Pseudometric Approach to Content Based Image Retrieval and Near Duplicates Detection*

© A. Goncharov, A. Melnichenko

Laboratory of Mathematical Methods of Artificial Intelligence
{ag.tsure, alexandramelnichenko}@gmail.com

Abstract

In this paper we investigate two approaches to content based image retrieval and their application to near duplicate detection in image collections. The first approach was proposed by C.E. Jacobs et al. [10]. It involves wavelet transformation of source image to extract features. The second approach is based on so called matrix of brightness variations which uses signs of partial derivatives of image brightness as features. Both approaches use some kind of pseudometric as similarity measure.

1. Introduction

Today the main way to search images in the Web is based on textual search by image annotations (tags, keywords) and textual description of the image extracted from the page. This approach cannot be applied for retrieve images without annotation, so the only way to retrieve unannotated images is content based image retrieval. Search using textual annotations has in addition two disadvantages. Manual definition of keywords is very labor-intensive procedure. Moreover, it cannot be completely unambiguous. In the contrary Content Based Image Retrieval (CBIR) is more objective. CBIR technique can be used either as additional tool for traditional textual search (use results of textual search as queries for content based search) or as the main image retrieval engine.

*This work is supported by RFBR, project #08-07-00129, #07-07-00067

The main applications of CBIR are detection of illegal usage of images, search of logos and search in medical image collections.

CBIR task is difficult by the following reasons. First of all, the similarity of images is extremely subjective and different people usually have different opinions about similarity of the same images. Second, image similarity may be treated in different ways, e.g. images may contain similar scene or the same object on different background, etc. Therefore, the search method should take into account the color characteristics of image as a whole and its individual parts as well as the presence of common details.

Nowadays there are many methods of searching by visual similarity, including those based on an analysis of color characteristics, of the contours of objects and those combining several of these opportunities. A popular method is the using of color histograms [9] and spatial histograms [20] of images. In [2] authors describe two methods based on color histograms, including technology of quadrotrees when methods of calculating and comparing the color histograms are applied not to the whole image, but to it quarters (one-sixteenth, etc.). The simplest method to analyze boundaries and forms of objects depicted on the basis of the Sobel operator and calculate the distance from the points of the contour to the center of the figures is regarded in [2]. The form and structure analysis of objects is described in [5]. The analysis of active contours for comparison of the images containing many objects is considered in [11]. In [17] authors propose an interesting method based on the construction of set of trees from the subwindows randomly allocated from the image. Principles of the search systems architecture are reviewed in [9].

The near duplicates detection task is a special kind of CBIR task and mainly is used to avoid duplicates in the answer of image retrieval system. The near duplicates detection problem is topical in Web search, because often the same image is posted on different pages (usually after some kind of preprocessing) and it leads to multiple appearance of the same image in retrieval result.

The range of methods used for near duplicates detection is very wide. In [25] authors consider a method based on the idea of representation of the composite parts of an image and relations between them with the help of a stochastic graph. The advantage is the ability to represent the spatial relations between parts of images, the method supports learning. In the work [19] the problem of finding of the near duplicates and of the frames of videosequences is solved by using hashing methods. The first approach is to build a global hierarchy of color histograms. In this case search uses locally-sensitive hashing, a hash table is built. The algorithm

builds lists of vectors that are with high probability in the neighborhood of the given vector. The second approach uses the representation of images using local descriptors. The search technology is based on the min-hash algorithm, adapted from a text search to image search. Discriminant classifier for automatic separation of documents is used in [18]. For the analysis both low-level features of images and their textual annotations are used. Common approaches to the problem of clustering documents of different nature are given in [3]. For solving near duplicates detection problem various modifications of the nearest neighbors method are often used, such as one in the [24]. There are also fuzzy versions of the C-means method, for example fuzzy C-means is used in the [22] to search for medical images. The main disadvantage of it is inability to automatically determine the number of clusters. Methods based on the theory of fuzzy sets and relations are also proposed in [12] and [7].

The core of the near duplicates detection and content based image retrieval tasks is feature extraction technique. A feature set has to be easy to extract from the image, compact for storage and real time computations, and discriminative. In the following section we investigate one of the most popular approaches to feature extraction from images based on wavelet decomposition and another technique we used for face recognition task.

