень символів представляється використання нейронної мережі, перевагою якої є можливість класифікації символів по неповному їхньому зображенню. Подібно тому як людина по обривках інформації про той чи інший об'єкт може, використовуючи асоціації, відновити повну картину, так і НМ може по частці вхідного образа відновити весь образ. Прикладом такої мережі є НМ Хопфилда, що реалізує так називану асоціативну пам'ять.

Дослідження нейроного класифікатора на основі мережі Хопфилда виявило три істотних недоліки, що роблять неможливим його використання для вирішення поставленої задачі:

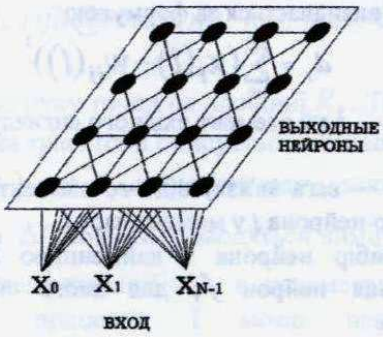
— НМ дуже чуттєва до перекручувань і зсувів зображень символів у полі зору.

— Рекурентність НМ Хопфилда приводить до того, що на "згадування" символу, який сильно відрізняється від еталонного, потрібно занадто багато часу.

— Якщо пропонований на вході НМ символ сильно відрізняється від еталонного, на виході може з'явитися так називаний "фантом" - символ не є жодним з еталонів. Як правило, це може бути інвертоване зображення еталона або сукупність декількох еталонів накладених один на одного.

У ході експериментування з різними моделями НМ було вирішено використовувати НМ Кохонена [3] відому також за назвою карти Кохонена, що самоорганізуються (Kohonen`s Self Organizing Map, SOM). Модель SOM була запропонована Кохоненом у 1984 році і до дійсного часу існує безліч її модифікацій з багатою математичною теорією навколо них. Суть моделі полягає в наступному.

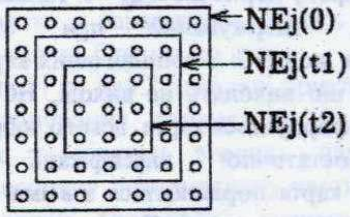
У мозку нейроні розташовуються у визначеному порядку так, що деякі зовнішні фізичні впливи викликають відповідну реакцію нейронів з визначеної області мозку. Наприклад, у тій частині мозку, що відповідає за сприйняття звукових сигналів, нейрони групуються відповідно до частот вхідного сигналу, на яких вони резонують. Хоча структура мозку в значній мірі визначається генетично, окремі структури мозку формуються в процесі самоорганізації. Алгоритм Кохонена до деякої міри нагадує процеси, що відбуваються в мозку.



Мал.4. Мережа Кохонена

Алгоритм Кохонена дає можливість будувати НМ для поділу векторів вхідних сигналів на підгрупи. Мережа складається з M нейронів, що утворять прямокутні грати на площині мал.4. Елементи вхідних сигналів подаються на входи всіх нейронів сітки. У процесі роботи алгоритму набудовуються синаптичні ваги нейронів.

Вхідні сигнали - вектора дійсних чисел - послідовно пред'являються НМ. Бажані вихідні сигнали не визначаються. Після того, як було пред'явлено достатнє число вхідних векторів, синаптичні ваги визначають кластери. Крім того,



Мал. 5. Зони топологічного сосідства на карті ознак в різні моменти часу

ваги організуються так, що топологічно близькі вузли чуттєві до схожих зовнішніх впливів (вхідним сигналам).

Для реалізації алгоритму необхідно визначити міру сусідства нейронів (міру близькості). На мал.5 показані зони топологічного сусідства нейронів на карті ознак у різні моменти часу. $NE_j(t)$ - безліч нейронів, що вважаються сусідами нейрона j у момент часу t. Зони сусідства зменшуються з часом.

Алгоритм Кохонена формування карт ознак передбачає наступні дії:

1) Ініціалізація мережі. Ваговим коефіцієнтам мережі привласнюються малі випадкові значення. Загальне число синаптичних ваг - $M*N$ мал.4.

Початкова зона сусідства показана на мал.5.

2) Пред'явлення мережі нового вхідного сигналу.

3) Обчислення відстані до всіх нейронів мережі.

Відстані d_j від вхідного сигналу до кожного нейрона j визначається за формулою:

$$d_j = \sum_{i=0}^{N-1} \left(x_i(t) - w_{ij}(t)\right)^2$$

де x_i — *i-ий* елемент вхідного сигналу в момент часу t;

$w_{ij}(t)$ — вага зв'язку від *i-го* елемента вхідного сигналу до нейрона j у момент часу t.

4) Вибір нейрона з найменшою відстанню. Вибирається нейрон j^*, для якого відстань d_j найменше.

5) Настроювання ваг нейрона j^* і його сусідів.

Виробляється підстроювання ваг для нейрона j^* і всіх нейронів з його зони сусідства NE. Нові значення ваг:

$$w_{ij^*}(t+1) = w_{ij}(t) + r(t)(x_i(t) - w_{ij}(t))$$

де $r(t)$ — крок навчання, що зменшується з часом (позитивне число, менше одиниці).

6) Повернення до п.2.

Слід зазначити, що в загальному випадку розмірність SOM може бути довільної.

У ході лабораторних досліджень також було запропоновано використовувати двовимірне стробування вихідного зображення символу. Вікно при стробуванні являє собою квадрат розміром $L_{стр}$, а величина кроку дорівнює $L_{стр}$ - 1. Кожен фрагмент зображення, одержуваний при стробуванні, подається на вхід НС. З топологічних карт окремих фрагментів, що виходять на виході, НС вихідного зображення формується карта всього зображення

Для остаточної класифікації отримана топологічна карта порівнюється з всіма еталонами, що зберігаються у пам'яті. Як рішення вибирається найбільш близький до карти еталон. Критерієм близькості є сума квадратів відхилень крапок реальної карти від еталонної.

ВИСНОВКИ

На мал. 6а,б представлені результати лабораторних досліджень нейроного класифікатора. Вихідним матеріалом служили 330 бінаризованих зображень символів отриманих з реальних фотографій автомобільних номерів. Класифікатору послідовно пред'являлися всі зображення символів і фіксувалося число успішних розпізнавань N. У ході моделювання мінялися розміри вікна стробування $L_{стр}$ і розміри топологічної карти a. Досліджувалися двовимірний і тривимірний варіант SOM. Оптимальними значеннями параметрів нейроного класифікатора було вирішено вважати наступну комбінацію:

число використовуваних вимірів — 2;

розмір вікна стробування $L_{стр}$ — 3;

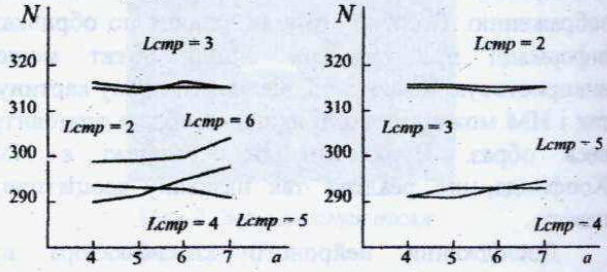
розмір карти a — 4.

Дана комбінація параметрів дозволила досягти імовірність успішного розпізнавання символу в 96% при мінімальній кількості обчислень.

Моделювання нейроного класифікатора без використання стробування тобто коли на вхід НМ подавалося все зображення символу цілком, виявило два істотних недоліки:

— імовірність успішного розпізнавання не перевищувала 66%;

— процес навчання такий НМ займав дуже багато часу (до 20 хв).



Мал. 6. Розподілення імовірності успішного роспізнавання символів при різних параметрах НМ: а — двовимірна SOM; б — тривимірна SOM

З іншого боку, застосування стробування зображення дозволило створити класифікатор на навчання якого потрібно усього кілька десятків секунд.

