

КВР-ПЕРЕТВОРЕННЯ ПРИ РОЗПІЗНАВАННІ ЗОБРАЖЕННЯ

Кожем'яко В. П., Понура О. І.

Вінницький державний технічний університет
286021, Вінниця, Хмельницьке шосе, 97

тел. (0432) 44-00-19, 44-01-25, факс (0432) 46-57-72, E-mail: VSTU@sovam.com

В представленій роботі запропоновано та описано спосіб розпізнавання зображень на базі квантронного (часового) перетворення сигналів. Для цього перетворення запропоновано використати логіко-часові функції та ієрархічну обробку інформації, що дозволяє добитися високої достовірності при розпізнаванні об'єктів.

ВСТУП

Для отримання різноманітних результатів інженерної діяльності необхідно розв'язати проблему обробки ісходної інформації. Найбільш інтересною задачею цієї проблеми є ідентифікування зображень. Тобто становиться мета розробки оптимальної системи технічного зору.

Найблищим прототипом технічного зору є око людини. Воно базується на мозковій діяльності. При аналізі такого підходу до обробки оптичної інформації з'являється проблема інтуїтивних рішень, що не базуються на стандартних логіко-математичних алгоритмах. Виходячи з цього, спроба створення оптимальних систем технічного зору, за своїми можливостями розпізнавання, які наближаються до людських, упирається в проблему створення неалгоритмічних методів обробки та аналізу інформації.

Таким чином, перед інженерами постає задача створення пристроїв, які базуються на фізико-технологічних основах перетворення інформації, що максимально наближені до природного паралельного сприйняття. Тобто, створення пристрою око-процесорного типу [1]. Подібний пристрій повинен містити в собі три основних блоки обробки інформації:

1. вводу та первинної обробки даних;
2. аналізу інформації, що отримується;
3. розпізнавання об'єкту.

1. ВВЕДЕННЯ ТА ПОПЕРЕДНЯ ОБРОБКА

Введення інформації від зорового об'єкту та її перетворення пропонується виконувати за допомогою оптико-електронних пристроїв, які дозволяють виконувати обробку сигналів двох типів: оптичних (світлових) та електричних. Вони дають можливість виконувати введення і обробку даних шляхом паралельного оптичного порівняння при часовому

зсуві та електроному поповненні енергії. Такі умови праці забезпечуються за допомогою принципу квантування часу світловим променем, що несе в своїй довжині інформацію, яка забезпечує максимально можливі швидкодню, інформативність та завадостійкість.

При цьому відмітимо, що основна вимога до сприймаючого елемента – це забезпечення найвищої точності перетворення параметрів зображень у відеосигнал [2].

1.1. ЗАСТОСУВАННЯ ЛВФ

Основою способу, що пропонується, є квантронне (часове) перетворення (КВР-перетворення) всіх сигналів, що надходять від зображення. Такий підхід дозволяє виконати універсалізацію пристроїв обробки, так як не залежно від типу сигналів всі вони перетворюються на деяку універсальну функцію шляхом квантування. Ця функція й підлягає подальшій обробці.

Необхідно, щоб така функція відповідала наступним вимогам: високе адаптування, згідність у високому темпі та з високою достовірністю максимально просто описувати зображення, яке необхідно розпізнати.

В умовах використання оптоелектронних пристроїв обробки інформації в якості універсальних функцій сигналів доцільно застосовувати функції логіко-часового (ЛЧФ) типу [1]. Такі функції не тільки прості в описанні, тому що їх математичний апарат базується на операціях векторної алгебри, але й дозволяє досягти високої швидкості обробки інформації. Така швидкість забезпечується за допомогою носія інформації – часового фактора.

Таким чином, для формування ЛЧФ виконується плоскостне перетворення у поточний момент часу величини світлового потоку у відповідні тривалості часових інтервалів. Такий підхід до обробки надзвичайно важливий в пристроях око-процесорного типу, які автоматизують процес обробки зображень, особливо в динамічних системах.

1.2. ПЕРЕТВОРЕННЯ ДАНИХ

При паралельному введенні оптичної інформації в пристрій, первинне зображення сканується на апертуру приймального пристрою, тобто кожною окремою чарункою апертури виконується прийом інформації від кожної окремої точки зображення.

При цьому формуються сигнали різноманітних типів, тобто кожен з них містить в собі доволі різноманітні характеристики зображення, яке розпізнається. Всі сигнали, які надійшли на приймальний пристрій, перетворюються за допомогою KVP-перетворення на ЛЧФ, які підлягають попередньої обробці.

Попередня обробка базується на способі паралельного складання часових інтервалів [3]. В даному випадку цей метод використовується як метод формування категорії визначника [4], у відповідності з яким буде виконуватися розпізнавання зображень. Спираючись на те, що спосіб паралельного складання дає можливість відокремлення загальних частин сигналів, з усіх ЛЧФ, що надходять, відокремлюється одночасно (паралельно) декілька загальних частин, кожна з яких характеризує деякий конкретний визначник розпізнавання.

З цього виходить, що спосіб, який пропонується, виконує обробку сигналів для розпізнавання не за одним конкретним визначником, а за деякою кількістю, при чому ця кількість не завдана спочатку, а визначається в процесі попередньої обробки інформації. Одночасно із процесом відокремлення загальних частин ЛЧФ виконується розділення характеристик визначників, що отримуються, на якісні та кількісні характеристики. Кожен з класів визначників подається на окремий канал обробки інформації у відповідності із якістю визначника. В кожному з каналів виконується одночасна кінцева обробка.

2. АНАЛІЗ ІНФОРМАЦІЇ

Як було відмічено вище, запропонований спосіб розпізнавання зображень базується на відокремленні визначників та формуванні їх систем для виконання кінцевого розпізнавання, тобто для визначення типу зображення. Але головною рисою, що відрізняється, цього способу є те, що за рахунок відокремлення одночасно декількох визначників значно підвищується достовірність розпізнавання зображень.

Після закінчення попередньої обробки інформації виконується паралельний аналіз сигналів в двох каналах: якісному та кількісному. Кожен з каналів виконує обробку сигналів з формуванням систем визначників зображень наприкінці аналізу.

В каналі кількісної обробки виконується формування визначників, які описують конкретні фізичні характеристики зображень (маса, вага, об'єм, форма, розміри та інші.).

В каналі якісної обробки виконується синтезування (генерування) визначників. Причому в основі генерації також покладений спосіб паралельного складання [3], в даному випадку виконується багатократне формування загальних частин. Така операція була названа ієрархічним складанням, та як операції паралельного складання здійснюється у відповідності з побудовою ієрархії. Побудова ієрархії

базується на умові достатності формування визначників, які здатні повністю описати зображення при кінцевому розпізнаванні.

Кожен ієрархічний рівень забезпечує генерацію деякого визначника зображення. Таким чином, робимо висновок, що чим більше рівней містить операція ієрархічного складання, тим точніше та повніше буде описано зображення. Для визначення достатності рівней ієрархії та кількості визначників вводиться деяка комутаційна функція, яка у випадку недостатності визначників виконує перекомутацію між окремими операціями ієрархії.

