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_МНП_ЦЕНТР за ДНТП "Образний комп'ютер"**

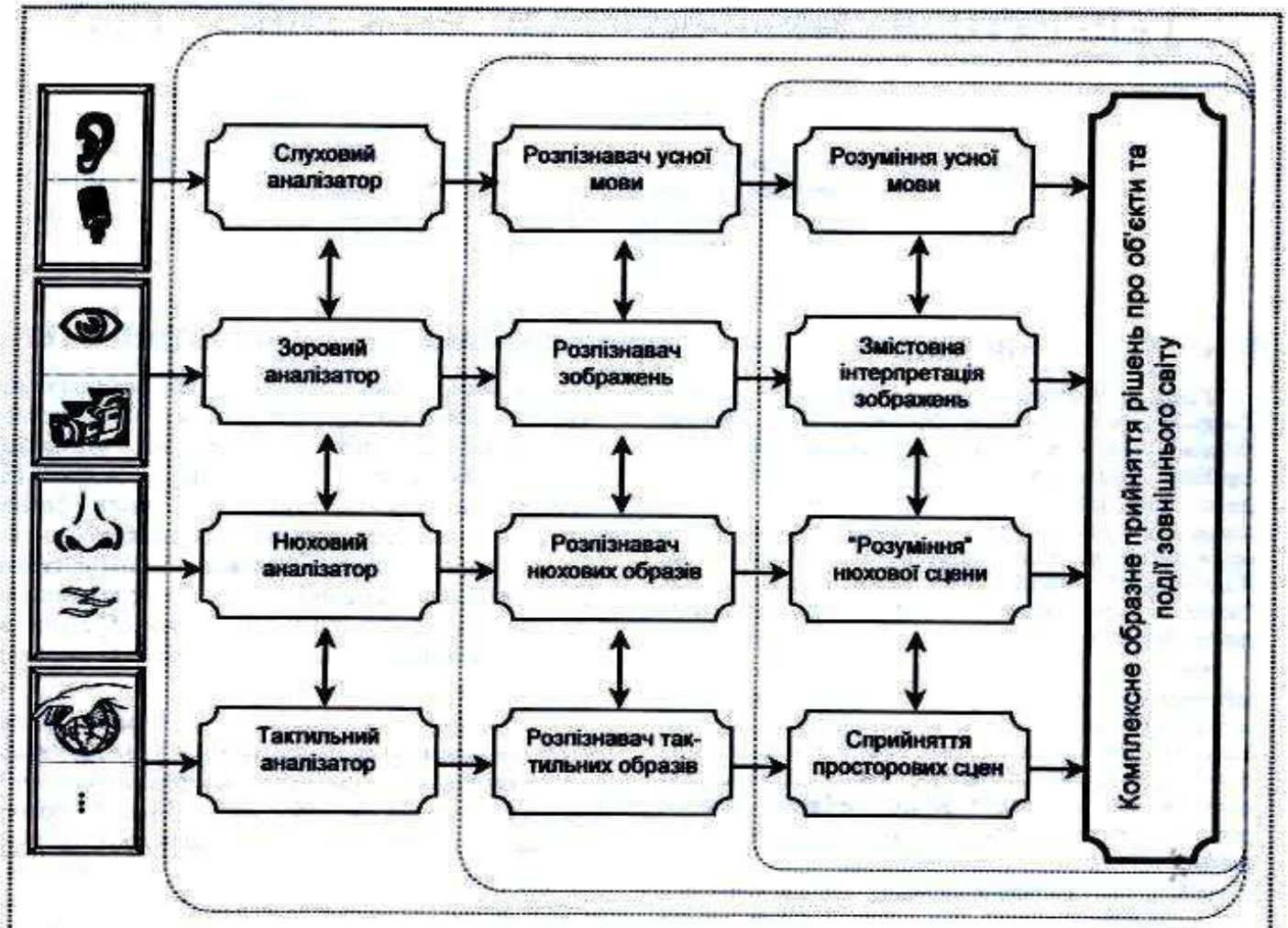

Рис. 1. Схематичне зображення структури образного комп'ютера. Зворотні зв'язки не показані.