2. Feature Extraction Technique

In this paper we investigate two approaches to extracting features from image. The first approach described by C.E. Jacobs et al. [10] based on wavelet decomposition of the image and selecting position of the most significant coefficient as image features. We also investigate the usage histograms of wavelet coefficients as feature vectors. The second approach is based on representation of source image as Matrix of Brightness Variation (MBV). Originally we proposed MBV for face detection and recognition tasks [6] and achieved satisfied results.

2.1 Wavelet Based Feature Extraction

Wavelet analysis has found many applications in the field of computer technology today. Particularly, using wavelet decomposition in the computer graphics is connected primarily with compression, filtering and editing of images.

The discrete wavelet transform (DWT) is performs for image analysis. DWT produce convolution of rows and columns of the image with special filters and consider obtained result as the result of transformation.

More details about continuous and discrete wavelet transform may be found in the works [23], [1] and [8].

Using wavelet in the field of image processing is based on the notion of multiresolution analysis. When there are at the same time both large and small objects (and both low-contrast and high-contrast) in the image, it may be useful to analyze this image at various resolutions. Beginning with the original resolution, we perform wavelet transformation and the resulting image resolution each time decreases in two times. (Similar reasoning and formulas for any scaling factor can be found in [1]. Some techniques to work with images, size of which is not a power of scaling factors are given in [8]. More information about multiresolution decomposition you can find in [15].)

Thus, when we work with an image, as a result of each phase of the transformation we get four matrices: a matrix of approximation, and three matrices containing vertical, horizontal and diagonal details. Each matrix has the size equal to half of the original one. Further transformation is applied to the approximation matrix, and then again to obtained approximation matrix, etc. The transformation can be performed until approximation matrix is represented by a single number, which describes the average color of the whole image. Thus, the maximum level of decomposition for the original image with size $n = 2^N$ equals to N.

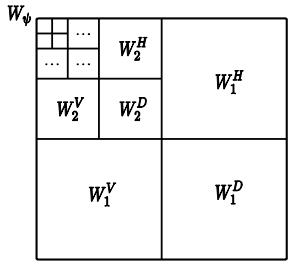

Figure 1. Schema of 2D wavelet decomposition

To implement the method we need to determine the type of transformation. Many wavelet transform bases, with different properties and applications are now well studied. One of the most popular is the Haar basis. Haar's wavelets are well located at the space and have a compact support. Haar transformation has rectangle basis, so we can expect that it will well interpret the orthogonal details and big homogeneous parts of image. There are many other wavelet families, for

instance, simlets or Dobechi's wavelets. The type of wavelets is the parameter of the method.

As a preprocessing of each image, we allocate the maximum square part, which belongs to it. Then we resize it to the same size for all images. After preprocessing we perform wavelet decomposition of each color channel of the image to the maximum level. The resulting matrix of wavelet coefficients of all levels are described as a vector. According to the algorithm proposed by Jacobs et al. [10] we build a feature vector for each image by extracting from it numbers and signs of coefficients which have the m largest absolute values. We save the positions and signs of these. This allows us to engage in seeking only information about the most significant details at all resolutions. The number of coefficients remaining after truncation is the result of a compromise between speed of computation and a sufficient level of recognition. Only information about the presence or absence of common details is important for us. As experiments shown, we can limit the number of coefficients up to about 20 for each color channel. Together with three values of average approximation for each color, they form a feature vector.

Another method related to the wavelet analysis used in this work is combining of the wavelet decomposition with analysis of histograms. For each color and level of decomposition, we compute a vector of wavelet coefficients according to the method described above. For selected number of levels of decomposition we build the histograms. Parameters of the method are number of levels of wavelet decomposition, for which the histogram are built and the number of bins in the histogram. Experimental results have shown that for stable work of the method it is sufficient to take the first 4-5 levels of decomposition and about 4 bins, because too big number of bins involves the appearance of too many zeros in the histogram. Then obtained histograms are combined together with the values of average approximation to form a feature vector.