Необхідно відмітити, що при такому підході до відокремлення визначників зображень всі визначники, що отримані, будуть неявно вираженими, тобто це будуть не тільки конкретні фізично визначенні величини, але й їх різноманітні опосередковані інтегровані комбінації.

Після закінчення синтезу (генерації) визначників, яке визначене за допомогою функції комутації, виконується формування систем окремих визначників. Формуються так звані кодери-визначники, які характеризують кожний конкретний визначник. В даному випадку в якості кодеру є ЛЧФ, яка описує властивості визначників за допомогою своєї формуючої функції.

Для виконання операції розпізнавання зображення необхідно створити єдину ЛЧФ системи визначників, яка б об'єднувала в собі всі властивості всіх кодерів визначників. В цьому випадку постає питання про аналітичну залежність єдиної функції розпізнавання від кожного з кодерів окремо при їх одночасному взаємовпливі. Для врахування впливу кожного кодеру один на одного вводиться нова математична функціональна залежність, яка називається оператором впливу. За його допомогою описується формування функції, яка буде кодувати конкретну систему визначників, – цільовий кодер:

$$F = \sum_{i=1}^n a_i \prod_{j=1}^m p_j, \quad (1)$$

де n – кількість вхідних сигналів;

a_i – інформація, яка міститься i -м сигналом;

$\prod_{j=1}^m$ – оператор впливу визначників один на одного;

m – кількість кодерів-визначників, що були синтезовані;

p_j – змінна, яка характеризує фізичний зміст визначника зображення.

З усього вище згаданого витікає, що цільовий кодер набуває якісну залежність від функції оператора впливу, при цьому він зберігає кількісну характеристику зображення. Завдяки впливу одного на одного різноманітних ЛЧФ робимо висновок, що функцію цільового кодеру можна представити у вигляді простіших логічних операцій.

Таким чином, при наявності функцій цільових кодерів кожної системи неявно виражених визнач-

ників з'являється можливість формування кінцевої єдиної ЛЧФ. Така функція у фізичному змісті повинна максимально повно описувати зображення, що розпізнається, формуючи в графічному вигляді своєрідний аналог «ключа-відмички». Це досягається за допомогою того, що змінні признаков p_j під дією оператора впливу діють на різноманітних рівнях суперпозиції, що дозволяє формувати ЛЧФ складної конфігурації (рис. 1).

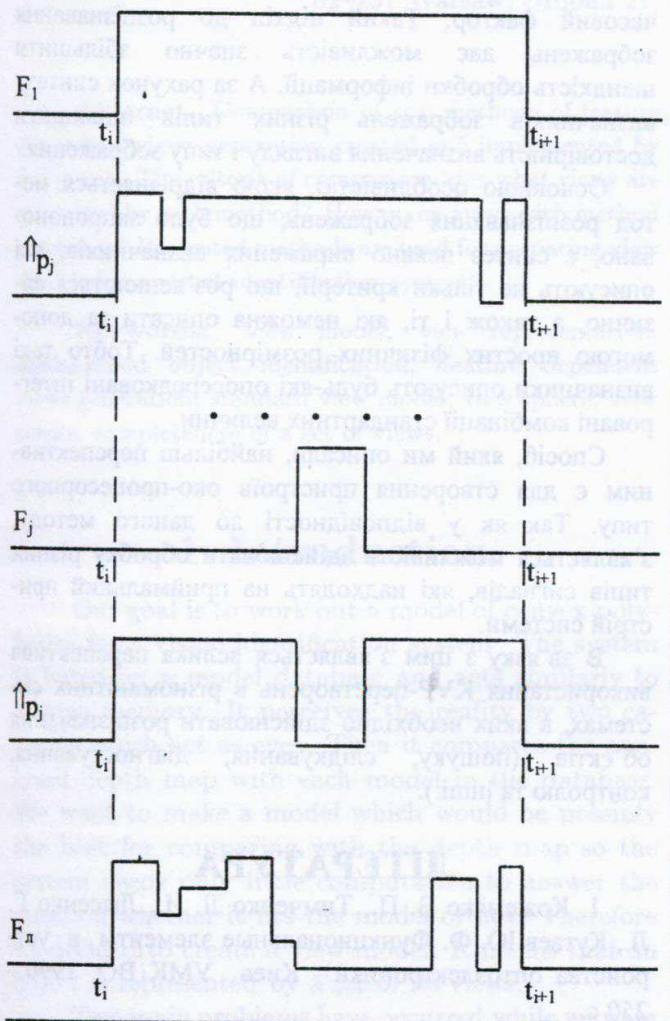


Рис. 1. Формування ключової функції.

Оператор впливу є певною функціональною залежністю, яка діє на конкретному проміжку часу. Операція дії оператора впливу та формування ключової функції описується за допомогою формули:

$$F = \int_m F_j = \int_m \left(\sum_{i=1}^n a_i \hat{p}_j \right), \quad (2)$$

де F_n – зведена інтегрована кількісно-якісна ЛЧФ (ключова функція);

$\int_m F$ – оператор узагальненого інтегрування кіль-

кісного результату паралельних вхідних змінних з визначенням фізичних розмірностей та неявно виражених визначників.

Після закінчення формування ключової функції, яка є унікальним відображенням конкретного

зображення, закінчується етап аналізу та обробки інформації.

3. РОЗПІЗНАННЯ ЗОБРАЖЕННЯ

Базою розпізнавання зображення є операція порівняння результатів, що отримані при обробці, із раніше відомими. Тому припускаємо, що існує деяка база знань, яка містить в собі еталонні зразки сигналів (тобто ЛЧФ) зображень, які обов'язково відносяться до класу зображення, яке необхідно розпізнати.

Перед тим, як приступити до кінцевого розпізнавання зображення, необхідно виконати аналіз цінності результату (ключової функції), тобто впевнитися в інформативності та відповідності сигналу, а саме сигнал має характеризувати зображення, а не сторонні завади чи шуми. У відповідності з цим виконується порівняння функції, що отримана в процесі формування ключової функції, з подібними функціями еталонів. У випадку наявності ідентичного еталону (аналогу) – зображення вважається розпізнаним. У випадку ж, коли аналогу до результату не знайдено розглядаються два можливих випадки: незначні відхилення результату від аналогу та відсутність аналогів взагалі.

В першому випадку необхідно виконати аналіз невідповідності та визначити наступне: чи є згадана невідповідність результатом дії завад при обробці інформації чи незначним змінням зображення з течією часу. Якщо неаналогічність – результат дії завад, то виконується повторне виділення визначників на раніше синтезованих структурах.

У випадку ж неспівпадання з причини часового фактору розпізнавання виконується у відповідності з імовірністю еволюційного розвитку об'єкту, що розпізнається. Якщо при пошуку аналогу в базі знань відсутні відповідні еталони, то у випадку позитивної оцінки цінності результату, новий результат записується в базу знань у якості нового еталону. Таким чином виконується еволюційний розвиток бази знань.