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 грунтується 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 одією 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

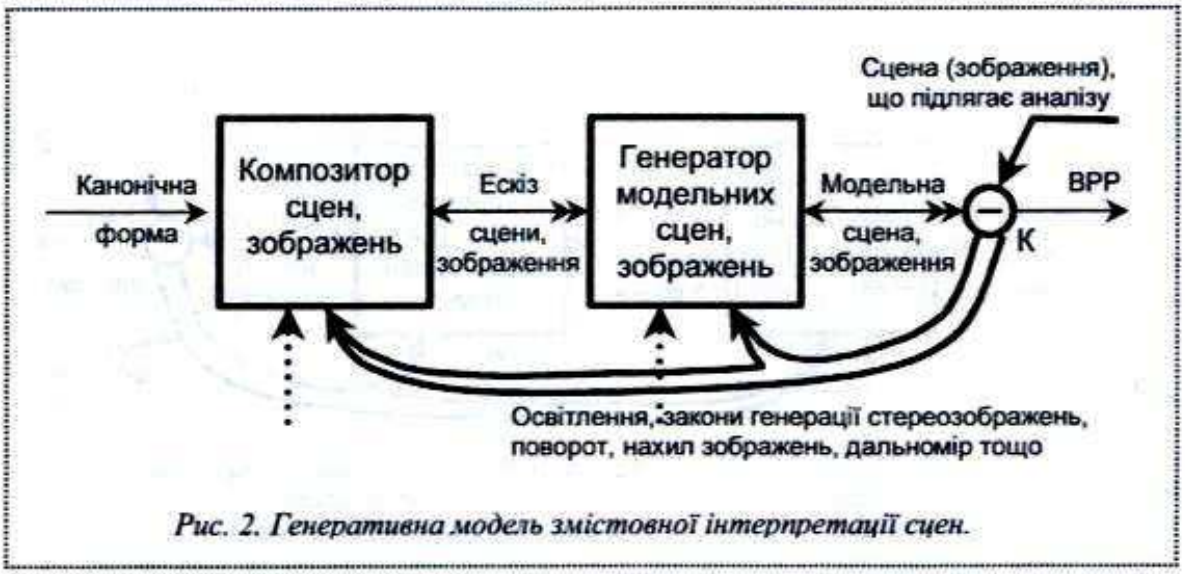

Рис. 2. Генеративна модель змістовної інтерпретації сцен.

