Improving CBIR Systems Using Automated Ranking

Goran J. Zajić, Nenad S. Kojić, and Branimir D. Reljin, *Senior Member, IEEE*

Abstract — **The most common way of searching images on the Internet and in private collections is based on a similarity measuring of a series of text words that are assigned to each image with users query series. This method imposes strong constraints (the number of words to describe the image, the time necessary to thoroughly describe the subjective experience of images, the level of details in the picture, language barrier, etc.), and is therefore very inefficient. Modern researches in this area are focused on the content-based searching images (CBIR). In this way, all described disadvantages are overcome and the quality of searching results is improved. This paper presents a solution for CBIR systems where the search procedure is enhanced using sophisticated extraction and ranking of extracted images. The searching procedure is based on extraction and preprocessing of a large number of low level image features. Thus, when the user defines a query image, the proposed algorithm based on artificial intelligence, shows to the user a group of images which are most similar to a query image by content. The proposed algorithm is iterative, so the user can direct the searching procedure to an expected outcome and get a set of images that are more similar to the query one.**

Keywords — **CBIR, feature, image databases, neural networks, ranking, searching.**

I. Introduction

INTENSIVE development and the use of multimedia and networking technology have recently provided a rapid growth in production and distribution of multimedia files. Standard multimedia databases are not capable anymore to serve users demands for database storage capacity, accessibility and searching functionalities. This problem can be overcome with the use of large multimedia databases with advanced functionalities hosted on personal computers or servers all over the Internet.

Because of their simple acquisition and distribution using the Internet, the images are the most numerous multimedia files. Today, every user of computers has his own image database hosted on a personal computer or on some of online image databases. In a couple of years, the average user can produce an impressive number of images using a digital camera. Searching images in large image databases can be very difficult and finding an efficient searching procedure represents a challenge for researchers.

There are two different approaches for image labeling in large image databases. The first approach uses semantic labeling based on image file name and a group of words that describe the content of the image. This approach is very time consuming because every image has to be examined and described. Searching process is based on matching the labeling words. Efficiency of this procedure depends on the language used for image description and number of used words. The results of the searching task are not usually satisfying for users. The second approach for image labeling is image annotation by content with low level features (color, texture, shape,…), and the systems which use annotated images in this way for searching procedure are named Content Based Image Retrieval Systems (CBIR systems) [1]-[3]. These systems provide automated annotation of the images and very simple measuring of similarity based on metric distance between feature vectors.

Searching results of CBIR systems based on an objective measure of similarity (metric distance) are not usually satisfying for users, because users have a subjective measure of similarity based on a perceptive impression of the image. This disadvantage is overcome by using the user's assistance in searching procedure with intelligent logic for the correction of searching results. This assistance is named relevance feedback [4]-[12]. The searching procedure with relevance feedback is very simple. After initial search based on the objective measure of similarity, the resulting images are shown to the user to assign them as relevant or irrelevant. Relevant images represent images that are most similar to the query image among the resulting ones. When the images are assigned, intelligent logic performs a new searching procedure as the correction based on the user's perception of similarity.

Procedure for the automated selection of relevant images can be based on intelligent logic [9], [12]. A CBIR system with the automated selection of relevant images based on automated ranking is presented in this paper. The system can be described as a three-phase system. The first phase represents feature extraction from images and feature database creation. The second phase consists of initial search and automated selection of relevant images. The third phase represents user relevance feedback and is iterative.

Initial search based on a query image and the selection of a representative group of images are realized by using three types low level features: color, texture and shape. Similarity measurements were performed using five color features, four texture features and one characteristic shape feature.

Goran J. Zajić is with the ICT College Belgrade, Serbia, (e-mail: goran.zajic@ict.edu.rs).

Nenad S. Kojić is with the ICT College Belgrade, Serbia, (e-mail: nenad.kojic@ict.edu.rs).

Branimir D. Reljin is with Innovation Center, School of Electrical Engineering, University of Belgrade, Serbia, (e-mail: reljinb@etf.rs).

In this paper the main focus is on the second phase, the pre-processing phase of CBIR system, because the other two phases are already presented in the previous works of the authors of this paper.