4. ПРАКТИЧНЕ РЕАЛІЗУВАННЯ МЕТОДУ

Для розпізнавання зображень був представлений спосіб KVP-перетворення, але для реалізування його необхідно виконати конструювання пристрою, який буде виконувати розпізнавання зображень на базі цього методу. Для цього необхідно врахувати всі вимоги до обробки зображень, що були представлені вище: до прийому інформації, перетворенню, аналізу, формуванню кінцевого результату (висновку про тип зображення).

Розглянемо припустимий пристрій розпізнавання, яке основане на відокремленні неявно виражених визначників та KVP-перетворенні. З принципу обробки інформації можна сказати, що в якості

основних блоків такого пристрою приймаються: блок перетворення сигналів на ЛЧФ та відокремлення типів визначників, блок синтезатора-генератора визначників та блок визначення типу зображення з системою формування бази еталонних знань.

На першому етапі обробки інформації в пристрої необхідно виконати не тільки перетворення на ЛЧФ сигналів, але й визначити типи визначників, згідно з якими потім буде виконуватися класифікування визначників та розпізнавання зображень. Для цього в якості первинного аналізатору інформації пропонується використати паралельний суматор [5], який базується на принципі паралельного складення.

При детальному аналізі можливостей цього суматора, робимо висновок, що він здатен не тільки виконувати простіші математичні операції (складення, віднімання) з лінійними функціями, але й синтезувати функції з інтегральним описом. Це дозволяє назвати цей суматор більш узагальненою назвою – функційно-інтегральний синтезатор.

Виходячи з принципу паралельного складення, функційно-інтегральний синтезатор виконує відокремлення загальних частин всіх сигналів, що надійшли. На етапі попередньої обробки ці загальні частини характеризують конкретні типи визначників, що дозволяє виключити помилку при отриманні двох ідентичних функцій для різних визначників. Для проведення детального аналізу інформації, що надійшла, до пристрою вводиться два каналу обробки: якісний та кількісний. Як вже говорилося вище, в кожному з каналів виконується обробка «своїх» конкретних визначників.

Основною часткою каналу якісної обробки є синтезатор-генератор [4] визначників зображень. Такий синтезатор-генератор виконує основну операцію відокремлення неявно виражених визначників – ієрархічне складення. Синтез визначників забезпечується завдяки будованню ієрархічної структури, елементарною чарункою якої є функційно-інтегральний синтезатор. Генерація максимально ємкісних визначників зображень передбачає адаптивний синтезатор-генератор самоналагоджуваний на оптимальну структуру розпізнавання для забезпечення максимальної ефективності перекомутації зв'язків між елементарними чарунками, в якості зв'язуючих ланцюгів використовується волоконно-оптична комутація.

Після виконаного аналізу інформації робиться кінцеве розпізнавання об'єктів. Для цього в пристрій вводиться блок визначення типу зображення із системою формування бази еталонних знань. При цьому необхідно врахувати, що для системи формування бази знань потрібна наявність динамічної пам'яті, яка забезпечить не тільки збереження даних, але й забезпечить еволюційний розвиток бази знань.

Найбільш доцільно при такому підході до обробки сигналів в якості динамічної пам'яті використати волоконно-оптичні лінії затримки. Вони

мають забезпечити всі умови для збереження та динамічного передавання даних (точність, швидкість, функціональність).

ВИСНОВКИ

Принцип розпізнавання, що було запропоновано, цікавий тим, що в його основі лежить квантронне перетворення (KVP-перетворення) сигналів, що дозволяє в якості носія інформації використовувати часовий фактор. Такий підхід до розпізнавання зображень дає можливість значно збільшити швидкість обробки інформації. А за рахунок синтезу визначників зображень різних типів підвищити достовірність визначення вигляду і типу зображення.

Основною особливістю, якою відрізняється метод розпізнавання зображень, що було запропоновано, є синтез неявно виражених визначників, які описують не тільки критерії, що роз'яснюються фізично, а також і ті, які неможна описати за допомогою простих фізичних розмірностей. Тобто такі визначники описують будь-які опосередковані інтегровані комбінації стандартних величин.

Спосіб, який ми описали, найбільш перспективним є для створення пристроїв око-процесорного типу. Так як у відповідності до даного методу, з'являється можливість здійснювати обробку різних типів сигналів, які надходять на приймальний пристрій системи.

В зв'язку з цим з'являється велика перспектива використання KVP-перетворень в різноманітних системах, в яких необхідно здійснювати розпізнавання об'єктів (пошуку, слідування, діагностування, контролю та інші.).

ЛІТЕРАТУРА

1. Кожемяко В. П., Тимченко Л. И., Лысенко Г. Л., Кутаев Ю. Ф. Функциональные элементы и устройства оптоэлектроники. – Киев УМК ВО. 1990. – 250 с.
2. Кожемяко В. П. Применение оптоэлектронных логико-временных сред в информационно-вычислительной технике. – К.: Знание. 1988. – 20 с.
3. А. с. СССР № 1119035, МКИ G 06 G 7/14. Способ параллельного сложения длительностей группы временных интервалов./ В. П. Кожемяко, Л. И. Тимченко, Т. В. Головань //Открытия. Изобретения. – 1984. – № 38.
4. Кожемяко В. П., Понура О. И., Кожемяко О. В. Универсальный способ генерации визначников для неалгоритмического распознавания зображень та реалізація його на логіко-часових середовищах. /Вимірювальна та обчислювальна техніка в технологічних процесах. № 2, 1997. – С. 13–19.
5. А. с. СССР № 1101817, МКИ G 06 F 7/50. Устройство для сложения./ В. П. Кожемяко, Л. И. Тимченко, Т. В. Головань, Н. Е. Фурдияк //Открытия. Изобретения. – 1984. – № 25.

Comparison of Two Chosen Methods of Feature Dependent Views Generation

Monika Dąbkowska, Wojciech S. Mokrzycki

ICS PAS, Graphics & Image Processing Group

02-237 Warsaw, Ordona 21, monika, wmokrzyc@ipipan.waw.pl

Abstract. Comparison of two methods of feature dependent views generation created and implemented by our group. The criteria of comparison are: what views are generated by each method? How many views each method generates? Presented methods are used for preparing view models for a visual identification system.

Keywords: view model, view representation, model-based object identification, feature dependent views generation, standard view model, view sphere, view points, completeness of a set of views.

1. Introduction

Our goal is to work out a model of convex polyhedra for a visual identification system. The system is based on a model database and acts similarly to human memory. It perceives the reality by two cameras which act as eyes. Then it compares the analysed depth map with each model in the database. We want to make a model which would be possibly the best for comparing with the depth map so the system needs only little computation to answer the question whether it fits the model or not. Therefore we decided to create a view model. It means that an object is represented by a set of its views.