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

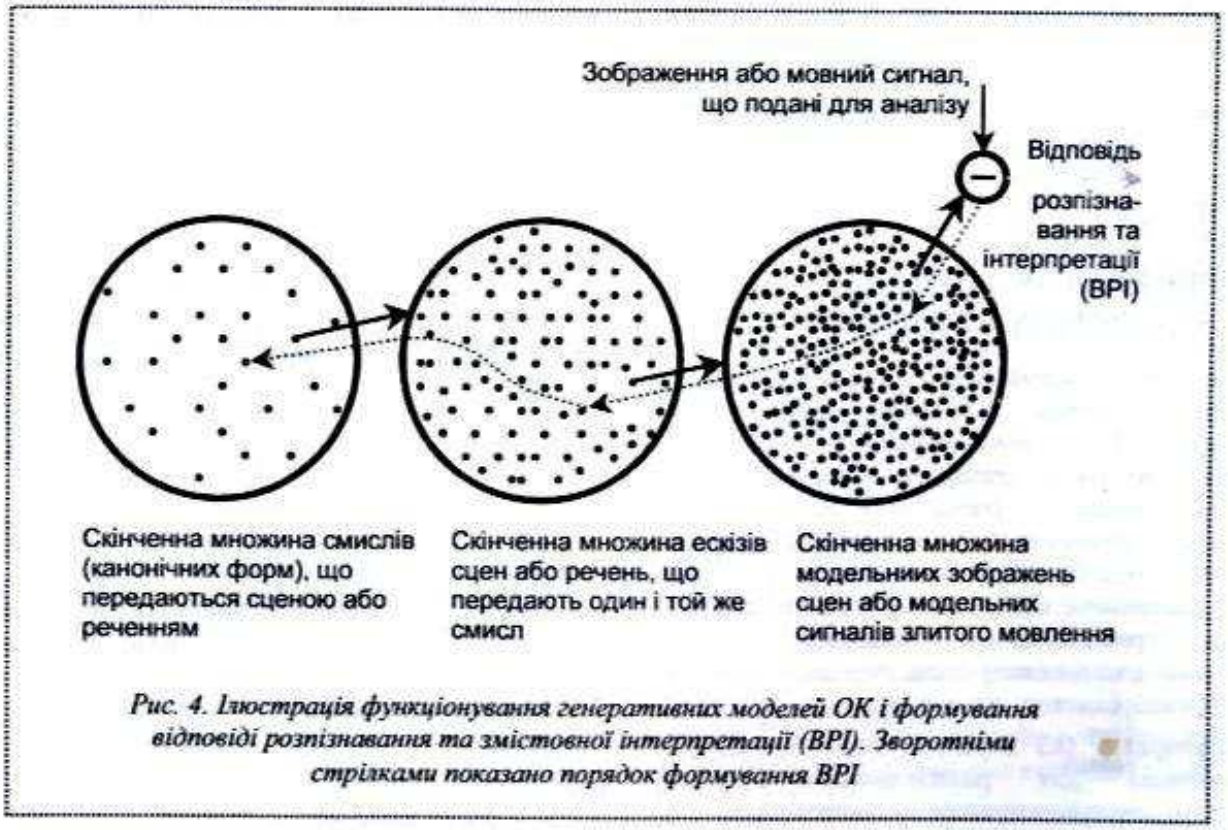

Рис. 4. Ілюстрація функціонування генеративних моделей ОК і формування відповіді розпізнавання та змістовної інтерпретації (ВРІ). Зворотніми стрілками показано порядок формування ВРІ