Таким чином, описана математична модель, дозволяє реально реалізувати АС для розпізнавання автомобільних номерів. Імовірність успішного розпізнавання одного символу складає 96%. Алгоритм роботи АС дозволяє виділяти номерний знак автомобіля незалежно від положення автомобіля в поле зору камери. Структура програмного забезпечення АС дозволяє легко змінювати різні види обробки, що є безсумнівним плюсом з погляду настроювання системи і її удосконалень у ході експлуатації.

ЛІТЕРАТУРА

1. ДСТУ 3650-97. Знаки номерні транспортних засобів.

2. Антощук С.Г., Крылов В.Н., Максимов М.В. Помехоустойчивая сегментация бинаризованных изображений символов // Тр. Одес. политехн. ун-та. — Одесса, 1998. — Вып. 2(6). — С. 111 — 114.

3. Уоссермен Ф. Нейрокомпьютерная техника: Теория и практика. — М.: Мир, 1992.

МЕТОД ГОЛОВНИХ КОМПОНЕНТ У ЗАДАЧАХ АВТЕНТИФІКАЦІЇ

Рифа В.М. Баклан Я.І. Баклан І.В.

Інститут прикладного системного аналізу НАН та МОН України. 03056 Київ 56, вул. Політехнічна 14, ННК ІПСА. Тел. 241 86 59. E-mail: rifa@relos. ntu-kpi. kiev. ua

ABSTRACT

Vasil Rifa, Yaroslav Baklan, Ihor Baklan. Principal component method to the problem authentication. The possibility of application of the principal component method [1] is considered to the problem of authentication of an operator of a complex process control system [2]. The numerical experiments results show the convergence of eigenvalues of covariance matrixes to steady state values on the set of parameters with increasing number of measurements.

ВСТУП

Пропоновані результати досліджень стосуються задачі автентифікації оператора системи керування у випадку, коли дії оператора фіксуються вимірами в замкненій обмеженій підмножині Евклідового простору R_n. Задача автентифікації полягає в підтвердженні даних (пароль, тощо), які оператор повідомляє системі керування про себе для того, щоб мати можливість взаємодії з процесом. Наведений в [1] метод головних компонент застосовується до задач класифікації. Різниця між цими задачами полягає в тому, що коли в задачі класифікації відомі координати простору вимірів, і саме завдяки аналізу цих вимірів проводиться класифікація, то в задачі автентифікації можна відслідкувати лише результати дій оператора; при цьому інформація, хто з операторів керує системою, може бути хибною.

УМОВИ ЕКСПЕРИМЕНТУ

Розглядається випадок, коли результати дій оператора системи керування можна відобразити траєкторією руху точки на площині R_2. Таким чином кожна точка траєкторії описується масивом векторів $\bar{r}_i(x_i, y_i, t_i)$, де t_i - моменти часу, взяті з деяким інтервалом Δt, в яких проводяться виміри (x_i, y_i). Інтервал дискретизації Δt є параметром системи керування процесом і може непередбачено змінюватись самою системою в межах 20 відсотків від найменшого значення.

В загальному випадку траєкторія $\bar{r}$ - це хаотичний рух точки на ділянці площини $(a_x \le x_i \le b_x; a_y \le y_i \le b_y), \forall i = 1 \div N$.

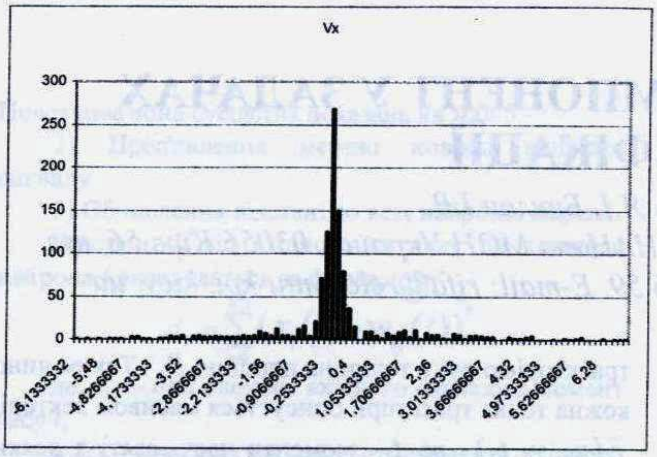
Основа застосування методу головних компонент до задачі автентифікації полягає в тому, щоб на вимірах $\bar{r}_i$ побудувати таку множину F параметрів траєкторії, які б відображували особливості поведінки оператора системи керування. Головною вимогою до параметрів є те, що кожен з них повинен мати гаусівський розподіл. Вибрані параметри наведені в таблиці 1.

Проведені дослідження довели, що гістограми усіх вибраних параметрів мають гаусівський розподіл або дуже близькі до нього.

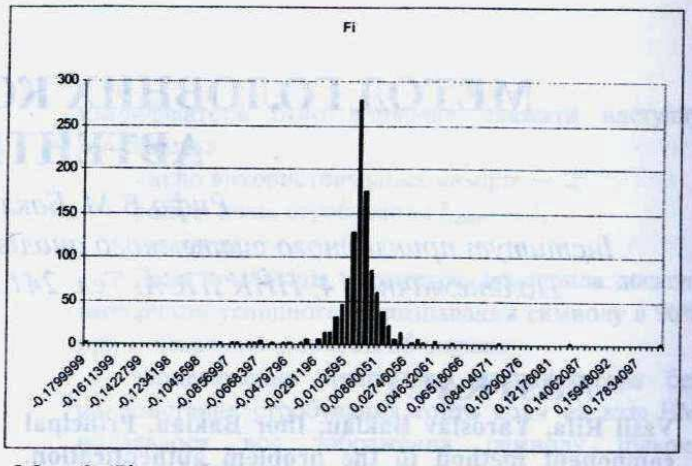
На мал.1 та мал.2 наведені приклади гістограм розподілів вибраних параметрів. Як бачимо, на гістограмах є ділянки значень параметрів, які складають основну інформативну частину, та досить широкі ділянки "шумових" значень параметрів, частота появи яких досить низька. При проведенні експериментів обчислення проводились без відсіювання викидів, на всій множині одержаних значень параметрів.

Таблиця 1.

$v_{xi} = \dfrac{x_{i+1} - x_{i-1}}{t_{i+1} - t_{i-1}}, i = 1 \div (n-1);$	$v_{yi} = \dfrac{y_{i+1} - y_{i-1}}{t_{i+1} - t_{i-1}}, i = 1 \div (n-1);$
$a_{xi} = \dfrac{v_{x(i+1)} - v_{x(i-1)}}{t_{i+1} - t_{i-1}}, i = 2 \div (n-2);$	$a_{yi} = \dfrac{v_{y(i+1)} - v_{y(i-1)}}{t_{i+1} - t_{i-1}}, i = 2 \div (n-2);$
$k_i = \dfrac{v_{xi} a_{yi} - v_{yi} a_{xi}}{\sqrt{\left(v_{xi}^2 + v_{yi}^2\right)^3}}, i = 2 \div (n-2);$	$\varphi_i = k_i \lvert v_i \rvert, i = 2 \div (n-2);$



Мал. 1. Гістограма параметра v_x.



Мал. 2. Гістограма параметра φ.