Two main problems have occurred while we were thinking. The first is how to place the view points? They can be disposed in a uniform way ([12, 15, 25]) or non-regularly in places depending on the object features ([5, 13, 19, 21, 23, 24]). Each object has an infinite number of views but for a computer database we can have only a finite number of views. We need to select as many views as necessary to have a good representation but, on the other hand, as few as possible to occupy a reasonable amount of memory. So, which views should be chosen to satisfy the conditions?

The second problem is to prove that the representation is complete. It means that for each view point the related view is in the database.

2. The View Sphere Concept

To simplify the task and to standardise the views we have decided to limit the set of views only to

these views which can be seen from a sphere (called **the view sphere**), which has the centre O in the centre of the object and its radius R is known, fig. 1. The radius of the view sphere depends on the size of the object and on the angle of the cone of vision 2α (which is a technical parameter of the cameras). The exact relation is described in [13]. This radius is fixed so that the model of an object for which we try to create the views fills the viewing space but also so that maximum elements of the model are visible (not too near and too far) and it is constant during the process of generating views of the model. This implicates a restriction on the cameras (during the process of recognition) which have to be placed in a proper distance from the objects on the scene, because the perspective projection makes extreme elements of the

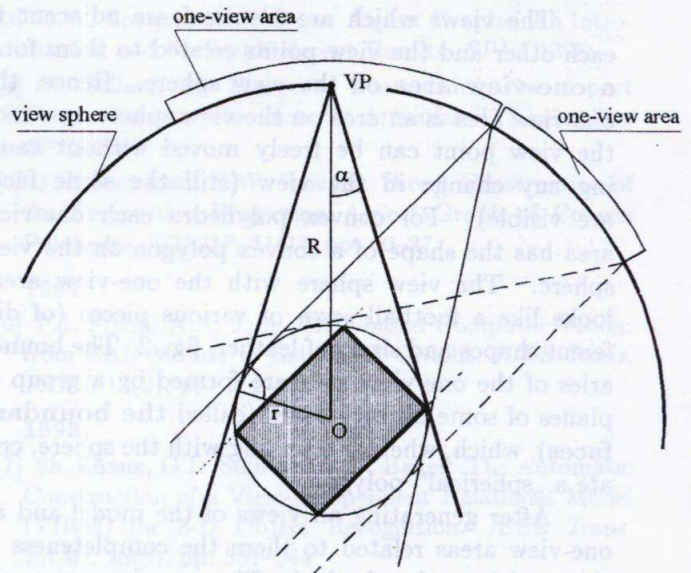


Fig. 1. The object, the cone of vision and the view sphere

visible part of the object disappear while increasing the distance. The distance between the cameras and the objects doesn't have to be exactly equal, but it should be comparable to the radius of the standard view sphere (used during creating the view model) of each object on the scene (it depends on the sizes of the objects and can be easily computed). This condition limits the usage of such a model database but it is still good for systems which can approach the scene,

for example when the cameras are placed above the table and they can be moved down if necessary or in the case of a robot which can decrease the distance to the object.

The set of views which can be seen from the surface of the view sphere is only a set of some standard views because of the constant distance to the centre of the model.

3. The View and the Completeness of the View Model

To begin with we tried to construct a model of convex and opaque polyhedra. We considered only geometrical and not e.g. photometrical features (color, texture...) of the objects. According to this it is sufficient to match the visible faces, edges and vertices of the object with a view of a model to recognise the class of objects. Because the visible edges and vertices are elements of the visible faces, therefore we have defined a **view** as a set of the visible faces of the object. It occurred automatically that many views are represented by only one view, so called the **characteristic view**. The characteristic view can be any view of the group of identical views (identical in the sense of the above mentioned definition).

The views which are identical are adjacent to each other and the view points related to them form a **one-view area** on the view sphere. Hence, the one-view area is an area on the view sphere on which the view point can be freely moved without causing any change in the view (still the same faces are visible). For convex polyhedra each one-view area has the shape of a convex polygon on the view sphere. The view sphere with the one-view areas looks like a football sewn of various pieces (of different shapes and sizes) of leather, fig. 2. The boundaries of the one-view area are formed by a group of planes of some object's faces (called the **boundary faces**), which, when intersecting with the sphere, create a „spherical” polygon.

After generating all views of the model and all one-view areas related to them the completeness of the model can be checked. The one-view areas are helpful here. If the football (view sphere) hasn't any gaps in its covering then it is ready to use (the model is complete), otherwise it can't be used in the game — some new pieces of leather (some new views and new one-view areas related to them) are needed to complete the covering.

The algorithm of views generation ([19]) can be divided into five phases. The first phase is generating and storing of the set of standard views. The second — determining the one-view areas for each stored view. The third — checking of the covering of the view sphere. The fourth — determining of the

gaps in the covering (indicating the gap boundaries). The fifth — generating of the missing views. The last phase isn't yet precisely specified. It may demand dividing the gap into smaller gaps and then generating the views. If it is possible to use here the sequence of instructions from the first phase then the algorithm could be taken in a loop.

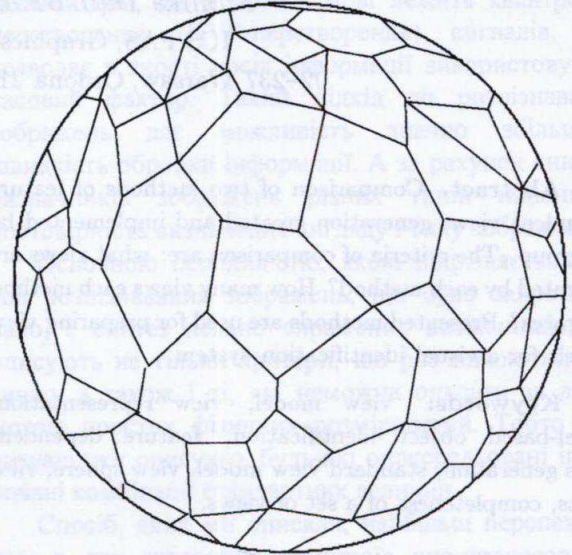


Fig. 2. Covering of the view sphere by the one-view areas

4. The Method of Three Types of Views

As written above we have made an assumption about the objects which can be recognised by the system: they can be convex and opaque polyhedra. According to the geometrical features of such objects — faces, edges and vertices — our first idea was to create the views dependent on those three features of the polyhedral object. The first method consists in determining the views of three types: face-centered, edge-centered and vertex-centered views, [13]. In the first case the viewing direction goes through the centre of the view sphere and the centre of the selected face. The view point lies on the sphere in the place where the so defined viewing direction intersects the sphere. In the case of the edge-centered views the viewing direction goes through the centre of the sphere and the centre of the selected edge. For the vertex-centered views it goes through the vertex. The sense of the viewing vector is always the same — toward the centre of the sphere.

After generating the face-centered view for each face of the polyhedron, the edge-centered view for each edge and the vertex-centered view for each vertex the one-view area for each view should be computed.

The reasoning for one-view areas, checking the covering, determining the gaps and the missing views should be repeated as in section 3.