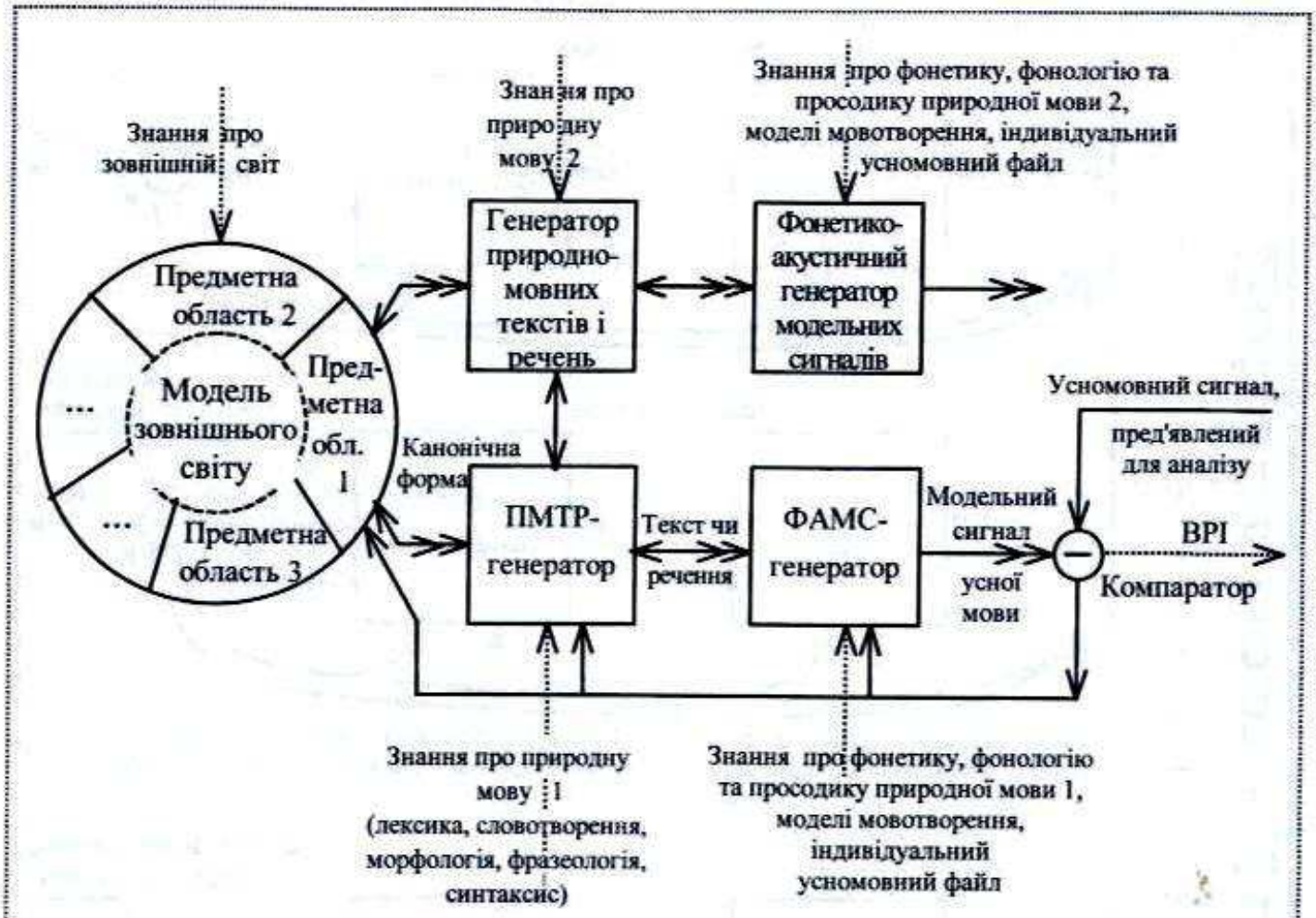

Рис. 5. Структура диктувальної машини та машини для усного перекладу.