Значення параметрів в момент t_i складають вектор

$$f_i^* = \{v_{xi}, v_{yi}, a_{xi}, a_{yi}, k_i, \varphi_i\}_i^*, \qquad f_i \in F_6.$$

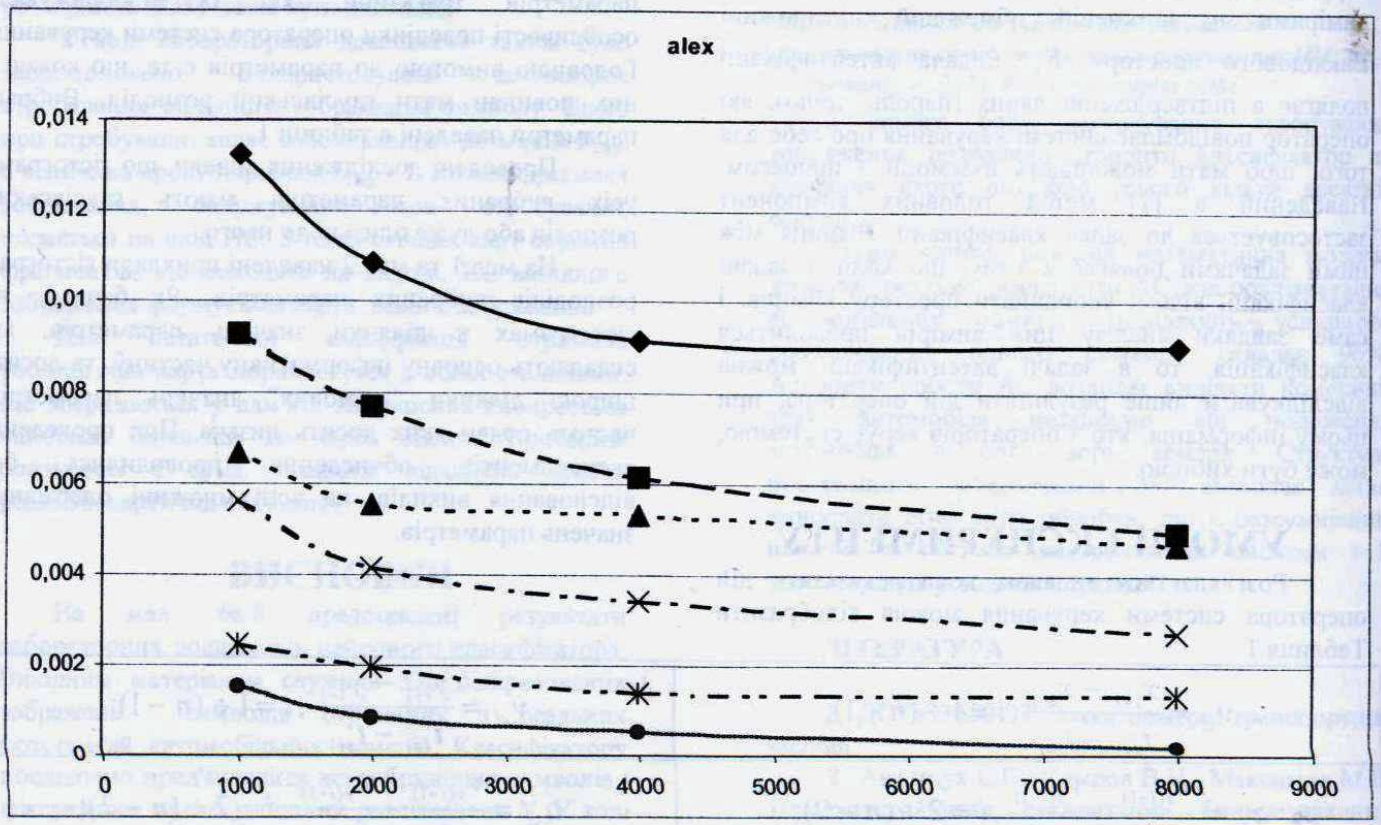
Виконавши умови нормування значень параметрів обчислюємо матрицю коваріацій.

$$\Sigma = \frac{1}{N}\sum_{i=1}^{N}(f_i - \bar{f})(f_i - \bar{f})^*,$$

де $\bar{f}$ - математичне сподівання розподілу вектора $f_i \in F$, а $(\bullet)^*$ - знак транспонування вектора.

РЕЗУЛЬТАТИ

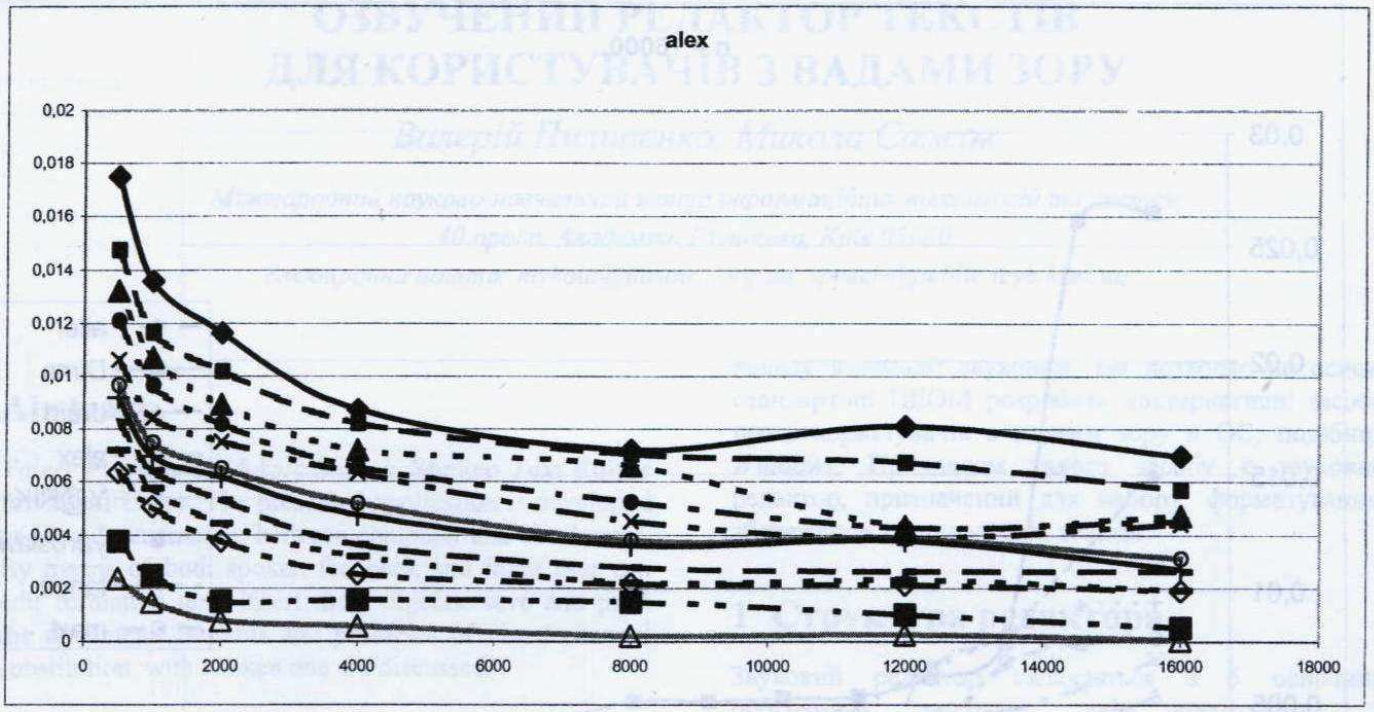
На мал.3 наведені результати чисельних експериментів, за якими можна простежити, що власні числа матриці коваріацій для R_6 збігаються до деяких сталих значень із збільшенням кількості точок вимірів траєкторії, тобто із збільшенням N.



Мал. 3. Збіжність власних значень оператора "alex" в R_6. По горизонталі кількість точок, що взята для проведення експерименту; по вертикалі значення власних чисел матриці коваріацій.

