
Bibliography

- [Fernández, 2005] L.Fernández, V.J.Martínez, F.Arroyo, L.F.Mingo, *A Hardware Circuit for Selecting Active Rules in Transition P Systems*, Workshop on Theory and Applications of P Systems. Timisoara (Romania), september, 2005.
- [Fernández, 2006] L.Fernández, F.Arroyo, J.Castellanos, J.A.Tejedor, I.García, *New Algorithms for Application of Evolution Rules based on Applicability Benchmarks*, BIOCOMP06 International Conference on Bioinformatics and Computational Biology, Las Vegas (USA), july, 2006 (accepted).
- [Paun, 1998] Gh.Paun, *Computing with Membranes*, Journal of Computer and System Sciences, 61(2000), and Turku Center of Computer Science-TUCS Report nº 208, 1998.
- [Paun, 2005] Gh.Paun, *Membrane computing. Basic ideas, results, applications*, Pre-Proceedings of First International Workshop on Theory and Application of P Systems, Timisoara, Romania, September 26-27, 2005, 1-8
- [Rasoul, 1991] S.R.Safavian, D.Landgrebe, *A Survey of Decision Tree Classifier Methodology*, IEEE Transactions on Systems, Man, and Cybernetics, Vol. 21, No. 3, pp 660-674, May 1991.
-

Authors' Information

Luis Fernandez – Natural Computing Group of Universidad Politécnica de Madrid (UPM); Ctra. Valencia, km. 7, 28031 Madrid (Spain); e-mail: setillo@eui.upm.es

Fernando Arroyo – Natural Computing Group of Universidad Politécnica de Madrid (UPM); e-mail: farroyo@eui.upm.es

Ivan Garcia – Natural Computing Group of Universidad Politécnica de Madrid (UPM); e-mail: igarcia@eui.upm.es

Gines Bravo – Natural Computing Group of Universidad Politécnica de Madrid (UPM); e-mail: gines@eui.upm.es

A PARTITION METRIC FOR CLUSTERING FEATURES ANALYSIS

Dmitry Kinoshenko, Vladimir Mashtalir, Vladislav Shlyakhov

Abstract: *A new distance function to compare arbitrary partitions is proposed. Clustering of image collections and image segmentation give objects to be matched. Offered metric intends for combination of visual features and metadata analysis to solve a semantic gap between low-level visual features and high-level human concept.*

Keywords: *partition, metric, clustering, image segmentation.*

ACM Classification Keywords: *1.5.3 Clustering - Similarity measures*

Introduction

There has been a tremendous growth of the image content analysis significance in the recent years. This interest has been motivated mainly by the rapid expansion of imaging on the World-Wide Web, the availability of digital image libraries, increasing of multimedia applications in commerce, biometrics, science, entertainments etc. Visual contents of an image such as color, shape, texture and region relations play dominating role in propagation of feature selection, indexing, user query and interaction, database management techniques. Many systems combine visual features and metadata analysis to solve a semantic gap between low-level visual features and high-level human concept, i.e. there arises a great need in self-acting content-based image retrieval task-level systems.

To search images in an image database traditionally queries 'ad exemplum' are used. In this connection essential efforts have been devoted to synthesis and analysis of image content descriptors. However, a user's semantic understanding of an image is of a higher level than the features representation. Low-level features with mental concepts and semantic labels are the groundwork of intelligent databases creation. Short retrieval time

independent of the database size is a fundamental requirement of any user friendly content-based image retrieval (CBIR) system. Characteristics of different CBIR schemes, similarities or distances between the feature vectors of the query by example or sketch and those of the database images are sufficiently full explored [see, e.g. 1-3]. To optimize CBIR schemes it is necessary to minimize a total number of matches at a retrieval stage. Thus there arises a problem of finding novel partition measures for the fast content-based image retrieval in video databases and we have to be able to compare different partitions obtained for instance as a result of segmentation.