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

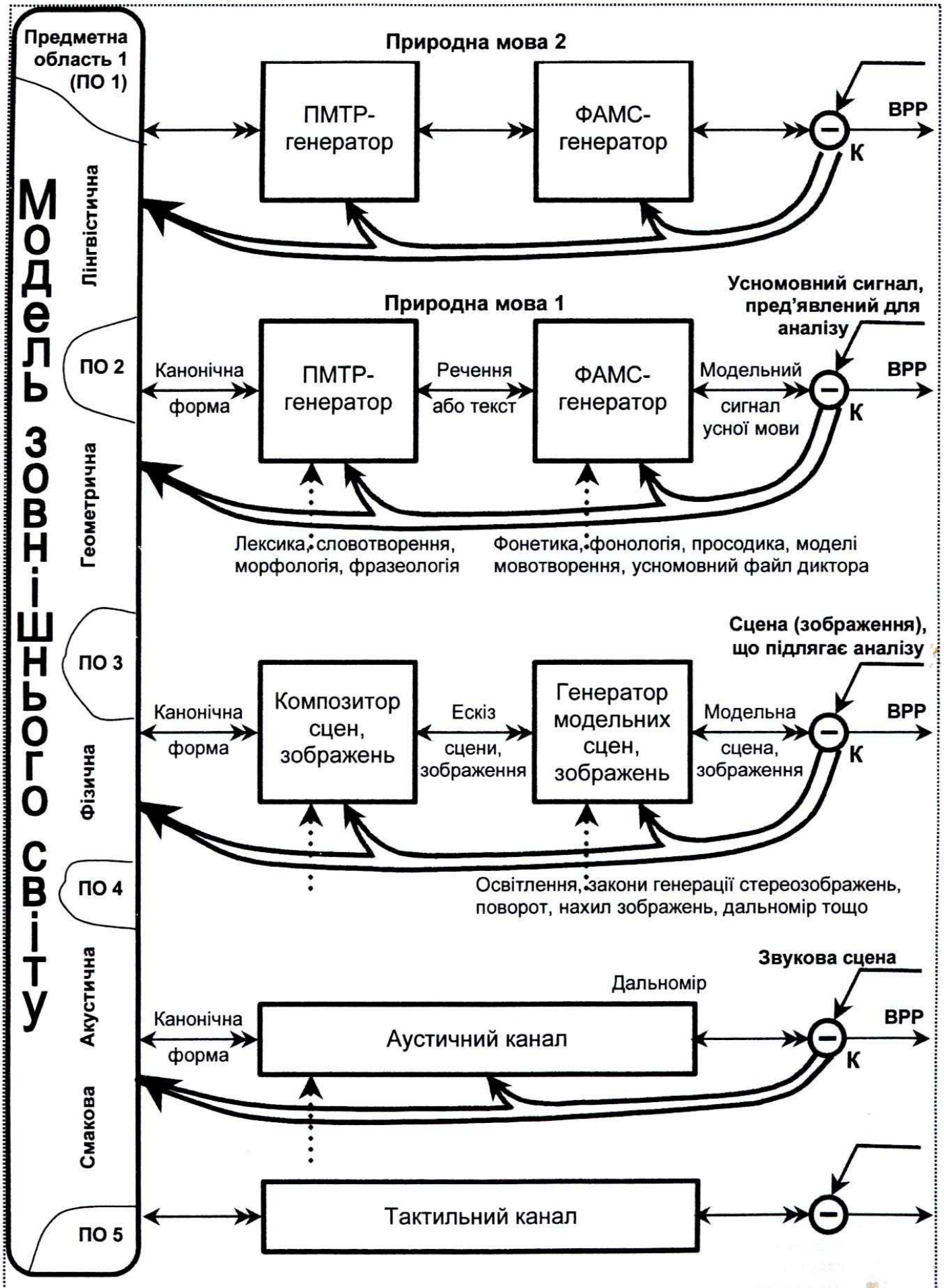

Рис. 6. Генеративна модель образного комп'ютера.

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 області 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 області. Тепер вирішення проблеми ПУД-машини для цієї предметної області 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

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

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

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

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

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

is valid, where

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

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

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

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

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

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

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

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 $\Phi(z)$,

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

2.2. Equivalent problems

Two problems

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

and

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

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

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

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

$$\langle T,\Im,K,(g^1_{tt'} \mid tt'\in\Im)\rangle ,$$
$$\langle T,\Im,K,(g^2_{tt'} \mid tt'\in\Im)\rangle$$

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

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

■

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 $G(\bar{k})$ does not depend on the problem of the class Z. The inequality

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

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

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

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:

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

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

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

where

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

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

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

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

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

where $g_{tt'}(k,k') = 0$, if $|k-k'| \le \Delta_{tt'}$, $= -\infty$, if $|k-k'| > \Delta_{tt'}$. (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 $\Phi(\vec{\mathbf{r}},t)$ 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

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

де $\vec{\mathbf{r}}_1$ і $\vec{\mathbf{r}}_2$ 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:

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

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

Таким чином, присутня в спектральному розкладі (2), (3) функція спектральної густини $S(\vec{\mathbf{r}}_1,\vec{\mathbf{r}}_2,\omega)$ 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 $\vec{\mathbf{r}}_i$ і $\vec{\mathbf{r}}_j$. 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.

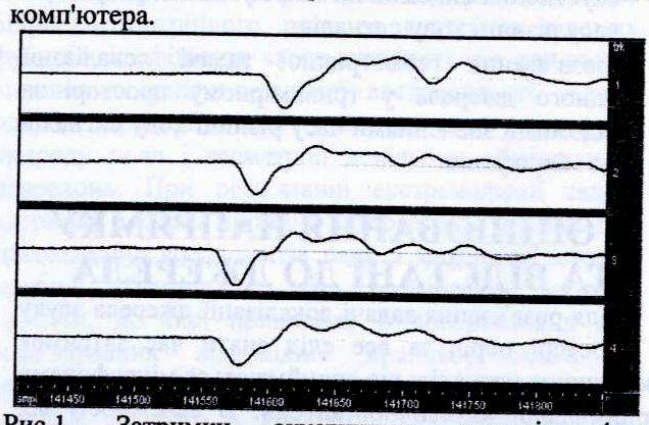

Рис.1. Затримки акустичних сигналів 4-ох мікрофонів після вмикання джерела звуку

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.