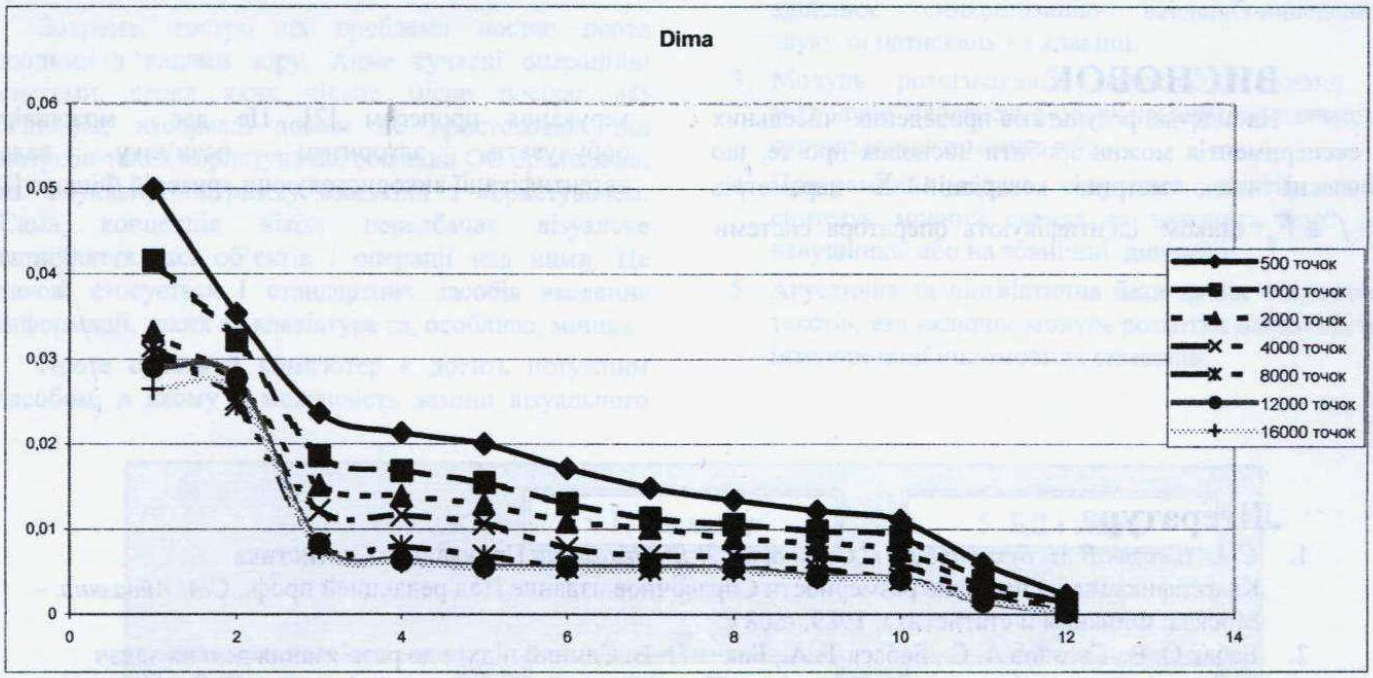
Експерименти також були проведені для дванадцяти параметрів. На рис. 4 наведені приклади збіжності власних чисел матриці коваріацій в R_{12}.

Як видно з графіків, матриці можуть мати декілька кратних власних чисел.



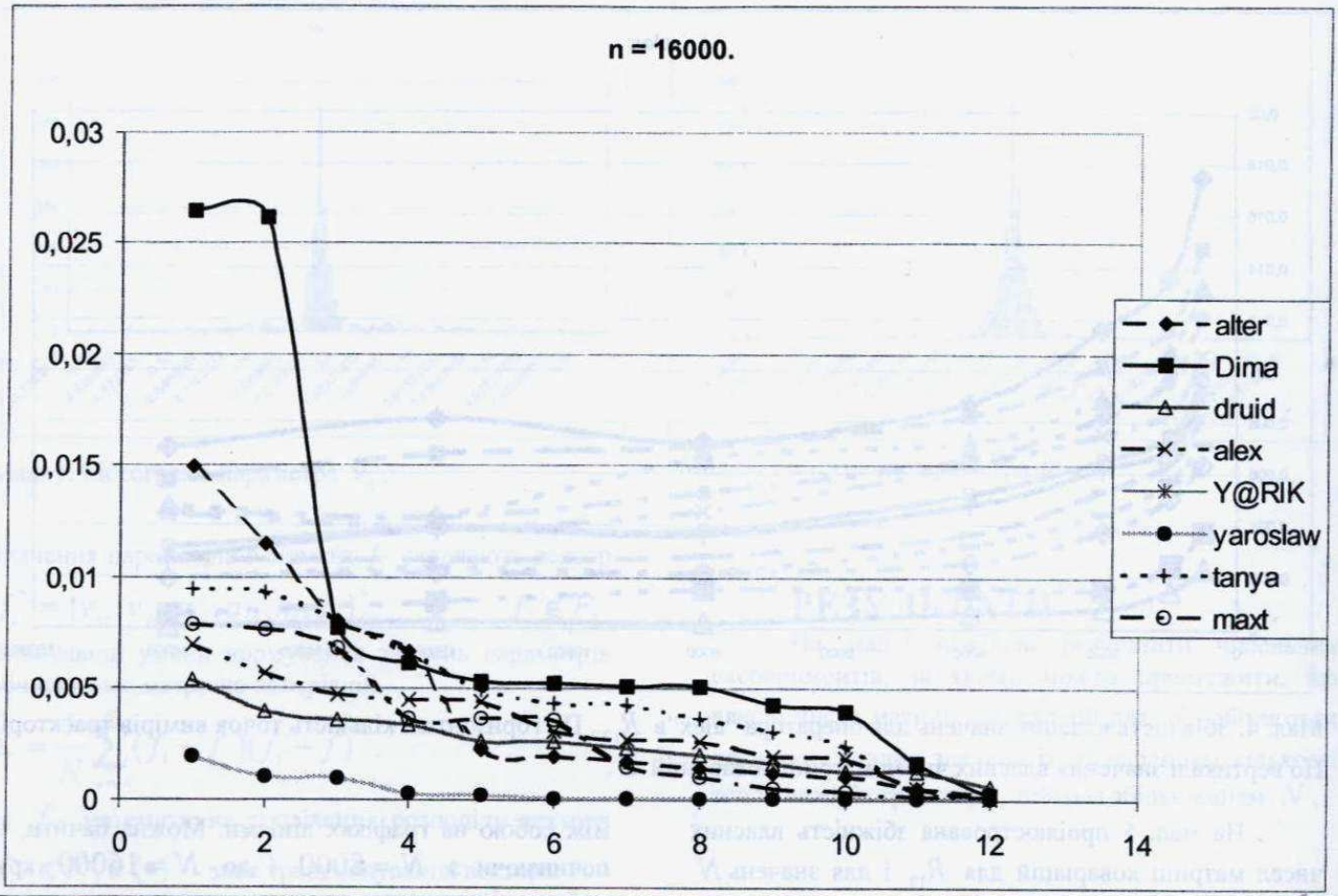
Мал. 4. Збіжність власних значень для оператора "alex"в R_{12}. По горизонталі кількість точок вимірів траєкторії. По вертикалі значення власних чисел матриці коваріацій Σ.

На мал. 5 проілюстрована збіжність власних чисел матриці коваріацій для R_{12} і для значень N від 500 до 16000. Власні числа впорядковані від найбільшого за №1 до найменшого №12 і сполучені між собою на графіках лініями. Можна бачити, що починаючи з $N = 8000$ і до $N = 16000$ криві майже співпадають.



Мал. 5. Збіжність власних значень матриці коваріацій оператора "Dima". По горизонталі номер власного числа. По вертикалі значення власних чисел матриці Σ.

На мал.6 подані криві, які відповідають власним числам матриць коваріацій $f \in R_{12}$ і $N = 16000$, для різних операторів, що дає можливість будувати критерії і алгоритми автентифікації оператора системи керування на основі цих кривих.



Мал. 6. Відмінність між власними числами різних операторів.

ВИСНОВОК

На підставі результатів проведених чисельних експериментів можна зробити висновок про те, що власні числа матриці коваріацій Σ параметрів $f \in F_n$ цілком ідентифікують оператора системи керування процесом [2]. Це дає можливість побудувати алгоритми розв'язку задачі автентифікації використовуючи критерій Фішера [1].

Література

1. *С.А. Айвазян,В.М. Бухштабер, И.С. Енюков, Л.Д. Мешалкин* Прикладная Статистика Классификация и снижение размерности Справочное издание Под редакцией проф.. *С.А. Айвазяна.* – Москва: Финансы и статистика, 1989. -608 с.
2. Бабак О. В., Гасанов А. С., Бабаєв Т. А., Баклан І. В. Єдиний підхід до розв'язання деяких задач дискримінантного та кластерного аналізу // Праці П'ятої Всеукраїнської міжнародної конференції УкрОБРАЗ'2000. – К.: 2000. – С. 67-70.