Motivation to Synthesis of a Partition Metric

As retrieval is computationally expensive, one of the most challenging moments in CBIR is minimizing of the retrieval process time. Widespread clustering techniques allow to group similar images in terms of their features proximity. The number of matches can be greatly reduced, but there is no guarantee that the global optimum solution is obtained. We propose clustering of image collections with objective function encompassing goals to number of matches at a search stage.

The problem is in that under given query $y \in Y$ one needs to find the most similar image (or images) $x_v \in X$. In other words, it is necessary to provide $\min_{v \in V} \rho(y, x_v)$ (here $\rho(o, o)$ is arbitrary distance function, V is an indexing set) during minimum possible warranted time. If $Y \subseteq X$, retrieval by exact matching is required. We shall name elements $[X]_\alpha, \alpha \in A$ of power set 2^X as clusters if they correspond to the partition of set X . Let us consider such partitions that any elements of one cluster do not differ from each other more than on ε , i.e. $\forall x' \neq x''$ we have $[x'] = [x'']$, if $\rho(x', x'') \leq \varepsilon$ and $[x'] \cap [x''] = \emptyset$ otherwise. The given or obtained value ε used at a clustering stage is connected with required accuracy of retrieve δ , if it is specified, as follows. There arise two cases:

$\delta > \varepsilon$ – any representative of the cluster nearest to the query y can be used as the image retrieval result, i.e. minimal number of matches is defined by the number of clusters; in other words it is necessary to provide

$$N_1 = \text{card} \{[X]_\alpha\} \rightarrow \min ; \tag{1}$$

$\delta \leq \varepsilon$ – the element of more detailed partition will be the result of the image retrieval. In simplest situations it is necessary to fulfill a single-stage clustering, i.e. to optimize retrieval under worst-case conditions we have to ensure

$$N_2 = \text{card} \{[X]_\alpha\} + \max(\text{card} [X]_\alpha) \rightarrow \min . \tag{2}$$

At the multilevel clustering the repeated clusters search inside of already retrieved clusters is fulfilled and only on the last step required image is searched by complete enumeration. Let us assume that the cluster $[X^{(i-1)}]_p$ is selected on $(i-1)$ level of hierarchy from a condition $\rho(y, [X^{(i-1)}]_q) \rightarrow \min, q = \overline{1, \text{card} \{[X^{(i-1)}]\}}$, i.e. $[X^{(i-1)}]_p = [X^{(i)}]_1 \cup [X^{(i)}]_2 \cup \dots \cup [X^{(i)}]_{\alpha_p}$ where for any k and l the equality $[X^{(i)}]_k \cap [X^{(i)}]_l = \emptyset$ holds. Then the minimization of matches amount is reduced to the clustering with the goal function

$$N_3 = \sum_{i=1}^{m-1} \{ \text{card} [X^{(i)}]_{p,(i)} \mid x \in [X^{(i-1)}]_{p,(i-1)} \} + \max(\text{card} [X^{(m-1)}]_{p,(m-1)}) \rightarrow \min , \tag{3}$$

where m is a number of hierarchy levels, $[X^{(0)}]_{1,(0)} = X$. Minimization task (1) was solved in [4], the solution of problem (2) was offered in [5], searching of (3) one could see in [6].

Content of an image may be often summarized by a set of homogeneous regions in appropriate feature space. Therefore, there is a great need for automatic tools to classify and retrieve image content on the base of segmentation.

Segmented images are formed from an input image by gathering its elements into sets likely to be associated with meaningful objects in the scene. That is, the main segmentation (clustering) goal is to partition the entire image into disjoint connected or disconnected regions. Unfortunately, the effectiveness of their direct interpretation depends heavily on the application area and characteristics of an acquisition system. Possible high-

level region-based interpretations are associated with a priori information, measurable region properties, heuristics, plausibility of computational inference. Whatever the case, often it is necessary to have dealings with a whole family of partitions and we must be able to compare these partitions, which are produced by a variety of segmentation algorithms. At least splitting and merging techniques make us to match segmentation results, which ultimately may be corresponded to indirectly images comparisons.