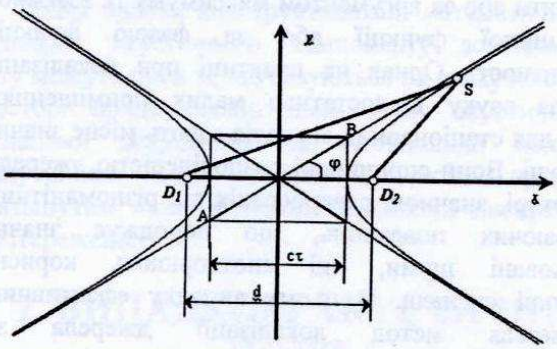

Рис. 2 Локалізація джерела звуку у площині

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.

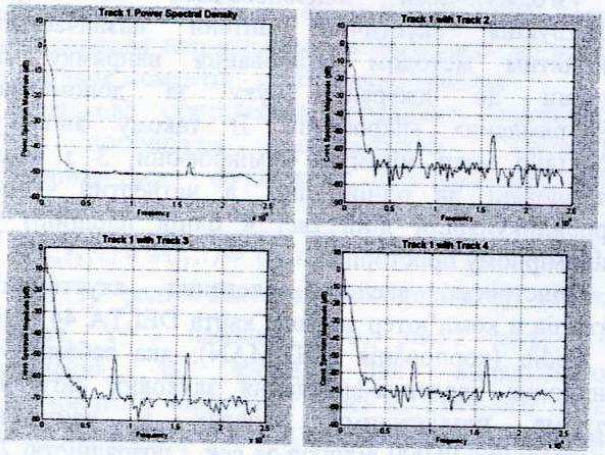

Рис. 3. Спектральні густини 4-ох акустичних сигналів

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.

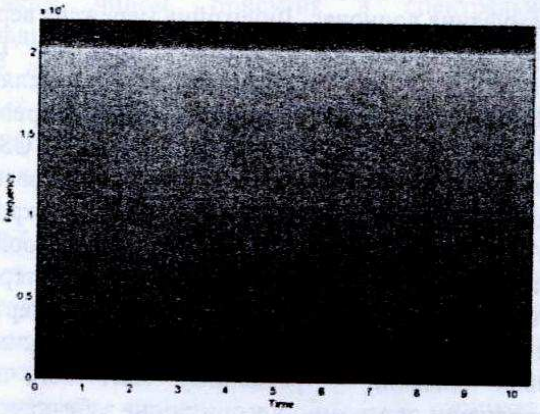

Рис. 4. Спектрограма сигналу 1-го мікрофона

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.

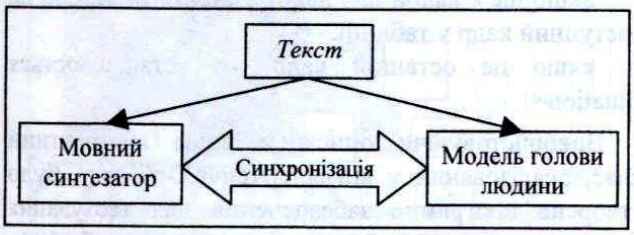

Рис.1.

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!".