5. The Method of Dispersing of the View Point

The second method is more accurate. At the beginning the first view is determined and registered. It can be any view computed by the first algorithm (described in section 4). In our implementation the face-centered view was chosen, because it is most possible that it won't be an incident view, [23].

After this the one-view area of that first view is computed. When all boundary faces (of the one-view area) are known, the view point is moved to the other side of each of them sequentially. The aim of this step is to find the neighbouring views for the first view. Firstly, a middling plane for the first view contour edge is computed. The middling plane is a plane containing the viewing direction (view point and the centre of the view sphere — point (0,0,0)) and the centre of the chosen (first) view contour edge. This plane intersecting with the view sphere determines the direction of the view point movement. The view point moves along the big circle formed by the middling plane and the sphere in the direction opposite to the chosen view contour edge. It stops after crossing the nearest boundary face (of the one-view area). Then the new view for the new location of the view point is computed and registered. Afterwards the view point comes back to its previous position, the middling plane for the next view contour edge is computed and the movement of the view point is continued along the middling plane.

When all view contour edges of the first view are used, the next view is chosen from the register and the whole procedure (determining the one-view area and the neighbouring views) is repeated until there are still any views in the register.

To be sure that the set of views is complete, the checking of the covering, as in section 4, should be done and if there are still any gaps, then the missing views should be computed.

6. Comparison

Both methods are implemented and have been tested on the same four objects: hexahedron (regular), pentahedron (saddle roof), heptahedron and octahedron. The results of the methods are in some cases different.

In the case of hexahedron and pentahedron both methods generate the same views. The hexahedron needs 26 views to be well represented (the view sphere is completely covered by the related one-view areas). Logically, the pentahedron, which has less faces than the hexahedron, should need less views: each method has generated the same 20 views of it, which also form the complete model. For more complex (with more faces and without symmetry axes) solids the situation is worse. The first method generates less

views of such solids than the second one. But still it is very frequent that both models of nonsymmetric solids are incomplete. There are still some missing views.

7. Conclusion

Both described methods are efficient for regular or symmetric solids. For nonsymmetric solids the second method is better. It provides more views. To complete both methods an algorithm for detecting boundaries of the gaps and a method of generating the missing views is needed. It is quite possible that for generating missing views the original method of generating views could be used.

References

- 1987
- [1] J.H. Connell, M. Brady: Generating and Generalizing Models of Visual Objects. *AI*, 31, pp. 159–183
- 1989
- [2] C. Hansen, T.C. Henderson: CAGD-Based Computer Vision. *IEEE Trans. PAMI*, 11(11), pp. 1181–1193
- 1990
- [3] K.W. Bowyer, Ch.R. Dyer: Aspect Graphs: An Introduction and Survey of Recent Results. *SPIE*, 1395
- [4] M. Sallam, K. Bowyer: Generalizing the Aspect Graph Concept to Include Articulated Assemblies. *SPIE*, 1395
- [5] J.H. Stewman, K.W. Bowyer: Direct Construction of the Perspective Projection Aspect Graph of Convex Polyhedra. *CVGIP*, 51(1), pp. 20–37
- 1991
- [6] P.J. Flynn, A.K. Jain: CAD-based Computer Vision: from CAD Models to Relational Graphs. *IEEE Trans. PAMI*, 13(2), pp. 114–132
- 1993
- [7] Sh. Zhang, G.D. Sullivan, K.D. Baker: The Automatic Construction of a View-Independent Relational Model (VIRM) for 3-D Object Recognition. *IEEE Trans. PAMI*, 15(6), pp. 531–544
- 1994
- [8] S. Petitjean: Algebraic Geometry and Object Representation in Computer Vision. *Object Representation in Computer Vision. Proc. Int. NSF-ARPA Workshop*, NY, pp. 155–165, Dec.
- 1995
- [9] P.A.R. Cole, M.S. Khan: Modelling 3D Rigid Objects Using the View Signature II Representation Scheme. *Proc. CAIP'95, LNCS*, 970, pp. 154–161, Sep.
- [10] A. Leonardis, S. Kovačič, F. Pernuš: Recognition and Pose Determination of 3D Objects Using Multiple Views. *Proc. CAIP'95, LNCS*, 970, pp. 778–783, Sep.

- [11] W.S. Mokrzycki: System of Stereovisual Perception — SSP. Description of Conception and Implementation. *ICS PAS Reports*, 789, Warsaw (in Polish)
- [12] O. Munkelt: Aspect-Trees: Generation and Interpretation. *Computer Vision and Image Understanding*, 61(3), pp. 365–386
- 1996**
- [13] M. Dąbkowska, W.S. Mokrzycki: Generating Views of 3D Objects for a Visual Identification System. *ICS PAS Reports*, 809, Warsaw (in Polish)
- [14] M. Dąbkowska, W.S. Mokrzycki: Determining One-view Areas for Convex Polyhedra. *ICS PAS Reports*, 823, Warsaw (in Polish)
- [15] V. Hlavač, A. Leonardis, T. Werner: Automatic Selection of Reference Views for Image-Based Scene Representations. 4th European Conference on Computer Vision, Cambridge, UK, Proceedings vol.1, *LNCS*, 1064, pp. 526–535, Apr.
- [16] K. Tarabanis, R.Y. Tsai, A. Kaul: Computing Occlusion-Free Viewpoints. *IEEE Trans. PAMI*, 18(3), pp. 279–292
- 1997**
- [17] M. Dąbkowska, W.S. Mokrzycki: Structures and Completeness of View Models in the Visual Identification System. *ICS PAS Reports*, 833, Warsaw (in Polish)
- [18] M. Dąbkowska: Spherical Representations of Objects for Visual Identification Systems. *Proc. of 3rd Symposium of Micro- and Optoelektronics Institute of Warsaw University of Technology „Technics of Image Processing” TPO'97*, Serock, Oct. 29–31 (in Polish)
- [19] M. Dąbkowska, W.S. Mokrzycki: A Multiview Model of Convex Polyhedra. *MG&V*, 6(4), pp. 419–450
- [20] C. Dorai, A.K. Jain: Shape Spectrum Based View Grouping and Matching of 3D Free-Form Objects. *IEEE Trans. PAMI*, 19(10), pp. 1139–1146
- [21] C.B. Madsen, H.I. Christensen: A Viewpoint Planning Strategy For Determining True Angles On Polyhedral Objects By Camera Alignment. *IEEE Trans. PAMI*, 19(2), pp. 158–163
- [22] I. Shimshoni, J. Ponce: Finite-Resolution Aspect Graph of Polyhedral Objects. *IEEE Trans. PAMI*, 19(4), pp. 315–327
- 1998**
- [23] M. Dąbkowska, W.S. Mokrzycki: A New View Model of Convex Polyhedra with Feature Dependent View. *MG&V*, 7(1/2), Proc. GKPO'98, Borki, Poland, pp. 325–334, May
- [24] M. Dąbkowska, W.S. Mokrzycki: Generating of Convex Polyhedra Views with Direct Tracing of One-View Areas. *ICS PAS Reports*, 861, Warsaw, (in Polish)
- [25] S. Kovačič, A. Leonardis: Planning Sequences of Views For 3D Object Recognition and Pose Determination. *PR*, 31(10), pp. 1407–1417