2.2 Matrix of Brightness Variations

The Matrix of Brightness Variation (MBV) is an original image representation which operates with signs of partial derivatives of image brightness. Let us consider $I(x, y)$ as grayscale image. Define matrix of brightness variation in the following way:

$$M(x, y) = \left[\operatorname{sgn} \frac{\partial I}{\partial x} \quad \operatorname{sgn} \frac{\partial I}{\partial y} \right]_{(x,y)},$$

so each item of MBV contains pair of elements corresponding to signs of partial derivatives from the image brightness. In case of color image we can extend our definition by computation MBV for each color channel.

Proposed image representation has interesting properties. First of all, it is stable to wide class of brightness transformation. Let us consider $\varphi(v)$ as function of brightness transformation and $\tilde{I}(x,y) = \varphi(I(x,y))$ as deformed variant of source image. If function φ is monotonically increasing then both images I and $\tilde{I}$ has the same matrix of brightness variation. Linear and logarithmic transformation of brightness which take place during image acquisition in digital photo sensors are both fulfill to specified requirement.

The second advantage of proposed representation is computation simplicity. Really, we can approximate signs of partial derivatives as result of logical operation of comparison of source image with its shifted variant.

The main disadvantage of MBV is dependence from image shifting.

To build feature vector we apply several steps of image preprocessing. The first step is image cropping which allow us to work with image of the same width and height ratio. The simplest way is cropping of the greatest square area of the center of image. In this case we loose some information, but usually the most significant image parts are situated close to the center.

The second step of preprocessing is downscaling with help of bilinear interpolation. In this step we significantly reduce dimension of feature space. For example, typical size of image posted in the Web varied from hundred thousands to millions of pixels and after downscaling we get image which contain approximately one or two thousands pixels. The second good effect of downscaling is smoothing and noise reduction.

Finally, to build feature vector we compute matrices of brightness variation for each color channel of the image, reshape them to vectors and concatenate to single vector. Note, the resulting feature vectors are binary vectors and dimension of feature space is varied from 5000 to 10000.

3. Content Based Image Retrieval

The main step of the content based image retrieval task is calculating of similarity measure between query image and images from collection. Let us consider similarity measures corresponding to the considered in previous section feature extraction techniques.

Popular metrics, such as Euclidean, cannot be applied to the CBIR problem, because they do not take into account human perception of image similarity. For this purpose, the pseudometrics are more flexible instrument.

According to the approach proposed by Jacobs et al. [10] feature vector is obtained from wavelet decomposition and measure of similarity between feature vector Q of query image and feature vector T of target image may be defined in the following way:

$$mes(Q,T) = \sum_{c=R,G,B} \left(w_{c,0} \left| Q_c[0] - T_c[0] \right| - \sum w_{c,lev} \left(Q_c[i] = T_c[i] \right) \right),$$

where w is a vector of weights, c is a color channel and lev is a level of the wavelet decomposition. In this form of pseudometric the smallest values are corresponded to the closest images. Weights can be estimated according to statistical methods (for example, classical regression and Bayesian logistic regression models outlined in [14]).

Obtained at the wavelet-histogram method feature vectors are compared using Mahalonobis metric:

$$d(x,y) = \sqrt{(x-y)^T S^{-1} (x-y)},$$

where x and y are two feature vectors, S is the covariance matrix.

Mahalonobis metric is based on correlations between vectors by which different patterns can be identified and analyzed. For using Mahalonobis metric we need of test set of vectors belonging to one of N classes. According to the given sample set of images we estimate the covariance matrix $S = (s_{ij})$ as following:

$$S = \overline{F}\,\overline{F}^T, \overline{F} = F - \overline{f},$$

where $\overline{f}$ is the average feature vector across the sample.

In case of feature vector obtained from the matrix of brightness variation we use Hamming's metric as similarity measure:

$$\rho(x,y) = \sum_i [x_i \neq y_i].$$

The value of Hamming's metric corresponds to the number of inequal elements in feature vectors. Note that Hamming's metric on feature vectors corresponds to the pseudometrics on source images, because as mentioned above different images may be represented by the same feature vector and from the equalities of feature vectors does not follow the equalities of images.

For extracting similar images from the database, for each query we find at the similarity matrix values with smallest values of pseudometrics and take corresponding images as the result. Number of similar images can either be defined or can be depended on a threshold, set on the similarity measure.