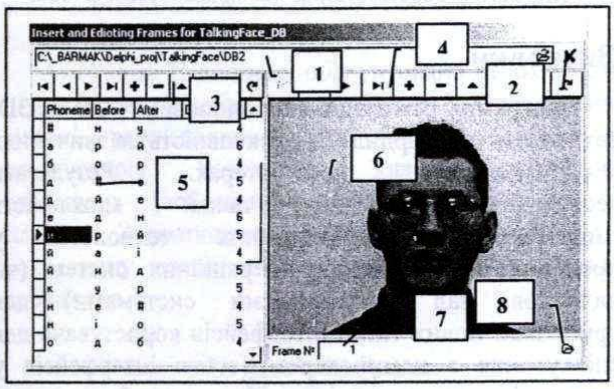

Рис. 3

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;***

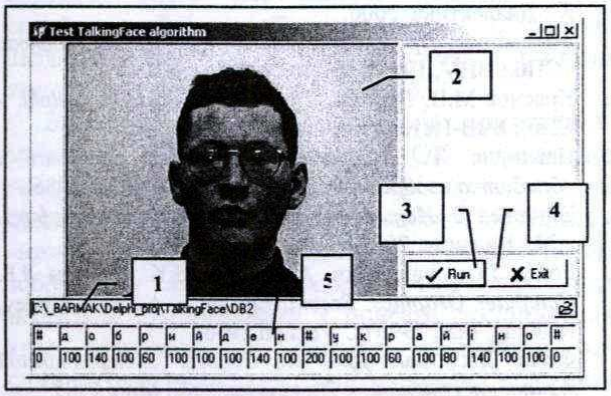

Рис.4

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).

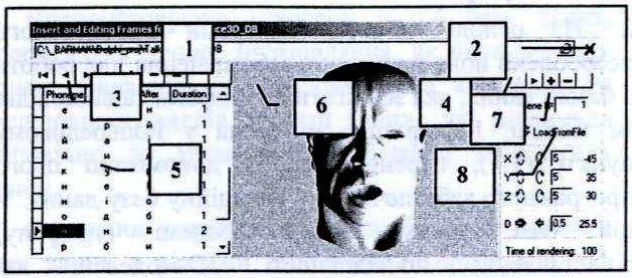

Рис. 6

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).

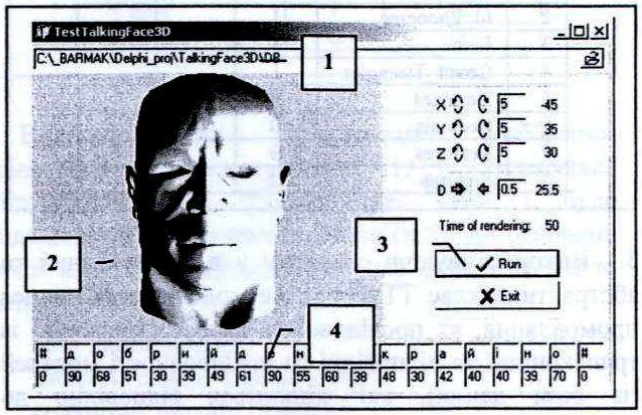

Рис. 7

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.

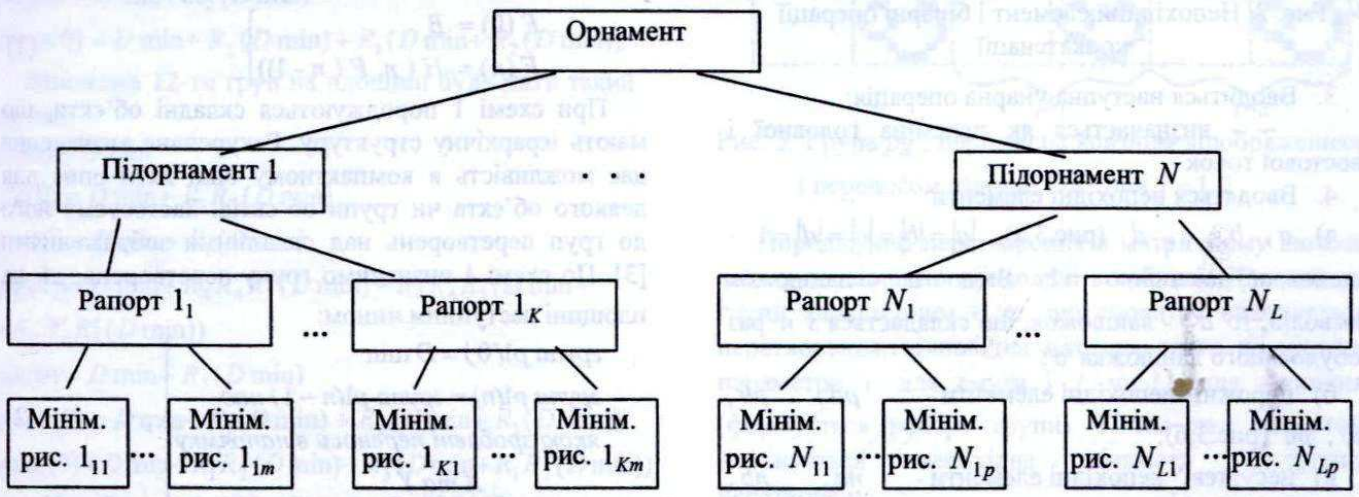

Рис. 1. Структура орнаменту

МОВА ОПИСУ ЗОБРАЖЕНЬ

Для представлення мінімального рисунку запропоновано мову опису зображень[3]

1. Кожен непохідний елемент помічається в двох різних точках - головній z 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".