ОЗВУЧЕНИЙ РЕДАКТОР ТЕКСТІВ ДЛЯ КОРИСТУВАЧІВ З ВАДАМИ ЗОРУ

Валерій Пилипенко, Микола Сажок

Міжнародний науково-навчальний центр інформаційних технологій та систем

40 просп. Академіка Глушкова, Київ 03680

Електронна пошта: mykola@uasoiro.org.ua, speech@public.icyb.kiev.ua

Abstract

Valery Pylypenko, Mykola Sazhok. **Spoken Text Editor for Blind Users**. The presented application is developed to provide interaction between computer and blind users. By means of both spoken feedback and input user can edit formatted text, insert there objects, save and print the document. Benefits and problems of visual channel substitution with spoken one are discussed.

Вступ

Персональні ЕОМ все ширше входять у життя людей, допомагаючи їм у праці та побуті, збагачуючи знаннями та досвідом. Проте існує чимала категорія людей, яким елементарне користування комп'ютером не доступне через різні фізичні вади.

Зокрема, гостро ця проблема постає перед людьми з вадами зору. Адже сучасні операційні системи, серед яких чільне місце посідає *MS Windows*, насправді зовсім не пристосовані під потреби таких користувачів, оскільки ОС орієнтовані на візуальну підтримку взаємодії з користувачем. Сама концепція вікон передбачає візуальне сприйняття цих об'єктів і операції над ними. Це також стосується і стандартних засобів введення інформації, таких як клавіатура та, особливо, мишка.

Проте сучасний комп'ютер є досить потужним засобом, в якому є можливість заміни візуального каналу взаємодії звуковим. Це дозволяє на основі стандартної ПЕОМ розробити альтернативні засоби праці користувачів з вадами зору в ОС, подібних *Windows*. Прикладом такого засобу є звуковий редактор, призначений для набору, форматування, збереження та друкування текстів.

1 Структура редактора

Звуковий редактор складається з 5 основних програмних модулів, які взаємодіють з комп'ютерними засобами введення-виведення, такими як мікрофон, звукові динаміки та клавіатура (Рис. 1).

1. Текстовий редактор забезпечує редагування, форматування, пошук, збереження, відкриття та друк документів.
2. Координатор звукових подій та клавіатури здійснює синхронізацію введення-виведення звуку та натискань на клавіші.
3. Модуль розпізнавання голосових команд з мікрофону забезпечує розпізнавання та навчання окремо вимовлюваних слів.
4. Програмний модуль озвучення текстів, який синтезує мовний сигнал та виводить його до навушників або на зовнішні динаміки.
5. Акустична та лінгвістична бази даних озвучення текстів, яка включає модуль розмітки наголосів та інтерпретації іншомовних символів.

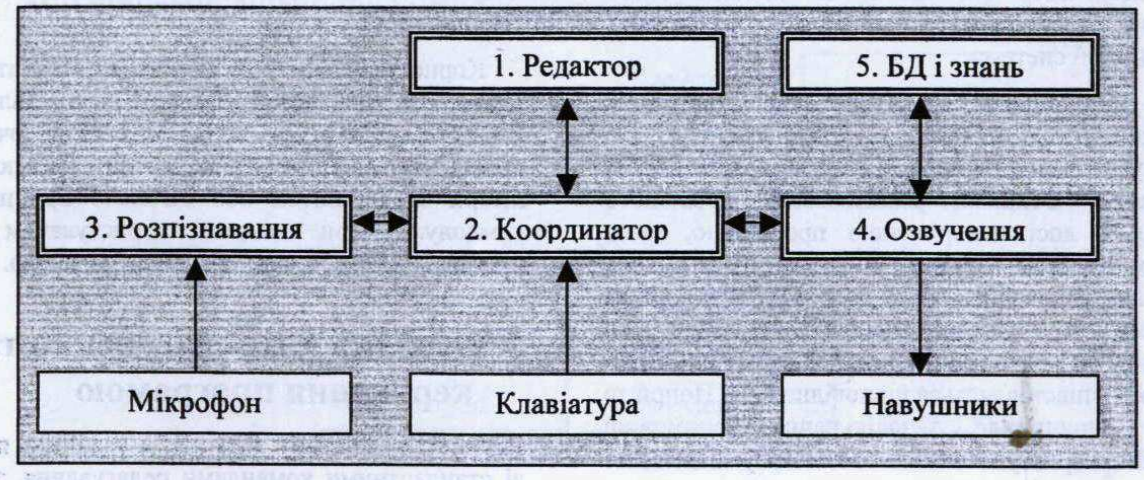


Рис. 1 Структура звукового редактора.

2 Засоби синтезу і розпізнавання усної мови

Синтезована усна мова виконує роль замінника візуального каналу взаємодії з користувачем. Усі дії користувача мають звуковий зворотній зв'язок, як то: вимовляння окремих літер та слів при натисканні клавіш, озвучення пунктів меню, діалогових вікон та елементів у ньому.

Користувач також повинен мати уяву про поточний стан системи: який файл він редагує, де розташовується точка введення тексту та які дії очікує комп'ютер від нього. Це забезпечується введенням додаткових команд, які оголошують поточний стан системи.

Ці та інші команди користувач може активізувати як за допомогою клавіатури, так і голосом, що забезпечується використанням модуля розпізнавання усної мови. Оскільки розпізнавання довільного мовлення є досі невирішеною проблемою, то в редакторі використовується система розпізнавання окремо вимовлюваних слів з фіксованого словника [1]. Тому було запропоновано використовувати побуквенне надиктовування текстів, де кожній літері та символу співставляється відповідне ім'я. Попри те що це спричинило заповільнення усномовної взаємодії з користувачем в свою чергу підвищило надійність розпізнавання та дозволило вводити довільні тексти при обмеженості словника модуля розпізнавання.

Для синтезу використовується синтезатор російської мови, створений на основі синтезатора розміченого тексту, розробленого в МДУ. Цей синтезатор був доповнений модулем транскрибування української та англійської мов, оскільки тексти, якими оперують користувачі здебільшого багатомовні.

Широко використовуються короткі звуки для позначення дій користувача під час редагування. Це і зміна мови введення тексту, початок-кінець документу, особливі символи такі як "Insert", "Home", "End", "Delete" тощо.