The Q-Transformation for Recognition of the Facial Images

Y. F. Kutaev¹, V. P. Kozhemyako², L. I. Timchenko², A. A. Gertsy², S. V. Chepornyuk², R. Hamdi².

¹State Scientific Enterprise "Astrophisika", Volokolamskoye shosse 95, Moscow, Russia

²Vinnitsa State Technical University, Khmel'nitskoye Shosse 95, Vinnitsa, 286021, Ukraine,

E-mail: chsv@cec.vstu.vinnica.ua, timchenko@yahoo.com

ABSTRACT

Both the methodology and results of experimental research performed for the facial images normalization by means of compact models description that describe image classes are considered in this paper. The problems of preliminary image processing based on the method of generalized Q-transformation are being solved. The image segmentation with the formation of connectivity matrices and the formal description of the resulting components are taken into consideration as well. The calculating algorithms based on the methodology of dichotomous balance of the images being prepared have been used for the classification of human facial images. The high efficiency of recognition caused by the compact description of object images was obtained as a result and in this way it helps to make the processing more noiseproof, to improve invariance to the 2D and 3D- transformations and to reduce the calculation time requirements.

1. INTRODUCTION

During the last 15 years the problem of facial recognition has been given more and more attention by researchers. The latest improvements in this field use a great variety of methods. Although each method has its own specific quality, the reliable recognition has not been practically achieved yet, even for the narrow range of conditions. The basic specific problems to be solved are as follows: 3D rotations, the invariance to facial expression, various sights and the integration of information from different scales.

The main goal of this research is to solve the problem of the integration of the high-level information into the low-level facial presentation [1]. This will result in higher recognition efficiency. The advantage of such recognition is compact face presentation which considers the reduced size of the data sets and faster efficiency.

A human being tends to concentrate upon some high level peculiarities while recognizing unfamiliar faces. On the contrary, most automatic systems process faces as a whole. Kirby and Sirovich [2], Lanitis and others [3] performed the analysis of principal components considering a face as a whole. Blackwell and others [4] noted that the preliminary processing of the whole image cannot solve the problems connected with

the integration of large data sets. The other system class processes located descriptors. Lades and others [5] recognized faces without integrating the regularity of a facial class. Penev and Atick [6] performed the analysis of a local peculiarity on the basis of each human's individual view.

2. METHODOLOGY

The general methodology of the compact image presentation is given in accordance with the scheme of *coarse - fine* processing. The coarse processing consists of image quantization and is followed by its division into spatially-connected segments [7]. The results of this coarse processing are separate segments structured in some fields according to their connectivity indexes. These image segments are formed from spatially connected pixels of a quantified image and then create connectivity fields. In the case of a human face, coarse processing these fields can reflect its distinctive components (eyes, cheeks, mouth, etc.). Then separate segments undergo the *fine* processing scheme. In this case, the processing of the segments which are less deformed because of different facial expressions becomes possible.

2.1 The fine scheme of image processing.

The fine processing scheme is based upon the separate segment analysis and includes image preparing, i.e. transforming the image into the matrix of 3-level binary preparations (positive, negative and zero) and their subsequent transformation by means of forming local and general equalization functions. Further correlation analysis of those functions is provided as well. The preliminary processing includes all subsequent operations. The input 2D grey-scale image is presented by the matrix of light intensity pixels $A = [a_{i,j}]$, $i = 1 \div N$, $j = 1 \div M$, where $M \times N$ is the input image dimensions. The general mean of the present matrix A is found as:

$$\bar{a} = \frac{1}{NM} \sum_{i,j} a_{i,j} \quad (1)$$

Then sets of differences between pixels and obtained means of the whole image or its segments, are determined:

$$R_{i,j} = a_{i,j} - \bar{a} \quad (2)$$

In order to obtain the set of preparations this differences are compared with the threshold δ [8, 9]:

$$q_{i,j} = \begin{cases} 1, & \text{if } R_{i,j} > \delta \\ -1, & \text{if } R_{i,j} < (-\delta) \\ 0, & \text{if } |R_{i,j}| \leq \delta \end{cases} \quad (3)$$

The experiments have shown that the usage of the function that equally distributes positive, negative and zero preparations is most useful for the facial image preparing. It makes it possible to mark areas with and without brightness differences (according to the threshold) in a proper way. The threshold itself is adapted to the image. It results in segmenting separate fields of an initial image for their further analysis. The threshold δ is chosen from the following condition [9]:

$$\text{MAX}(N_t^{(1)}, N_t^{(-1)}, N_t^{(0)}) \quad (4)$$

where $N_t^{(1)}, N_t^{(-1)}, N_t^{(0)}$ is the number of positive, negative and zero preparations within each grey level $-t$.

The matrix of preparations $Q^x = [q_{i,j}^x]$ is formed as a result of the above mentioned operations.

2.2 The image segmentation.

In order to get certain features [7] from the obtained matrices the connectivity indexes are determined separately for positive, negative and zero preparations. The stages of image segmentation in accordance to the meaning of their connectivity indexes are mentioned below.

The general connectivity index of the whole prepared image is found. The connectivity spectrum of the image with the $M*N$ -dimensions is defined as

$$W = \sum_{v=1}^8 \sum_{i,j} q_{i,j}^v$$

where $q_{i,j}^v$ is an element of the image with ij -coordinates and v -connectivity.

2.3 The formal description of image parts.

The following decomposition based on dichotomous principle could be used to analyse the obtained connectivity spectrum. Let us divide the whole prepared matrix in a column (row) direction so that two

divided parts will have equal connectivity indexes. The operation of column (row) equalization should be introduced for each level of this dichotomous decomposition. The local column equalization function $U^* = [u_j^*]$ has to meet the following requirement:

$$\sum_{i=1}^{u_j^*} q_{i,j}^v = \sum_{i=u_j^*+1}^N q_{i,j}^v, \quad u_j \in \{1, 2, \dots, N\} \quad (6)$$

Then, the general column equalization function $U^{**} = [u_j^{**}]$ is determined as follows:

$$\sum_{j=1}^{u_j^{**}} \sum_{l=1}^j q_{i,j}^v = \sum_{j=u_j^{**}+1}^N \sum_{l=1}^j q_{i,j}^v \quad (7)$$

Besides, this equalization process is performed for each row and column of the prepared image. Furthermore, the obtained connectivity indexes are added to each of the divided fields.

The equalization procedure is carried out during each step of this decomposition. The operations being described are performed separately for positive, negative and zero preparations. Then a classifying analysis method has to be used as described in [10] for recognition and identification of human faces with the usage of operations based on the network structure.