For region-based similarity analysis novel approaches are required since usually early processing scheme consists of following steps: images are segmented into disjoint (or weakly intersecting) regions, features are extracted from each region, and the set of all features is used for high-level processing. It should be emphasized that quite often simultaneous processing of partitions or coverings is wanted to produce reliable true conclusion. In this connection we propose and vindicate a new metric providing arbitrary partitions (and consequently images) matching. Generally, these partitions are results of any clustering procedures.

A Metric for Partitions Matching

Let Ω be arbitrary measurable set with given measure $\mu(\circ)$, which can be interpreted as length, area, volume, mass distribution, probability distribution, cardinality, etc. Consider a set Π_Ω of finite (from the standpoint of the factor set cardinality) partitions of Ω , i.e. $\alpha \in \Pi_\Omega$, $\alpha = \{X_1, X_2, \dots, X_n\}$, $X_i \subseteq \Omega$, $i = \overline{1, n}$, $\Omega = \bigcup_{i=1}^n X_i$, $\forall i, j \in \{1, 2, \dots, n\} : i \neq j \Rightarrow X_i \cap X_j = \emptyset$, $\forall i \in \{1, 2, \dots, n\} \Rightarrow \mu(X_i) < \infty$. We denote subsidiary partitions which will be used in further proofs as $\beta = \{Y_1, Y_2, \dots, Y_m\}$ and $\gamma = \{Z_1, Z_2, \dots, Z_l\}$.

Note, these partitions could be image segmentation results, representing pairwise disjoint family of nonempty subsets whose union is the image and each subset may contain required target, may belong to a carrier of object image or may be a part of it. Generally, partitions are results of arbitrary clustering.

Let us introduce on $\Pi_\Omega \times \Pi_\Omega$ the functional

$$\rho(\alpha, \beta) = \sum_{i=1}^n \sum_{j=1}^m \mu(X_i \Delta Y_j) \mu(X_i \cap Y_j) \tag{4}$$

(here $X_i \Delta Y_j = (X_i \setminus Y_j) \cup (Y_j \setminus X_i)$ is a symmetric difference) and prove that the functional $\rho(\alpha, \beta)$ may be interpreted as a distance function for partitions matches. Before we verify that (4) is a metric we shall begin with subsidiary formal propositions.

Lemma 1. The functional $\rho(\alpha, \beta)$ can be represented in the tantamount form

$$\rho(\alpha, \beta) = \sum_{i=1}^n [\mu(X_i)]^2 + \sum_{j=1}^m [\mu(Y_j)]^2 - 2 \sum_{i=1}^n \sum_{j=1}^m [\mu(X_i \cap Y_j)]^2 . \tag{5}$$

Proof. It should be noted that for arbitrary measurable sets P and Q the equality

$$\mu(P \Delta Q) = \mu(P) + \mu(Q) - 2\mu(P \cap Q) \tag{6}$$

takes place. Indeed, from the definition of a symmetric difference we directly get $P = (P \setminus Q) \cup (P \cap Q)$ and $Q = (Q \setminus P) \cup (Q \cap P)$ where sets $P \setminus Q$ and $P \cap Q$ do not intersect. Then by virtue of measure $\mu(\circ)$ is additive we have

$$\begin{cases} \mu(P) = \mu(P \setminus Q) + \mu(P \cap Q), \\ \mu(Q) = \mu(Q \setminus P) + \mu(P \cap Q). \end{cases} \tag{7}$$

Adding up equations (7) and taking into account that $P \Delta Q = (P \setminus Q) \cup (Q \setminus P)$ we obtain

$$\mu(P) + \mu(Q) = \mu(P \setminus Q) + \mu(Q \setminus P) + 2\mu(P \cap Q) = \mu(P \Delta Q) + 2\mu(P \cap Q). \tag{8}$$

It is clear that (8) and (6) are equivalent equations.