4. Near Duplicates Detection

The near duplicate detection task consists in extraction very similar pictures from image collection. Examples of near duplicates may be two sequential images in photo series or neighbor frames in video or slightly changed variants of original image. Note that clusters of image duplicates may have nonempty intersection. It is obvious from the example of video sequence: every two or three neighbor frames usually are near duplicates, but if we handle long sequence of frames, all of them are not near duplicates, but every short subsequence is.

Figure 2. Examples of near duplicates (images are taken from ROMIP collection [21])

One of the main problems of the near duplicates detection task is computation complexity and high requirements to memory storage. The main step of the task is selecting from image collection subsets of images (clusters) with high similarity measure between items of the cluster. To carry out this step we need firstly compute measure of similarity for every pair of images in collection that lead to squared complexity and secondly it is necessary to store similarity matrix that lead to allocation of huge amount of memory. For example, to store similarity matrix for collection of million images we need to allocate memory for about 3.7 TB (if a similarity measure is symmetrical) and it quadratic depends on the collection size.

To outperform storage problem and reduce memory requirements we use sparse matrix representation and store only high values of similarity measure. Because in real collections there are duplicates for a small part of images and number of duplicates per image is quite small, amount of

memory needed to store sparse similarity matrix is grow up linearly with number of images in collection.

After calculation of similarity matrix it is necessary to extract clusters of duplicate images. It can be easily done from the similarity matrix in the following way. Firstly we should look throw the similarity matrix and build new cluster from images which have high similarity measure to the current image. Secondly, we unite the clusters which have quite strong intersection and do not unite clusters with small intersection, because it may lead to the chain process of joining of the big number of clusters. To avoid this situation we compute centers of clusters and compare each added vector with center of current cluster. If value of pseudometrics between them is more than given threshold, we do not add this vector to the cluster.

In our experiments we tried three described above feature sets: coordinates of the most significant wavelet coefficients, histograms of wavelet coefficients and matrix of brightness variations. At preprocessing stage we apply cropping of the biggest central square part of the image and scaling to the same size, e.g. 32×32 or 64×64 pixels.

5. Experimental Results

5.1 Content Based Image Retrieval

The image collection used for judgment of CBIR a task is subset of Flickr.com image database and contains 20,000 of color and grayscale images of various sizes. All images go without any annotations or keywords.

The quality assessment was carried out by 250 randomly chose queries with pool depth 19. Every pool entry was judged by two assessors as strong relevant, weak relevant or not relevant to the query. In case of "or" metric resulting image is treated relevant if at least one assessor marks it as relevant. In case of "and" metric resulting image is treated relevant if both assessors mark it as relevant

The following tables contain results of the quality assessment for six different CBIR systems, where "jade-2" corresponds to the MBV-based approach, "jade-6" corresponds to the wavelet-based approach which used histograms of wavelet coefficients as feature vectors, "jade-7" corresponds to the wavelet-based approach proposed by Jacobs et al. [10], and "xxxs-1", "xxxx-2", "xxxx-3" correspond to other CBIR systems tested by ROMIP.

Table 1. CBIR results for weak relevance (“or” metric)

Run ID / Metric	xxxx-1	xxxx-2	xxxx-3	jade-7	jade-2	jade-6
Precision(10)	0.2078	0.1332	0.1410	0.0340	0.0447	0.0086
Bpref-10	0.2376	0.1523	0.1611	0.0367	0.0435	0.0056
Bpref	0.1482	0.0894	0.0964	0.0231	0.0286	0.0045
Recall	0.3906	0.2429	0.2520	0.0595	0.0682	0.0080
Average. precision	0.1457	0.0824	0.0884	0.0165	0.0226	0.0033
Precision	0.1959	0.1189	0.1245	0.0410	0.0476	0.0133
R-precision	0.1996	0.1268	0.1344	0.0381	0.0408	0.0064
Precision(5)	0.2123	0.1549	0.1639	0.0352	0.0525	0.0131

Table 2. CBIR results for weak relevance (“and” metric)