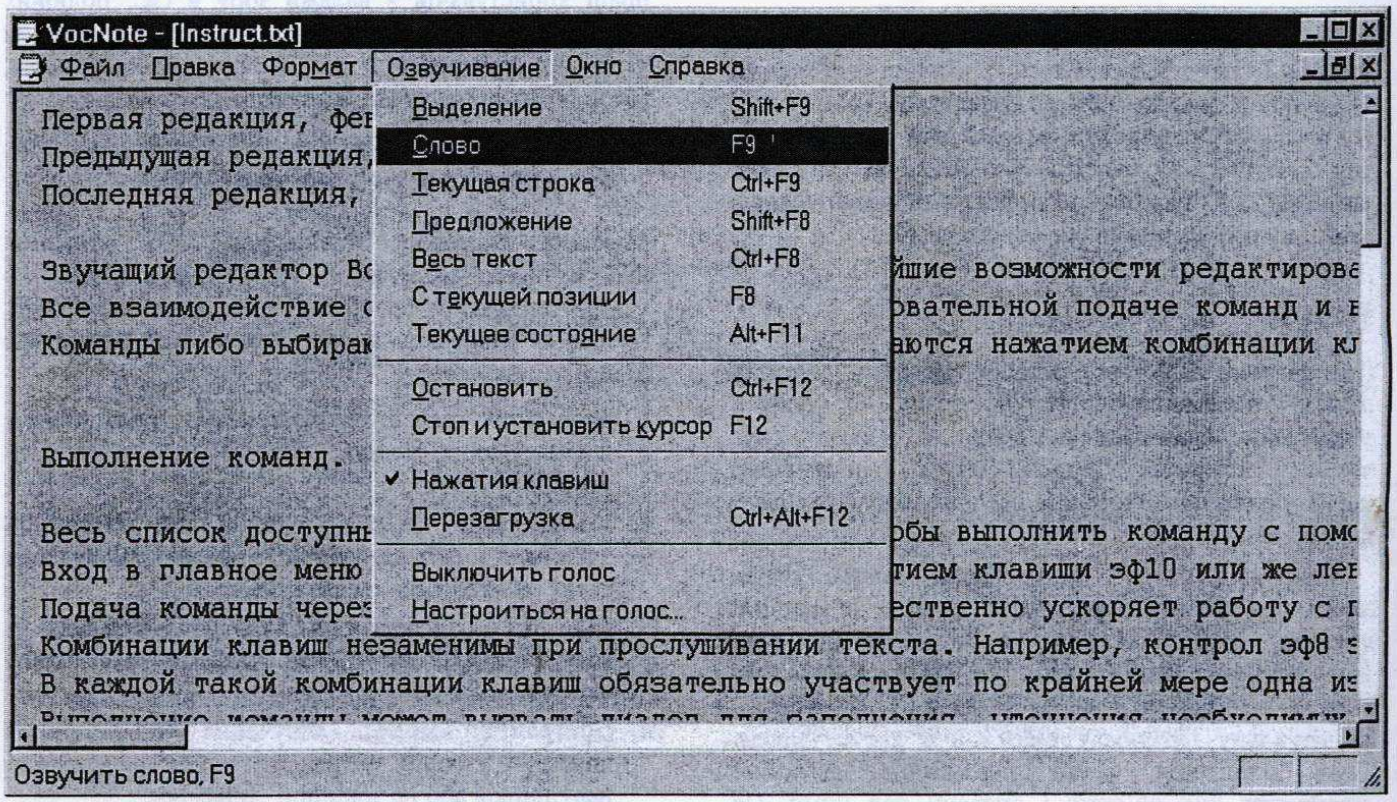


Рис. 2. Зовнішній вигляд голосового редактора користувачів з вадами зору "VocNote".

Користувач має можливість прослухати окреме слово, речення, абзац або весь текст цілком. При переміщенні курсору відтворюється поточна буква, при цьому для синтезу використовується повна назва літери за алфавітом або загально прийнята назва символу. При введенні користувачем символу пробілу озвучується останнє введене слово.

3 Особливості редагування та керування програмою

Користувач з вадами зору має можливість працювати зі стандартними командами редагування, такими як набір тексту, пошук тексту, заміна, збереження

тексту на твердому дискові, друк (Рис. 2). Також є можливість вибору різних шрифтів для редагування, в тому числі математичних. Передбачені додаткові команди для з'ясування характеристик виділених символів, як то колір, назва, висота, нахил шрифту, мову.

Дуже важливу роль виконують команди роботи з текстовими блоками. Стандартний спосіб виділення за допомогою натисненої клавіші Shift, а тим більше за допомогою мишки не підходить для користувачів з вадами зору. У редакторі реалізований додатковий набір команд для виділення і оперування блоками тексту, який виконується за допомогою натискання клавіш, вибором пунктів меню або голосом. Користувачі виявили бажання використати накопичувальний буфер обміну (кишеню), в якому вони можуть умістити декілька порцій тексту.

Операції редагування, що можуть привести до сильних змін тексту, заблоковані або запитують підтвердження на виконання. Наприклад, заблокована заміна виділеного блоку одним символом, що дозволяється у стандартних редакторах.

Всі операції з текстом: взяття в кишеню, вставка з кишені, видалення і т.п. супроводжуються короткими звуками або докладними коментарями (за бажанням користувача).

Озвучення взаємодії з користувачем також відбувається і при роботі зі вставними об'єктами, такими як редактор математичних формул.

Режими озвучення за бажанням користувача перемикаються від максимального, коли відтворюються всі символи і повідомлення редактора, до мінімального, коли всі додаткові повідомлення не озвучуються. Максимальний режим озвучення корисний для навчання редагуванню, а мінімальний необхідний для досвідчених користувачів.

4 Настроювання на голос користувача

Дуже гостро постає проблема навчання на диктора, оскільки система розпізнавання, використана в редакторі, є дикторозалежною.

У звуковому редакторі застосовується навчання на голос користувача за допомогою звукової підказки (Рис. 3). Користувач викликає режим навчання і програма підказує наступне слово для навчання. Після промовляння слова, система аналізує сигнал і в випадку занадто гучного, або слабкого, сигналу запрошує промовити це слово ще раз.

Але система не має можливості скасувати промову, якщо користувач промовляє зовсім інше слово ніж підказує система. Рекомендується протестувати навчання на словник, для виявлення невдалого промовляння тих чи інших слів та скоригувати навчання.

Результати настроювання автоматично зберігаються в окремому файлі та викликаються, у випадку активації голосового керування, звуковим редактором. Можливе настроювання як на одного користувача так і на декількох.

5 Взаємодія з середовищем *Windows*

Звуковий редактор реалізований у середовищі *Windows* та використовує стандартні засоби даної системи. Для введення та виведення звуку використовується стандартна звукова карта типу *SoundBlaster*. Все це дає можливість встановлювати звуковий редактор на довільний комп'ютер, який обладнаний звуковою картою, та працює в середовищі *Windows*.

Рис. 3. Діалогове вікно для настроювання на голос користувача.

Розпізнавання ти синтез мови виконуються в реальному масштабі часу та не потребують від користувача занадто довгого очікування тієї чи іншої події. Попри очевидні переваги при введенні та виведенні даних, звуковий канал не може повністю замінити візуальний за швидкістю та зручністю роботи користувача.

Віконна операційна система використовує багато візуальних засобів для підвищення працездатності та зручності користування комп'ютером. Багато з цих засобів не можливо адекватно передати в звуковому вигляді. В даному редакторі переслідується прагматична ціль - надати можливість користувачеві з вадами зору працювати з текстами.

Для повноцінної роботи користувачеві з вадами зору доводиться дуже довго опановувати концепції ОС *Windows*, прийняті для роботи з діалоговими вікнами. Тому в даному редакторі максимально збережений стиль діалогових вікон *Windows* для того, щоб користувачі поступово засвоювали цей стиль, оскільки в майбутньому вони неминуче користуватимуться ОС *Windows* і такими складними редакторами як *Word*, *Excel*.

Наприклад для відкриття файла, його збереження використовуються стандартні діалогові вікна, в яких були прибрані всі зайві в певному розумінні елементи. Це дозволило трохи спростити процес задавання потрібних користувачеві опцій, таким чином поступово його вводячи у світ сучасних операційних систем.

6 Досвід користування

Даний редактор використовується членами асоціації "Вікно у світ" [2] у їх повсякденній роботі при написанні листів, статей, звітів тощо.

Користувачі мають можливість порівняти звуковий редактор з іншими засобами допомоги. Брайлевський дисплей дозволяє адекватно передавати графічні об'єкти, але має значну собівартість та малу швидкодію. Також потребує значних зусиль по навчанню користувача на брайлевській техніці.

Універсальні засоби для озвучення графічного операційного середовища дозволяють користуватися довільними додатками *Windows*. З іншого боку користувач зіштовхується з явищем занадто великої кількості елементів, які не використовуються при звичайній роботі. Також потребує опанування користувачем складних концепцій графічного інтерфейсу.

Звуковий редактор потребує значно меншого часу для освоювання, та дозволяє зробити перший крок в опануванні концепцій середовища *Windows*.

При асоціації "Вікно у світ" відкритий комп'ютерний клас, в якому користувачі навчаються роботі зі звуковим редактором. Для цього розроблені спеціальні методики навчання, якими користуються вчителі при навчанні учнів.

7 Перспективи розвитку

Подальший розвиток редактора полягає у тому, щоб спростити роботу з таблицями, формулами та іншими сучасними засобами створення текстів. Такі складні об'єкти потребують розроблення додаткових засобів звукового відображення.

Надто гостро постає проблема оброблення малюнків, маніпулювання графічною інформацією, якими рясніють текстові документи.