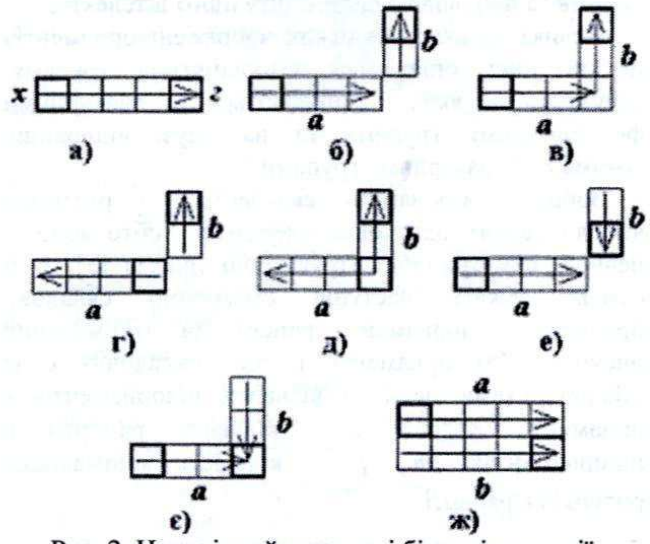

Рис. 2. Непохідний елемент і бінарні операції конкатенації

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.

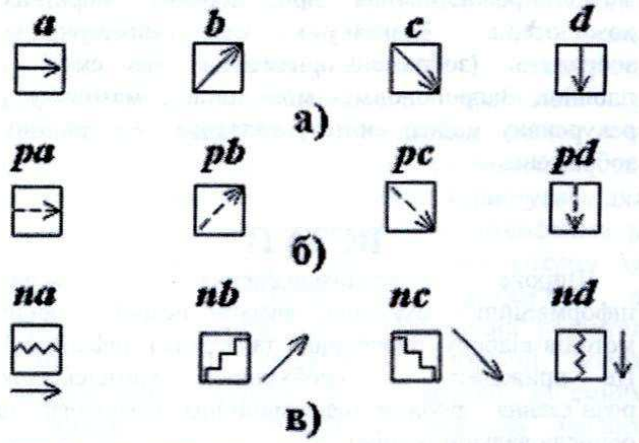

Рис. 3. Непохідні елементи: а)основні; б)«порожні»; в)«несуттєві»

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:

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

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.

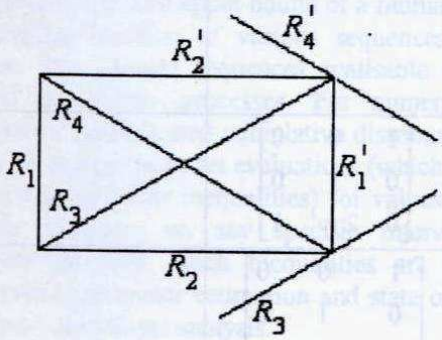

Рис. 4. Представлення осей відображень

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.

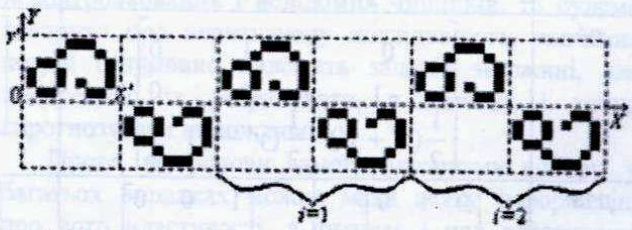

Рис. 5. Група pg, породжена ковзним відображенням і переносом вздовж OX при $i = 1, 2$.

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.