Run ID / Metric	xxxx-1	xxxx-2	xxxx-3	jade-7	jade-2	jade-6
Precision(10)	0.1005	0.0557	0.0585	0.0120	0.0169	0.0033
Bpref-10	0.1954	0.1169	0.1216	0.0178	0.0304	0.0049
Bpref	0.0840	0.0359	0.0500	0.0073	0.0114	0.0041
Recall	0.4282	0.2366	0.2368	0.0482	0.0661	0.0061
Average precision	0.1218	0.0599	0.0708	0.0083	0.0155	0.0035
Precision	0.0897	0.0460	0.0477	0.0148	0.0179	0.0072
R-precision	0.1085	0.0487	0.0663	0.0126	0.0180	0.0050
Precision(5)	0.1093	0.0689	0.0721	0.0098	0.0208	0.0055

Table 3. CBIR results for strong relevance (“or” metric)

Run ID / Metric	xxxx-1	xxxx-2	xxxx-3	jade-7	jade-2	jade-6
Precision(10)	0.0748	0.0331	0.0417	0.0110	0.0157	0.0039
Bpref-10	0.1746	0.0948	0.1064	0.0329	0.0462	0.0087
Bpref	0.0589	0.0281	0.0379	0.0045	0.0062	0.0080
Recall	0.4295	0.1945	0.2144	0.0789	0.0928	0.0093
Average precision	0.1055	0.0535	0.0620	0.0156	0.0224	0.0077
Precision	0.0688	0.0290	0.0319	0.0137	0.0142	0.0099
R-precision	0.0751	0.0365	0.0452	0.0080	0.0106	0.0093
Precision(5)	0.0803	0.0441	0.0441	0.0094	0.0157	0.0063

Table 4. CBIR results for strong relevance (“and” metric)

Run ID / Metric	xxxx-1	xxxx-2	xxxx-3	jade-7	jade-2	jade-6
Precision(10)	0.0395	0.0105	0.0211	0.0026	0.0079	0.0000
Bpref-10	0.1467	0.0665	0.1111	0.0085	0.0382	0.0000
Bpref	0.0351	0.0526	0.0658	0.0000	0.0000	0.0000
Recall	0.4167	0.1382	0.2434	0.0702	0.1053	0.0000
Average precision	0.0851	0.0606	0.0945	0.0061	0.0217	0.0000
Precision	0.0360	0.0111	0.0166	0.0055	0.0073	0.0000
R-precision	0.0439	0.0526	0.0658	0.0000	0.0000	0.0000
Precision(5)	0.0368	0.0158	0.0316	0.0000	0.0053	0.0000

5.2 Near duplicates detection task

Near duplicates detection task was judged on the image collection contained frames of several video sequences, so there are a lot of natural duplicates in the collection. Total amount of images in the collection is 37,800.

The following procedure is used for near duplicates detection quality assessment. A random image is chose and all images are marked by systems as near duplicate for the choused one are placed to the pool. Neighbor frames of the chose image are placed to the pool too. Assessor judges the pool entries and classifies them by the 20 or less clusters. Assessor may miss some images from the pool during judgment. In our case 45 randomly selected images are used to form pools.

Table 5. Results of quality assessment of the near duplicates detection task

Run ID	FPR	FNR	Precision	Recall
xxxxx-1	4.91354E-06	0.591639	0.975427	0.408361
xxxxx-2	4.07497E-04	0.585925	0.931302	0.414075
xxxxx-3	2.55687E-06	0.661169	0.986575	0.338831
xxxxx-4	3.15882E-05	0.464711	0.910422	0.535289
xxxxx-5	2.45281E-04	0.367309	0.698379	0.632691
xxxxx-6	6.50414E-04	0.629230	0.770902	0.370770

As result 526 clusters with total number of images 3,765 were built by assessor. The following table contains results of quality assessment provided by ROMIP for several systems, where run ID "xxxxx-1" and "xxxxx-3" corresponds to the MBV-based approach, "xxxxx-2" corresponds to the wavelet-based approach with features proposed by Jacobs et al. [10]. Quality is assessed in terms of the following metrics: False Positives Rate (FPR, Type I Error) vs. False Negatives Rate (FNR, Type II Error) and Precision vs. Recall.

ROMIP Organization Committee also provides values of the metrics mentioned above for every cluster. Based on these values we estimate recall and precision graph by calculating average precision over all clusters for specified recall level.

To estimate dependence between false positives rate and false negatives rate we calculate average FPR over all clusters for specified FNR.