Також необхідно створювати додаткові можливості, такі як голосове управління електронною поштою, скануванням книг і навіть навігацією в Інтернеті. Інтегрування додаткових засобів може дати можливість виконувати різноманітну корисну роботу, використовуючи єдиний засіб.

Висновки

Реалізовано звуковий редактор текстів для користувачів з вадами зору в середовищі *Windows*. Вказана категорія користувачів комп'ютера має змогу редагувати тексти, зберігати, викликати та друкувати їх.

Візуальний канал взаємодії замінюється звуковим за допомогою синтезу та розпізнавання мови.

Звуковий редактор використовується користувачами з вадами зору у їх повсякденній роботі.

Література

1. Винцюк Т.К. Анализ, распознавание и смысловая интерпретация речевых сигналов. — Киев: Наукова думка, 1987, 264 с.
2. Пилипенко В., Сажок М. Озвучений редактор текстів з елементами голосового керування в середовищі *Windows*. // Праці 1-го міжнародного семінару "Інформаційні технології у соціально-трудовій реабілітації інвалідів" – Київ, 2001, с 121-123.

ВИЯВЛЕННЯ ОСНОВНИХ РИТМІВ РЕГУЛЯРНОЇ СКЛАДОВОЇ ЕЛЕКТРОКАРДІОГРАМИ МЕТОДОМ НАЙМЕНШИХ КВАДРАТІВ

О. В. Заболотний♥, О. І. Яворський♦
♥Фізико-механічний інститут ім. Г. В. Карпенка НАН України
МСП, вул. Наукова 5, м. Львів, 79601
тел: 8 (0322) 65-43-40, факс: 8 (0322) 64-94-27,
e-mail: ZOV@ipm.lviv.ua
♦Академія Управління Персоналом
Пр. Чорновола 57 м. Львів, 79000
тел: 8 (0322) 52 33 83, факс: 8 (0322) 978836
e-mail: Iavor@ipm.lviv.ua

Застосовано модель полі-періодично корельованих випадкових процесів для опису сигналу електрокардіограми. Адаптовано метод найменших квадратів для виявлення основних частот коливань складових її математичного сподівання

1. ВСТУП

Для більшості процесів, що протікають в біологічних системах, притаманна майже періодична повторюваність і стохастичність. Це обумовлено тим, що фізіологічний стан біологічного об'єкта дуже тісно пов'язаний зі станом оточуючого середовища, в якому він перебуває (сезонна, добова мінливість температури, зміна сонячної активності, річна ритміка явищ в атмосфері та гідросфері, тощо). Тому застосування до аналізу медико-біологічних сигналів стаціонарних методів обробки не є достатньо обґрунтованим, оскільки, як показали дослідження, імовірнісні характеристики біо-процесів залежать від часу, що не передбачається в рамках теорії стаціонарних процесів. Краще наближення для медико-біологічних сигналів дає модель майже періодичних корельованих випадкових процесів (МПКВП) [1]. Відразу зазначимо, що віднісши сигнал до класу МПКВП, ми припускаємо часову інваріантність основних частот коливань складових його імовірнісних характеристик. Для біологічних систем це припущення справедливе лише на невеликому часовому інтервалі. В загальному випадку частоти сильно залежать від фізіологічного стану об'єкта досліджень і є діагностичними параметрами. Водночас, для того, щоб для аналізу сигналу можна було застосувати розроблені методи МПКВП, потрібно з високою точністю задати частотні набори для математичного сподівання і кореляційної функції [2]. Таким чином, при вивченні біологічних процесів, важливим етапом є визначення їхніх періодів корельованості.

2. ОГЛЯД ЧАСТОТНИХ ХАРАКТЕРИСТИК ЕЛЕКТРОКАРДІОГРАМИ

Типовим медико-біологічним сигналом з поліритмічною структурою є електрокардіограма (ЕКГ) (рис. 1).

Рис. 1 Електрокардіограма людини

Частота скорочень міокарда залежить від багатьох факторів: нервової активності, фізичного навантаження, атмосферного тиску, тощо. Тому, усереднений спектр ЕКГ на великій довжині реалізації, в стаціонарному наближенні, дуже складний (рис. 2).

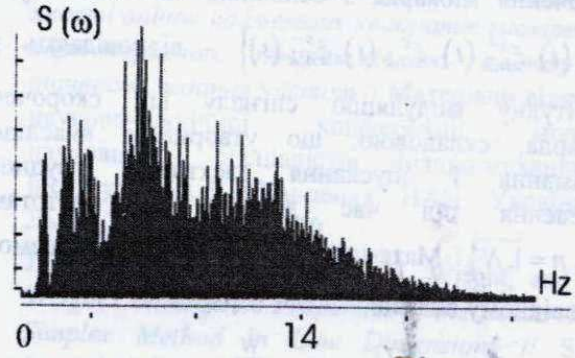


Рис. 2 Усереднений спектр ЕКГ за 100 с.

Прослідкувати зміну частотних характеристик сигналу в часі можна за допомогою частотно-часового представлення (рис. 3).

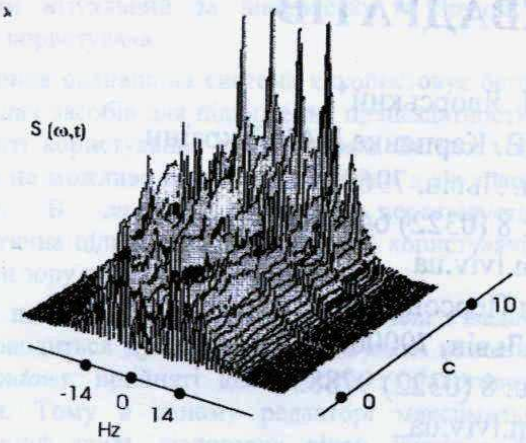


Рис. 1 Частотно-часове представлення ЕКГ людини

3. ПОБУДОВА МПКВП- МОДЕЛІ СИГНАЛУ ЕКГ

На невеликому часовому інтервалі (5-10 основних періодів) сигнал ЕКГ можна наближено представити у вигляді МПКВП

$$\xi(t)=\xi_0(t)+$$

$$+\sum_{k=1}^{L}\left[\xi_{0,k}^{c}(t)\cos k\omega_0 t+\xi_{0,k}^{s}(t)\sin k\omega_0 t\right]+$$

$$+\sum_{k=1}^{L}\sum_{n=1}^{N}\left[\xi_{n,k}^{c+}(t)\cos(k\omega_0+\omega_n)t+\xi_{n,k}^{s+}(t)\sin(k\omega_0+\omega_n)t\right]$$

$$+\sum_{k=1}^{L}\sum_{n=1}^{N}\left[\xi_{n,k}^{c-}(t)\cos(k\omega_0-\omega_n)t+\xi_{n,k}^{s-}(t)\sin(k\omega_0-\omega_n)t\right]$$

де $\{\xi_0(t),\xi_{n,k}^{c}(t),\ \xi_{n,k}^{s}(t),\ \xi_{n,k}^{c+}(t),\xi_{n,k}^{s+}(t),\ \xi_{n,k}^{c-}(t),\xi_{n,k}^{s-}(t)\}$ - стаціонарні випадкові процеси. Причому $\{\xi_0(t),\xi_{n,k}^{c}(t),\ \xi_{n,k}^{s}(t)\}$ - відповідають за скорочення міокарда з основною частотою ω_0, а $\{\xi_{n,k}^{c+}(t),\xi_{n,k}^{s+}(t),\xi_{n,k}^{c-}(t),\xi_{n,k}^{s-}(t)\}$ - відповідають за амплітудну модуляцію сигналу від скорочень міокарда, складовою, що утворилася внаслідок піднімання і опускання електродів грудного відведення під час дихання із частотами $\{\omega_n,\ n=\overline{1,N}\}$. Математичне сподівання запишемо у відповідному вигляді.

$$\mathrm{m}(t)=\mathrm{m}_0(t)+\sum_{n=1}^{N}\mathrm{m}_n(t),$$

де

$$\mathrm{m}_0(t)=\mathrm{m}_0+\sum_{k=1}^{L}\left[\mathrm{m}_{0,k}^{c}\cos k\omega_0 t+\mathrm{m}_{0,k}^{s}\sin k\omega_0 t\right]$$

$$\sum_{n=1}^{N}\mathrm{m}_n(t)=\sum_{k=1}^{L}\sum_{n=1}^{N}\left[\mathrm{m}_{n,k}^{c+}\cos(k\omega_0+\omega_n)t+\right.$$

$$\left.+\mathrm{m}_{n,k}^{s+}\sin(k\omega_0+\omega_n)t\right]+$$

$$\sum_{k=1}^{L}\sum_{n=1}^{N}\left[\mathrm{m}_{n,k}^{c-}\cos(k\omega_0-\omega_n)t+\mathrm{m}_{n,k}^{s-}\sin(k\omega_0-\omega_n)t\right],$$

де компоненти математичного сподівання $\mathrm{m}_{(\square)}^{(\square)}=E\xi_{(\square)}^{(\square)}(t)$.