With due regard (6) we could rewrite (4) with reference to X_i and Y_j as

$$\begin{aligned} \rho(\alpha, \beta) &= \sum_{i=1}^n \sum_{j=1}^m \mu(X_i \cap Y_j) [\mu(X_i) + \mu(Y_j) - 2\mu(X_i \cap Y_j)] = \\ &= \sum_{i=1}^n \sum_{j=1}^m \mu(X_i) \mu(X_i \cap Y_j) + \sum_{i=1}^n \sum_{j=1}^m \mu(Y_j) \mu(X_i \cap Y_j) - 2 \sum_{i=1}^n \sum_{j=1}^m [\mu(X_i \cap Y_j)]^2 = \quad (9) \\ &= \sum_{i=1}^n \mu(X_i) \sum_{j=1}^m \mu(X_i \cap Y_j) + \sum_{j=1}^m \mu(Y_j) \sum_{i=1}^n \mu(X_i \cap Y_j) - 2 \sum_{i=1}^n \sum_{j=1}^m [\mu(X_i \cap Y_j)]^2 . \end{aligned}$$

In (9) we distinguish components $\sum_{j=1}^m \mu(X_i \cap Y_j)$ and $\sum_{i=1}^n \mu(X_i \cap Y_j)$. Emphasize sufficiently obvious but significant property of similar sums for all $i \in \{1, 2, \dots, n\}$, $j \in \{1, 2, \dots, m\}$

$$\begin{cases} \sum_{j=1}^m \mu(X_i \cap Y_j) = \mu(X_i), \\ \sum_{i=1}^n \mu(X_i \cap Y_j) = \mu(Y_j). \end{cases} \quad (10)$$

The correctness of (10) immediately follows from fig. 1 (according to $\beta, \alpha \in \Pi_\Omega$ and measure $\mu(\circ)$ additivity).

Substituting (10) into (9) we finally get

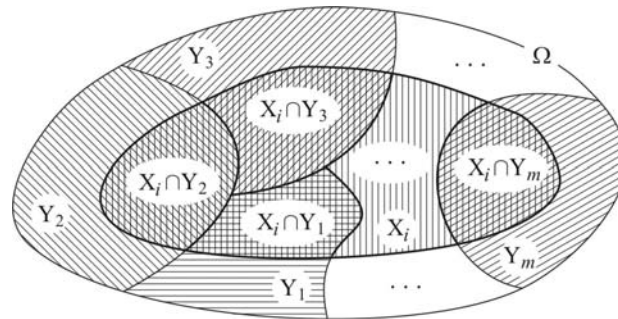


Fig. 1. To explanation of any factor set element intersections with any partition

$$\rho(\alpha, \beta) = \sum_{i=1}^n \sum_{j=1}^m \mu(X_i \Delta Y_j) \mu(X_i \cap Y_j) = \sum_{i=1}^n [\mu(X_i)]^2 + \sum_{j=1}^m [\mu(Y_j)]^2 - 2 \sum_{i=1}^n \sum_{j=1}^m [\mu(X_i \cap Y_j)]^2,$$

which was required, i.e. we get equivalent form of functional (4).

Let us introduce triple intersections of partition elements into consideration, viz sets of following kind

$$X_i \cap Y_j \cap Z_k, \quad i = \overline{1, n}, \quad j = \overline{1, m}, \quad k = \overline{1, l}.$$

Lemma 2. Values $v_{ijk} = \mu(X_i \cap Y_j \cap Z_k)$ are in accordance with the equalities

$$\mu(Z_k) = \sum_{i=1}^n \sum_{j=1}^m v_{ijk} , \quad (11)$$

$$\mu(X_i \cap Y_j) = \sum_{k=1}^l v_{ijk} , \quad (12)$$

$$\mu(X_i \cap Z_k) = \sum_{j=1}^m v_{ijk} , \quad (13)$$

$$\mu(Y_j \cap Z_k) = \sum_{i=1}^n v_{ijk} . \quad (14)$$

Proof. First let us consider $\alpha, \beta \in \Pi_\Omega$. Notice that $\alpha \cap \beta = \{X_i \cap Y_j\}_{i=\overline{1, n}, j=\overline{1, m}} \in \Pi_\Omega$, i.e. an intersection of partitions is a partition also. Indeed, if we choose arbitrary element $\omega \in \Omega$ we get that, owing to α, β are partitions, there exist indices $i \in \{1, 2, \dots, n\}$, $j \in \{1, 2, \dots, m\}$ s.t. $\omega \in X_i$ and $\omega \in Y_j$ or $\omega \in X_i \cap Y_j$, i.e.