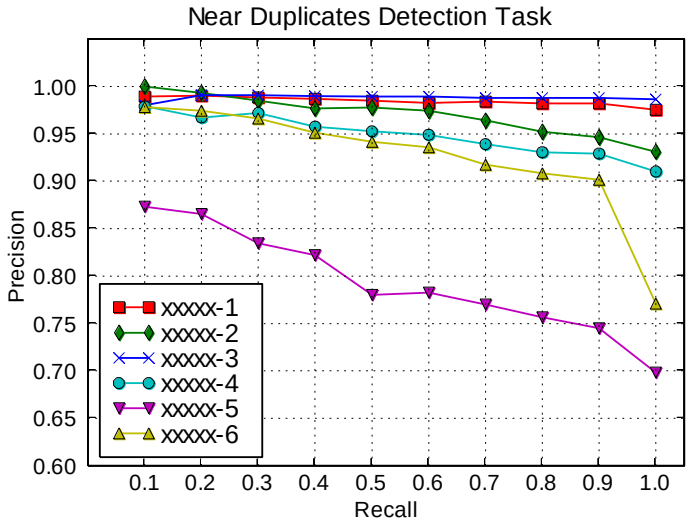

Figure 3. Recall-Precision graph of near duplicates detection task

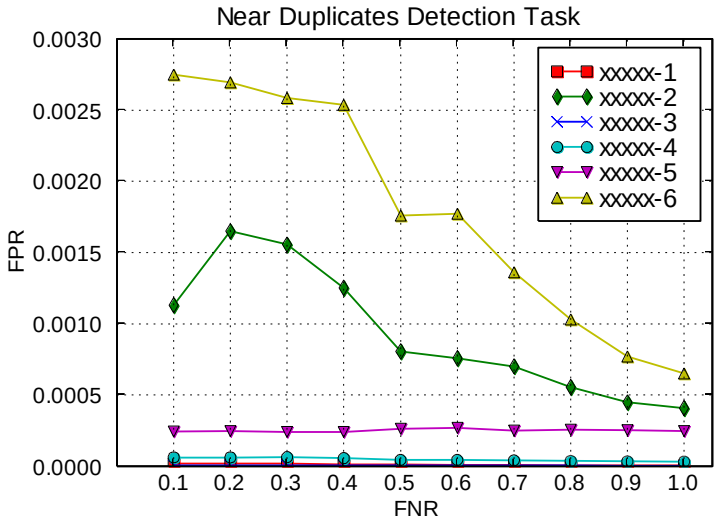

Figure 4. False Negatives Rate vs. False Positives Rate of the near duplicates detection task

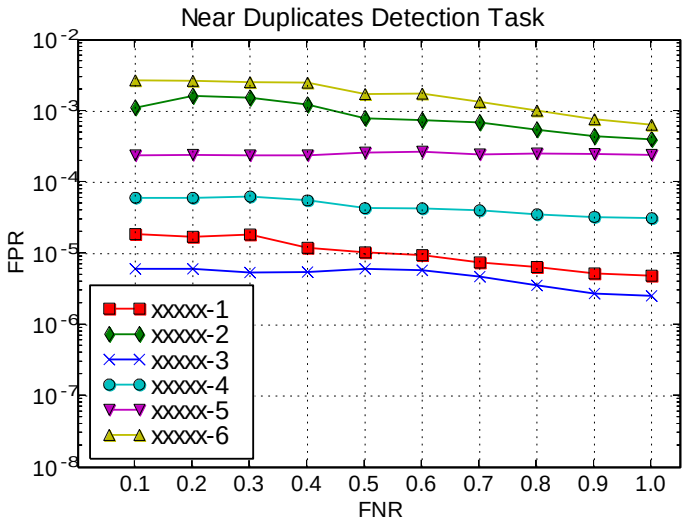

Figure 5. False Negatives Rate vs. False Positives Rate (in logarithmic scale) of the near duplicates detection task

6. Conclusion

As experimental results show considered approaches are not so good for content based image retrieval but quite good for near duplicates detection task. The main reason of such results consists in feature extraction techniques that are not invariant to affine transformations for all considered approaches.