4. ОЦІНКА ЧАСТОТ МАТЕМАТИЧНОГО СПОДІВАННЯ

Частоти $\{\omega_0,\omega_n,n=\overline{1,N}\}$ шукатимемо з умови мінімуму середнього квадратичного відхилення:

$$F\left[\{\omega_0,\omega_n\},\hat{\mathrm{m}}_{(\square)}^{(\square)}\right]=\frac{1}{2\theta}\int_{-\theta}^{\theta}\left[\xi(t)-\hat{\mathrm{m}}(t)\right]^2 dt, \quad (1)$$

де $[-\theta,\theta]$ - відрізок часу на якому задано реалізацію ЕКГ $\xi(t)$, $\hat{\mathrm{m}}(t)$ - оцінка математичного сподівання $\mathrm{m}(t)$, а компоненти $\hat{\mathrm{m}}_{(\square)}^{(\square)}$ - оцінки компонент математичного сподівання $\mathrm{m}_{(\square)}^{(\square)}$.

Згідно отриманих теоретично результатів оцінка частот математичного сподівання, отриманих шляхом мінімізації функціоналу (1), в першому наближенні є незміщена і слушна.

Для мінімізації функціоналу (1) використовувався симплекс метод [3], одним із вхідних параметрів якого є наближене значення шуканої частоти. Грубу оцінку частоти знаходили зі спектральної густини потужності, яку обчислювали використовуючи алгоритм швидкого перетворення Фур'є. Хоч дисперсія оцінки найменших квадратів частот математичного сподівання згасає пропорційно θ^{-3}, реалізації довжиною 5-10 основних періодів виявилося недостатньо для одночасного розділення усіх ритмів. Алгоритм оцінки періодів кореляції було розділено на два етапи. Спочатку було оцінено основну частоту ω_0 скорочень міокарда, і відповідні амплітуди гармонік математичного сподівання $\hat{\mathrm{m}}_0(t)$: $\{\hat{\mathrm{m}}_{0,k}^{(\square)},k=\overline{1,L}\}$ ($K=40$) (рис. 4). Фактично при цьому мінімізувавши функціонал

$$F_0\left[\omega_0, \hat{m}_{0,(\cdot)}^{(\cdot)}\right] = \frac{1}{2\theta}\int_{-\theta}^{\theta}\left[\xi(t) - \hat{m}_0(t)\right]^2 dt,$$

де $\hat{m}_0(t)$ - оцінка складової $m_0(t)$.

Щоб прискорити процедуру оптимізації функціоналу $F_0[\bullet]$ по частоті ω_0, було враховано встановлений експериментальним шляхом факт, що основна потужність сигналу знаходиться в частотному інтервалі $[0,20]$ Hz

Рис. 4 Компоненти математичного сподівання на $\hat{m}_0(t)$

Це дозволило при оцінюванні частоти ω_0 обмежитися $L = 20$ компонент. За початкове значення частоти ω_0 вибирали частоту, для якої спостерігався максимум спектральної густини потужності ЕКГ на частотному інтервалі $[0.9, 1.5]$ Hz

На другому етапі проводилася оцінка частот $\omega_n, n = \overline{1,N}$. Після центрування сигналу $\xi(t)$ на оцінку $\hat{m}_0(t)$

$$\xi_1(t) = \xi(t) - \hat{m}_0(t)$$

в спектрі сигналу $\xi_1(t)$ знаходили глобальний максимум на деякій частоті ω. Знаходимо початкове наближення для ω_1

$$\omega_{10} = \left|\omega_0 - \{\omega, \omega_0\}\right|,$$

де оператор $\{\omega, \omega_0\}$ - рівний остачі від ділення ω на ω_0 і знаходимо уточнене значення ω_1 з умови мінімуму функціоналу

$$F_1\left(\omega_1, \hat{m}_{1,(\cdot)}^{(\cdot)}\right) = \frac{1}{2\theta}\int_{-\theta}^{\theta}\left[\xi_1(t) - \hat{m}_1(t)\right]^2 dt, \quad (2)$$

і компоненти складової математичного сподівання $\hat{m}_1(t)$.

Аналогічно робимо і для усіх решти $\hat{m}_n(t)$, $n = 2,3,..,N$, поки потужність $\xi_N(t)$ не буде меншою за деяке порогове значення δ.

В даному випадку ітераційний процес продовжувався, поки з сигналу ЕКГ не було забрано 70% енергії $N = 4$. Оцінку математичного сподівання, яку ми отримали при цьому, зображено на рис. 5

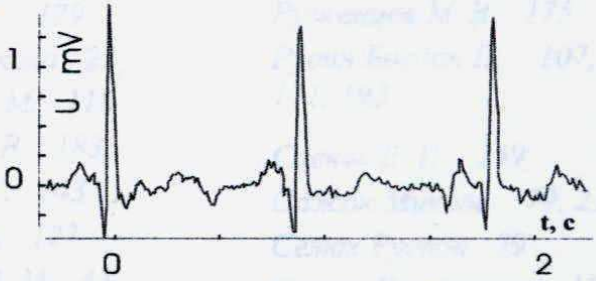


Рис.5 Оцінка математичного сподівання ЕКГ (рис.1)

5. ВИСНОВКИ

Умовне розбиття алгоритму оцінювання частот математичного сподівання сигналу ЕКГ на описані два етапи, дозволило звести процедуру оптимізації функціоналу (1) з $N+1$ нелінійним параметром, до мінімізації $N+1$ функціоналів типу (2) з одним нелінійним параметром. Це дозволило зменшити розмірність задачі, що в свою чергу покращило збіжність алгоритму оптимізації. За рахунок рознесення шуканих параметрів в частотній області на відстань кратну ω_0 в кожній ітерації нам вдалося уникнути ситуації оцінювання близьких частот, що має дуже важливе значення при обробці сигналів на короткій довжині реалізації.

Дана модифікація алгоритму оцінювання частот за методом найменших квадратів дозволяє вивчати поліритмічну структуру сигналу зі складною мультиплікативною взаємодією двох процесів. В нашому випадку такими процесами є процес скорочення серцевого м'яза і процес розширення і звужування грудної клітки під час дихання.

6. ЛІТЕРАТУРА

1. Заболотний О.В., Михайлишин В.Ю., Яворський І.М., *Метод найменших квадратів при статистичному аналізі поліритміки* // Доповіді національної академії наук 2000, №8 – с.93-101.
2. О.В.Заболотний, М.В.Михайлишин *Залежність якості оцінок найменших квадратів імовірнісних характеристик поліритмічного сигналу від точності задання частот* // Матеріали відкритої науково-технічної конференції молодих науковців і спеціалістів Фізико-механічного інституту ім. Г.В.Карпенка НАН України. – Львів, 2002. – С.118-120.
3. J.C. Lagarias, J.A. Reeds, M.H. Wright, and P.E. Wright *Convergence Properties of the Nelder-Mead Simplex Method in Low Dimensions* // SIAM Journal of Optimization, Vol. 9 Number 1, pp. 112-147, 1998.

АЛФАВІТНИЙ ПОКАЖЧИК. AUTHOR INDEX