$$\bigcup_{i=\overline{1, n}, j=\overline{1, m}} (X_i \cap Y_j) = \Omega. \quad (15)$$

On the other hand, we can write $(X_i \cap Y_j) \cap (X_{i'} \cap Y_{j'}) = (X_i \cap X_{i'}) \cap (Y_j \cap Y_{j'})$ due to virtue of set intersection associativity and commutativity. Since pairs (i, j) and (i', j') do not coincide, inequalities $i \neq i'$ and

$j \neq j'$ hold separately or both. Thereby at least one of sets $X_i \cap X_{i'}$ or $Y_j \cap Y_{j'}$ is empty one as they belong to partitions α and β correspondingly. As a result we have

$$\forall (i, j), (i', j') \in \{1, 2, \dots, n\} \times \{1, 2, \dots, m\} \Rightarrow (X_i \cap Y_j) \cap (X_{i'} \cap Y_{j'}) = \emptyset. \tag{16}$$

The validity of (15), (16) suggests correctness of the inclusion $\alpha \cap \beta \in \Pi_\Omega$, which was required. Now applying first equality from (10) to the sets Z_k and $(X_i \cap Y_j)$ respectively we obtain

$$\mu(Z_k) = \sum_{i=1}^n \sum_{j=1}^m [Z_k \cap (X_i \cap Y_j)] = \sum_{i=1}^n \sum_{j=1}^m v_{ijk},$$

i.e. equality (11). Absolutely analogously using by pairs sets $X_i \cap Y_j, X_i \cap Y_k, Y_j \cap Z_{kj}$ and γ, β, α pro tanto we get equalities (12)–(14), q.e.d.

Now we are ready to prove that $\rho(\alpha, \beta)$ is the distance function.

Theorem. Let Ω be arbitrary measurable set with given measure $\mu(\circ)$ and let Π_Ω be a set of its partitions then for arbitrary $\alpha, \beta \in \Pi_\Omega$ functional (4) is a metric.

Proof. To prove the theorem it is sufficiently to show that (4) is nonnegative and satisfies axioms of reflexivity, symmetry and the triangle inequality.

Note, the nonnegativity and the symmetry directly follow from the definition.

To prove reflexivity it is necessary to verify that $\rho(\alpha, \beta) = 0 \Leftrightarrow \alpha = \beta$. First we consider $\rho(\alpha, \alpha)$ for arbitrary $\alpha \in \Pi_\Omega$. In accordance with the symmetry property we have

$$\rho(\alpha, \alpha) = \sum_{i=1}^n \sum_{j=1}^n \mu(X_i \Delta X_j) \mu(X_i \cap X_j) = \sum_{i=1}^n \mu(X_i \Delta X_i) \mu(X_i \cap X_i) + 2 \sum_{\substack{i, j=1 \\ i > j}}^n \mu(X_i \Delta X_j) \mu(X_i \cap X_j).$$

First term from the right consists of n summands with the identical partition elements and so it is equal to zero as $\mu(X_i \Delta X_i) = 0$ for all $i = \overline{1, n}$. The second one consists of $(n^2 - n)$ different partition elements only and it also equals to zero since for all $i, j = \overline{1, n}, i \neq j$ we obtain $\mu(X_i \cap X_j) = 0$, which establishes direct reflexivity.