3. EXPERIMENTAL RESULTS.

Two basic noise distributions most common in the tasks of image processing - Gaussian (normal) distribution and uniform noise distribution have been chosen to analyse a noise invariancy of this method. 7 different gradations with such levels as 0%, 5%, 10%, 15%, 20%, 25%, 30% were organized for the experiment being held. To implement image preparing, the proper threshold value was selected from the range

$$\delta \in \left\{ 1, 2, \dots, \frac{I_{\max}}{2} \right\}, \quad I_{\max} = 255 \text{ gray scale levels.}$$

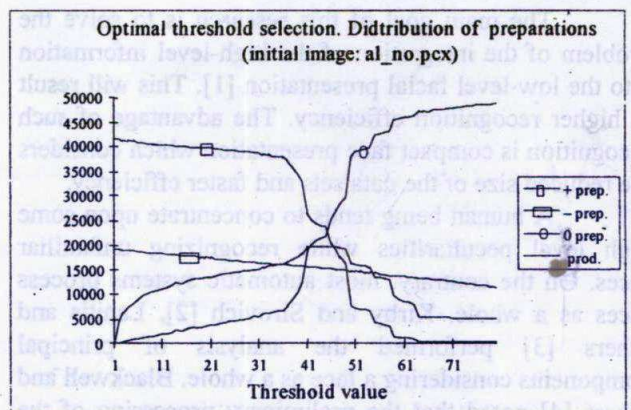


Fig.1 The preparation distributions for the initial image

Those threshold values that exceed this range do not meet the above mentioned requirements because the number of zero preparations prevails. Image preparing was performed as a test for each current value of the threshold with simultaneous calculation of positive, negative and zero preparation. The preparation distributions for the initial image are shown in Fig. 1 and 2.

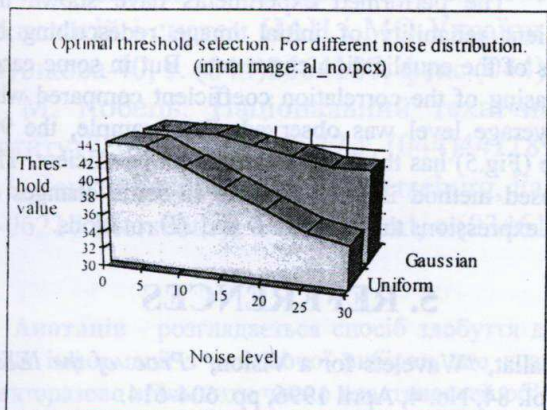
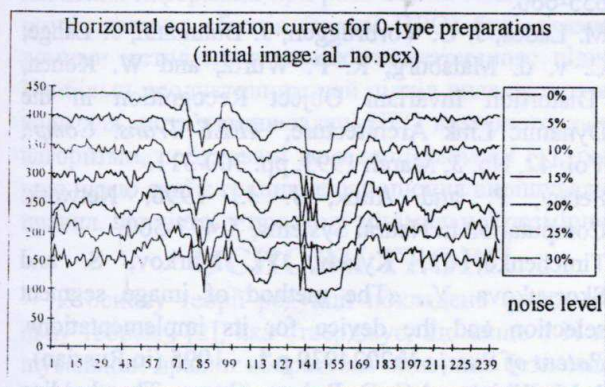
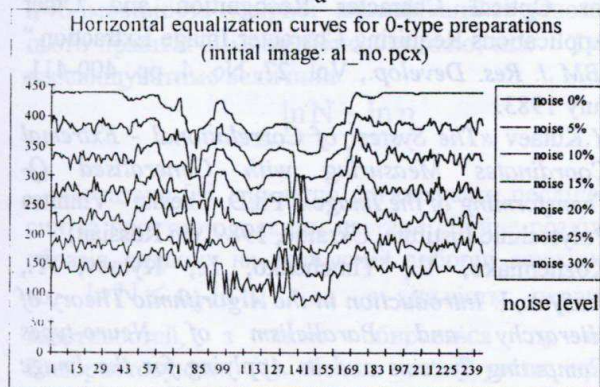


Fig.2. The distributions of optimum threshold for noise-added images

Fig. 1 shows that the optimum value for the initial image is 44. The analogous operations for all noise levels were performed during the experiments. The higher noise level causes decreasing the threshold value, especially in case of the uniform noise distribution. The vertical and horizontal equalization curves were obtained for the initial and noise-added images during the experiment.



a

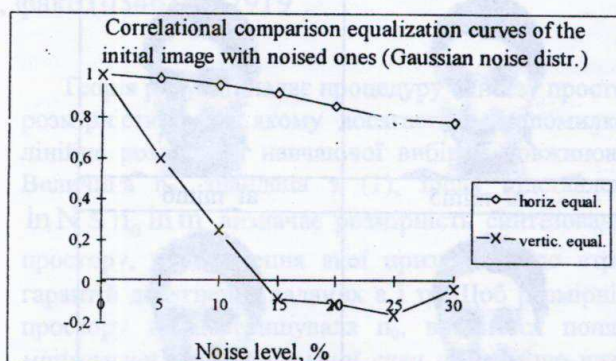


b

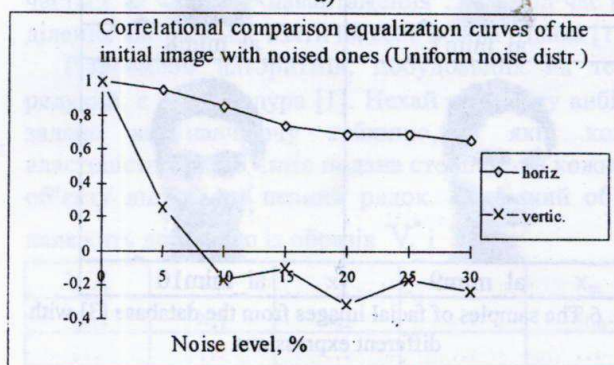
Fig.3 Horizontal equalization curves. a) Gaussian noise distribution; b) Uniform noise distribution

The resulting curves with offset are shown in Fig.3, where the highest curve corresponds with the initial image without noise. The results of the correlational comparison of the initial and noise-added image curves can be found in Fig. 4.