References

[1] Astafieva N. M. Wavelet analysis: the foundations of theory and examples of applications // Successes of the physical sciences – 1996 – Vol. 166, No. 11 (in Russian).
[2] Baigarova N. S., Bukhshtab Yu.A., Evteeva N.N., Koryagin D.A. Various Questions Connected with Content-Based Search of Visual Information and Videoinformation // Inst. Appl. Math., the Russian Academy of Science, 2002 (in Russian).
[3] Berikov V.and Lbov G., Modern approaches in the cluster analysis, 2008 (in Russian).
[4] Bloom B.. Space/time Trade-Offs in Hash Coding with Allowable Errors. In Communications of ACM, volume 13(7), pages 422-426, 1970.
[5] Datta R., Joshi D., Li J., Wang J.Z. Image Retrieval: Ideas, Influences, and Trends of the New Age // ACM Transactions on Computing Surveys, Vol. 40, No. 2, April 2008.
[6] Goncharov A., Gorban A., Karkishchenko A., Lepskiy A. Content Based Facial Image Search // http://download.yandex.ru/IMAT2007/goncharov.pdf, 2007. (In Russian)
[7] Goncharov M., Clustering on the base of fuzzy relations. Fuzzy relation clustering algorithm, 2005 (in Russian).
[8] Gonzalez R. C., Woods R. E. Digital Image Processing, 2002, Pearson Education, Inc.
[9] Hove Lars-Jacob. Evaluating Use of Interfaces for Visual Query Specification // Department of Information Science and Media Studies University of Bergen.
[10] Jacobs C. E., Finkelstein A., Salesin D. H. Fast Multiresolution Image Querying, ACM SIGGRAPH, New York, 1995.
[11] Katare A., Mitra, Suman K.; Banerjee A. Content Based Image Retrieval System for Multi Object Images Using Combined Features // Theory and Applications – 2007 - Pages: 595 – 599.
[12] Kaya M., An algorithm for image clustering and compression, Turkey J Electronic Engine, vol. 13, no. 1, pp. 79-91, 2005.

[13] Kotoulas L., Andreadis, Circuits I. Colour histogram content-based image retrieval and hardware implementation // Devices and Systems - 2003-Vol. 150, No. 5, Pages: 387-93.
[14] Ksantini R., Ziou D., Colin B., and Dubeau F. Weighted Pseudometric Discriminatory Power Improvement Using a Bayesian Logistic Regression Model Based on a Variational Method // IEEE Transactions on pattern analysis and machine - 2008 - VOL.30, NO. 2.
[15] Mallat St. G. A Theory for Multiresolution Signal Decomposition. The Wavelet Representation // IEEE Transactions on pattern analysis and machine intellegence - 1989 - VOL.11, NO. 7.
[16] Marchionini G. and Hersch W. Proceedings of the Second ACM/IEEE-CS Joint Conference on Digital Libraries (eds.). New York, Association for Computing Machinery, 2002.
[17] Maree R., Geurts P., Wehenkel L. Content-based Image Retrieval by Indexing Random Subwindows with Randomized Trees // University of Liege, Belgium, Proc. ACCV 2007, LNCS.
[18] Nickolov R. and Collins-Thompson K., A clustering-based algorithm for automatic document separation, 2001.
[19] Philbin J., Isard M., Zisserman A., and Chum O., Scalable near identical image and shot detection, CIVR, 2007.
[20] Rao A., Srihari R.K., Zhang Z. Spatial color histograms for content-based image retrieval // Tools with Artificial Intelligence -1999 - Vol. 30, Pages: 183 – 186.
[21] ROMIP Web site, 2005. http://romip.narod.ru
[22] Shihab A. I., Fuzzy Clustering Algorithms and their application to Medical Image Analysis}, PhD thesis, 2000.
[23] Stollnitz E. J., DeRose T.D., Salesin D.H. Wavelets for Computer Graphics. Theory and application // Morgan Kaufmann Publishers, San Francisco.
[24] Yianilos P. N., Data structures and algorithms for nearest neighbor search in general metric spaces, NEC Research Institute, Princeton, 2000.
[25] Zhang D.-Q. and Chang S.-F., Detecting image near-duplicate by stochastic attribute relational graph matching with learning, Department of Electrical Engineering, Columbia University, New York, 2003.