Now let us show the validity of reflexivity in a reverse order. Let $\rho(\alpha, \beta) = 0$ for arbitrary $\alpha, \beta \in \Pi_\Omega$ such that $\alpha \neq \beta$. By virtue of terms nonnegativity in (4) we have equality of each summand to zero. Choose an element X' in the partition α . Note, in (4) it belongs to zero summands of kind $\mu(X' \Delta X_j) \mu(X' \cap X_j) = 0$ where $j = \overline{1, m}$. Suppose X' does not belong to β hence the inequality $\mu(X' \Delta Y_j) \neq 0$ is fulfilled for all $Y_j \in \beta$. Then the equality $\mu(X' \cap Y_j) = 0$ holds true for all indexes j . Though it is possible while as $X' = \emptyset$ since $X' \subset \Omega$ and family $\beta = \{Y_1, Y_2, \dots, Y_m\}$ covers the set Ω . But X' is the element of covering α of the same set Ω and naturally $X' \neq \emptyset$ then there exist elements $Y_1^*, Y_2^*, \dots, Y_p^* \in \beta$ which cover subset $X' \subset \Omega$ and have nonempty intersection with it, i.e. $\mu(X' \cap Y_r^*) \neq 0, r = \overline{1, p}$. We get a contradiction, i.e. one can assert that any element X' from α is an element of β , i.e. $\alpha \subset \beta$. Due to symmetry we have $\beta \subset \alpha$ and finally $\alpha = \beta$ which proved reflexivity.

Now we are going to examine whether (4) satisfies the triangle inequality.

Let us consider three arbitrary partitions α, β, γ of the set Ω . We have to prove $\rho(\alpha, \beta) \leq \rho(\alpha, \gamma) + \rho(\gamma, \beta)$. Suppose $\alpha = \{X_1, X_2, \dots, X_n\}, \beta = \{Y_1, Y_2, \dots, Y_m\}, \gamma = \{Z_1, Z_2, \dots, Z_l\}$. Using lemma 1 (expression (5)) we have

$$\begin{aligned} \rho(\alpha, \gamma) + \rho(\gamma, \beta) - \rho(\alpha, \beta) &= \\ &= \sum_{i=1}^n [\mu(X_i)]^2 + \sum_{j=1}^m [\mu(Y_j)]^2 + 2\sum_{k=1}^l [\mu(Z_k)]^2 - \\ &- 2\sum_{i=1}^n \sum_{k=1}^l [\mu(X_i \cap Z_k)]^2 - 2\sum_{k=1}^l \sum_{j=1}^m [\mu(Z_k \cap Y_j)]^2 - \\ &- \sum_{i=1}^n [\mu(X_i)]^2 - \sum_{j=1}^m [\mu(Y_j)]^2 + 2\sum_{i=1}^n \sum_{j=1}^m [\mu(X_i \cap Y_j)]^2 \geq 0. \end{aligned}$$

Collecting like terms and dividing by 2 we arrive at the equivalent expression

$$\sum_{k=1}^l [\mu(Z_k)]^2 + \sum_{i=1}^n \sum_{j=1}^m [\mu(X_i \cap Y_j)]^2 \geq \sum_{i=1}^n \sum_{k=1}^l [\mu(X_i \cap Z_k)]^2 + \sum_{k=1}^l \sum_{j=1}^m [\mu(Z_k \cap Y_j)]^2 .$$

Further, if we apply lemma 2 (expressions (11)–(14)) this inequality can be rewritten as follows

$$\sum_{k=1}^l (\sum_{i=1}^n \sum_{j=1}^m v_{ijk})^2 + \sum_{i=1}^n \sum_{j=1}^m (\sum_{k=1}^l v_{ijk})^2 \geq \sum_{i=1}^n \sum_{k=1}^l (\sum_{j=1}^m v_{ijk})^2 + \sum_{j=1}^m \sum_{k=1}^l (\sum_{i=1}^n v_{ijk})^2. \quad (17)$$