The correlative comparison of equalization curve sets (Fig.5) obtained in the row direction was carried out for different facial expressions. These images are shown in Fig.6. The image with normal expression (al_no1) was used as a reference for further comparison with other images which contain 9 other facial expressions and 3D rotations of the same face (al_mim2..al_mim10). The optimal thresholds were obtained for these images. The results of correlational comparison are shown in Fig. 7.



a)



b)

Fig.4. Results of the correlational comparison: a) Gaussian noise distribution; b) Uniform noise distribution

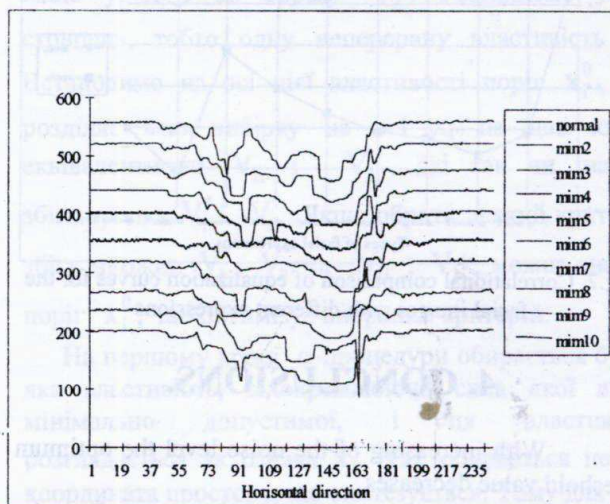


Fig.5 Horizontal equalization curves of zero-preparations for the facial images with different expressions (see Fig. 6)

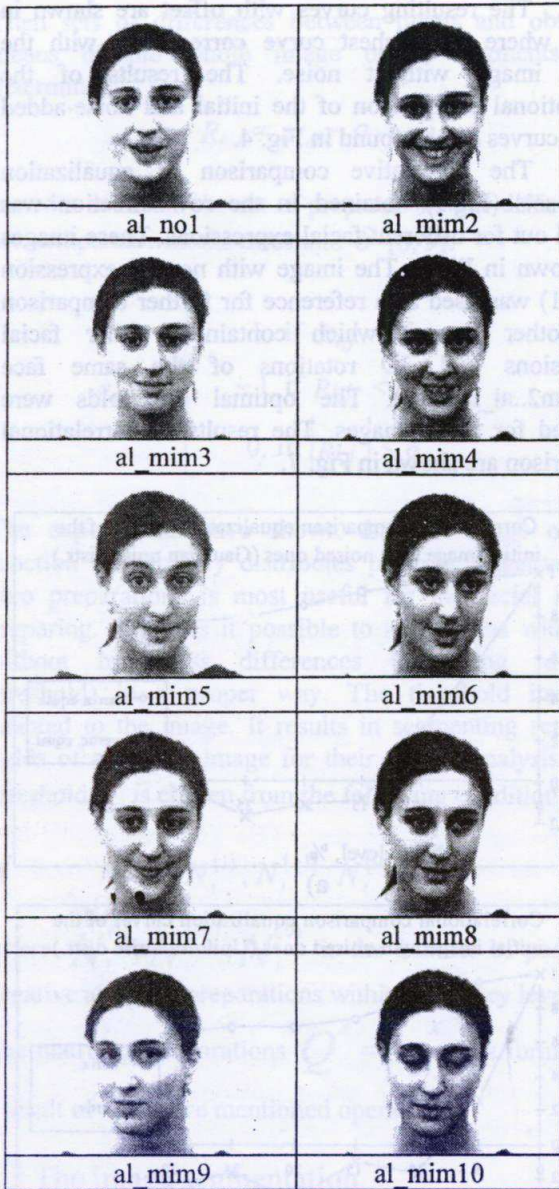


Fig. 6 The samples of facial images from the database [3] with different expressions

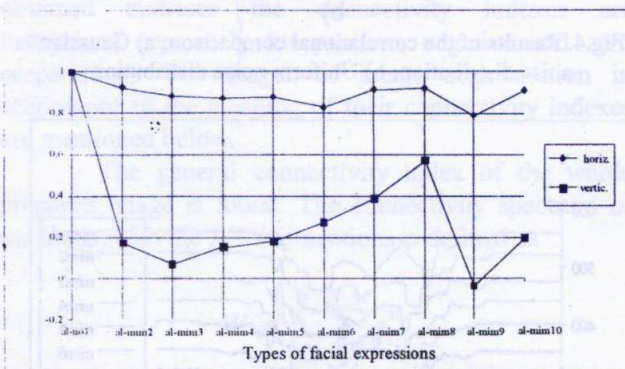


Fig. 7. Correlational comparison of equalization curves for the facial images with different expressions

4. CONCLUSIONS.

With increasing of the noise level the optimum threshold value decreases.

The horizontal equalization curves reflect the peculiarities of facial images much better. In further investigations, the horizontal equalization operations will be used in tasks of recognition and identification.

In the case of the noise level having been increased to 30 %, the correlation coefficient remains rather high - 0.8 (Gaussian noise) and 0.65 (uniform noise); that is caused by the usage of operations of addition which minimize the noise influence.

Due to the transformation of the initial 2D image into a 1D equalization curve set the further correlational calculations becomes faster.

The performed experiments have shown the sufficient reliability of initial image redescribing by means of the equalization curve sets. But in some cases decreasing of the correlation coefficient compared with the average level was observed. For example, the 9th image (Fig.5) has the lowest correlation coefficient. The proposed method is more reliable to some changes of facial expressions than to the 2D and 3D rotations.

5. REFERENCES

1. Mallat, "Wavelets for a Vision," *Proc. of the IEEE*, Vol. 84, No. 4, April 1996, pp. 604-614.
2. M. Kirby and L. Sirovich, "Application of the Karhunen-Loeve Procedure for the Characterization of Human Faces," *IEEE Trans. Pattern Anal. Machine Intell.*, Vol. PAMI-12, 1990.
3. A. Lanitis, C. J. Taylor, and T. F. Cootes, "Automatic Face Identification System Using Flexible Appearance Models," *Image and Vision Computing*, Vol. 13, No. 5, June 1995, pp. 393-401.
4. Blackwell, K. T., Vogl, T.P., Hyman, S.D., Barbour, G.S., and Alkon, D.L. 1992. *Pattern Recognition*, 25, 655-666.
5. M. Lades, J. C. Vorbruggen, J. Buhmann, J. Lange, C. v. d. Malsburg, R. P. Wurtz, and W. Konen, "Distortion Invariant Object Recognition in the Dynamic Link Architecture," *IEEE Trans. Comp.*, Vol. 42, No. 3, March 1993, pp. 300-311.
6. Penev, P. and Atick, J. J., 1996, *Network: Computation in Neural Systems*, 7, 477-500.
7. Timchenko, L., Kytaev, Y., Markov, S. and Skoryukova, Y., «The method of image segment selection and the device for its implementation». *Patent of Russia №2024939 p.1. - 1995.* (in Russian).
8. J. M. White and G. D. Rohrer, "Image Thresholding for Optical Character Recognition and Other Applications Requiring Character Image Extraction," *IBM J. Res. Develop.*, Vol. 27, No. 4, pp. 400-411, July 1983.
9. Y. Kutaev «The System of Correlational - Extremal Coordinates Measuring with Generalised Q-Transforming of the Image». Ph.D. Thesis: - Vinnitsa Polytechnic Institute, Ukraine, 1989, (in Russian)
10. Kozhemiako, V., Timchenko, L., Kytaev, Y., Ivasyuk, I. *Introduction in the Algorithmic Theory of Hierarchy and Parallelism of Neuro-types Computing Systems and its Applying for the Image Transformation. Part II.* - Kiev, Ukraine, 1994, pp 65-205 (in Russian)