There is no difficulty in understanding that at raising to the second power all summands v_{ijk}^2 in both sides of this inequality are canceled. Furthermore, all doubled products in the right part are contained in the first summand at the left of (17). Indeed, with the notation $M = \{1, 2, \dots, n\} \times \{1, 2, \dots, m\}$, the first item has the form

$$\sum_{k=1}^l (\sum_{i=1}^n \sum_{j=1}^m v_{ijk})^2 = \sum_{k=1}^l \sum_{i=1}^n \sum_{j=1}^m v_{ijk}^2 + 2 \sum_{\substack{(i,j), (i',j') \in M \\ (i,j) \neq (i',j')}} v_{ijk} v_{i'j'k} . \quad (18)$$

The first summand at the right of (17) one can rewrite as follows

$$\sum_{i=1}^n \sum_{k=1}^l (\sum_{j=1}^m v_{ijk})^2 = \sum_{i=1}^n \sum_{k=1}^l \sum_{j=1}^m v_{ijk}^2 + 2 \sum_{\substack{(j,j') \in M \\ j \neq j'}} v_{ijk} v_{ij'k} . \quad (19)$$

It is clear that all pairs $v_{ijk} v_{ij'k}$ lie among doubled products $v_{ijk} v_{i'j'k}$ since in (19) all variations are connected with alteration of the second index whereas both the first and the second indices are varied in (18). The same property is fulfilled for the second summand in the right part of (17) as all variations are connected with alteration of the first index i . As a result we get that all items at the right of (17) are canceled and all summands at the left are nonnegative, i.e. inequality (17) holds. Hence the triangle inequality is valid and (4) is a metric, q.e.d.

Thus, we have proved that functional (4) is a metric. We think that obtained results have to be applied to the set coverings that will provide analysis of arbitrary weak clustering when elements can belong to more that one cluster concurrently.

Results and Conclusion

An intensive experimental exploration with the collection of histologic specimens images with final goal classification as an aid in cancer detection vindicates the efficiency of proposed metric.

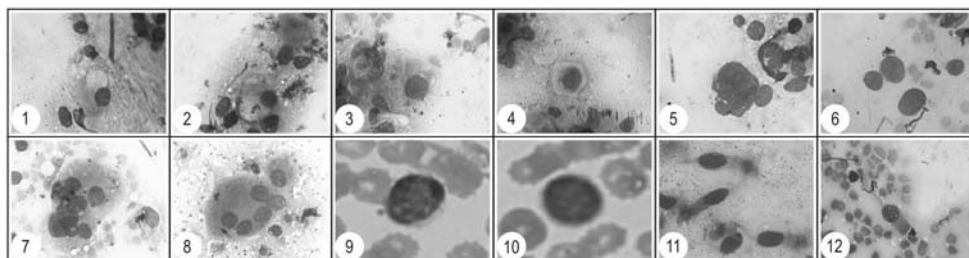


Fig. 2. Examples of input images

The analysis of experimental results has shown that the application of partitions as features provides a sufficient relevance at retrieval of the images in database with queries 'ad exemplum'. Figures 2 and 3 illustrate images

and partitions that were compared by means of traditional and proposed metrics. Examples of dependences, query image and its partition are shown in fig. 4. One can see comparability of obtained results for Euclidean metric and distance function (4). The reliability of matching can be increased by an intellectual processing (via relations analysis of region-based models) which provides conditions for entirely correct and complete segmentation.

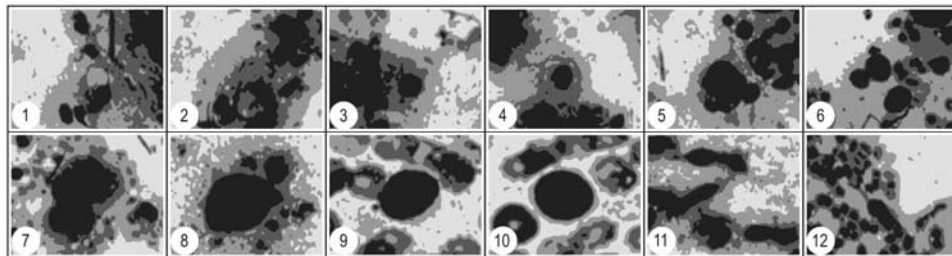


Fig. 3. Multithresholding segmentation of images shown in fig. 1

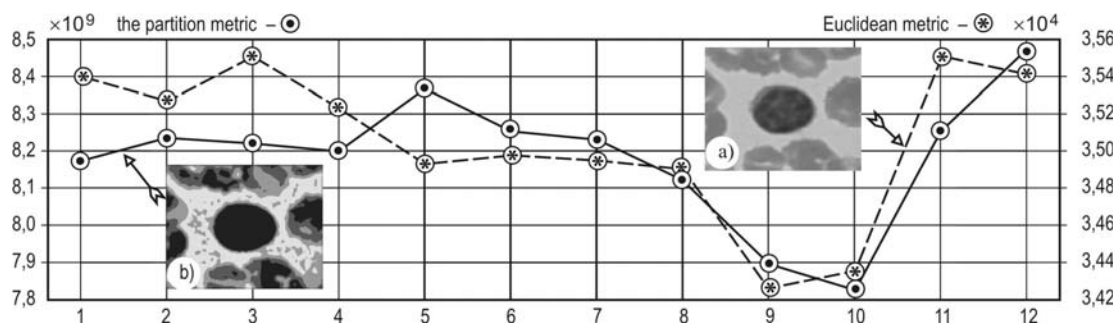


Fig. 4. Examples of image and partitions matches (query a) and b) correspondingly)

Bibliography

1. Müller H., Müller W., Squire, D.McG., Marchand-Maillet S., Pun T. Performance Evaluation in Content-Based Image Retrieval: Overview and Proposals. *Pattern Recognition Letters*, Vol. 22, 2001, pp. 593-601.
2. Li J., Wang J.Z. Automatic linguistic indexing of pictures by a statistical modeling approach. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 25, 2003, pp. 1075-1088.
3. Käster T., Wendt V., Sagerer G. Comparing Clustering Methods for Database Categorization in Image Retrieval. *Lecture Notes in Computer Science*, Springer-Verlag, Berlin Heidelberg, Vol. 2781, 2003, pp. 228-235.
4. Mashtalir V.P., Yakovlev S.V. Point-Set Methods of Clusterization of Standard Information. *Cybernetics and Systems Analysis*, Kluwer Academic Publishers, Vol. 37, 2001, pp. 295-307.
5. Kinoshenko D., Mashtalir V., Yegorova E. Clustering Method for Fast Content-Based Image Retrieval. *Computer Vision and Graphics*, K. Wojciechowski et al., eds., *Computational Imaging and Vision*, Springer, Vol. 32, 2006, pp. 946-952.
6. Kinoshenko D., Mashtalir V., Yegorova E., Vinarsky V. Hierarchical Partitions for Content Image Retrieval from Large-Scale Database. *Machine Learning and Data Mining in Pattern Recognition*, P. Perner., A. Imly, eds., *Lecture Notes in Artificial Intelligence*, Springer-Verlag, Vol. 3587, 2005, pp. 445-455.

Authors' Information

Kinoshenko Dmitry – Ph.D. student of Computer Science Department, Kharkiv National University of Radio Electronics, Lenin Ave., 14, Kharkiv, Ukraine, 61166, e-mail: kinoshenko@kture.kharkov.ua

Mashtalir Vladimir – Doctor of Technical Sciences, Professor of Computer Science Department and Dean of Computer Science Faculty, Kharkiv National University of Radio Electronics, Lenin Ave., 14, Kharkiv, Ukraine, 61166, e-mail: mashtalir@kture.kharkov.ua

Shlyakhov Vladislav – Candidate of Technical Sciences, Senior Researcher of Computer Science Department, Kharkiv National University of Radio Electronics, Lenin Ave., 14, Kharkiv, Ukraine 61166.