This paper is organized in five chapters. After the introduction, the principles of neural network and a description of the proposed system are given in chapter two. The used features are explained in the third chapter in detail. The results are presented in chapter four. The fifth chapter presents conclusions and further guidelines in the work.

II. SYSTEM DESCRIPTION

A CBIR system with relevance feedback and selection of relevant images by ranking is realized in three phases. A functional scheme of the system is shown in Fig. 1. Each block on the scheme represents a corresponding step in the searching process and these blocks are grouped into phases according to their functional description.

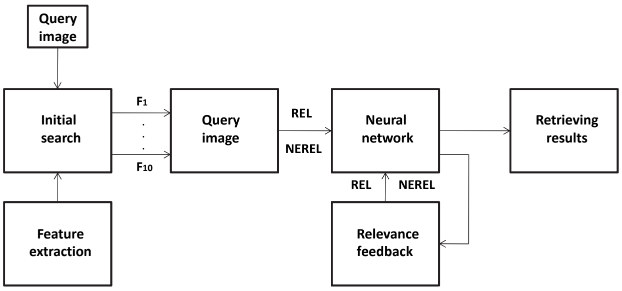

Fig. 1. Functional scheme of the proposed CBIR system based on automated selection of relevant images.

A. Phase One

The first phase represents the procedure of preparation of an image feature database. Low level features for color, texture and shape are extracted from images, and stored in feature vectors. An image feature database consists of several feature matrices created from individual feature vectors. All feature matrices are normalized.

B. Phase Two

In the second phase, the user selects a query image and starts the image searching procedure. This phase is a set of steps in a searching process that results in the selection of a number of relevant images based on the query one. Since after the initial search selected relevant images are not displayed to the user, the second phase is named preprocessing.

The initial search is performed for each feature group (color, texture, shape), and individually by every feature. Based on the Euclidean distance from query image features the closest N1 = 20 images are selected from each individual feature matrix. The images in the database are identified only by their ordinal number (ID). A series of IDs is formed for each image feature (sequences F1, F2,..., F10). New arrays of image IDs A, B and C are obtained by concatenation of arrays F1-F10 based on feature group membership. Corresponding lengths of arrays A, B and C are 100, 80 and 20 (5x20, 4x20, 1x20) coordinates respectively. A visual presentation of this procedure is given in Fig. 2.

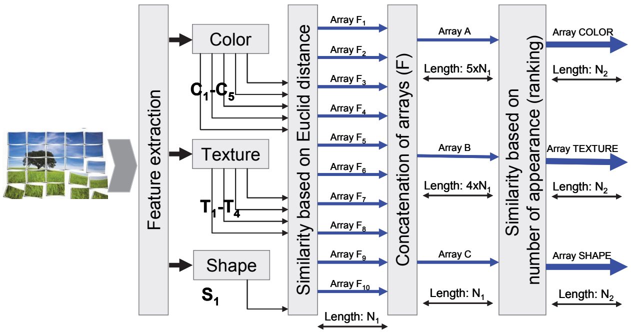

Fig. 2. Initial searching procedure based on individual features and Ranking algorithm.

In the next step the arrays of image IDs (A, B and C) are sorted based on the number of appearances (ranking) of image IDs in each array. The number of repeating image IDs in the appropriate group of features indicates the similarity to the query image by various criteria. Because of a number of the coordinates and their variance there is a problem with the dominance of certain characteristics when the Euclidean distance between feature vectors is used as the only criterion for similarity in the initial search. This problem is overcome with the indication of similarity by various criteria in the proposed algorithm. After sorting, N2 = 15 image files with the highest number of appearances are selected from each group of features (A,B and C), and moved with a preserved rank order to a new set of arrays COLOR,TEXTURE and SHAPE respectively, as it is shown in Fig. 2.

Merging of arrays is realized by using a special procedure. The resulting array D is obtained by concatenation of the first 2xN3 image IDs from array COLOR, the first N3 image IDs from array TEXTURE and the first N3 image IDs from array SHAPE, shown in Fig. 3.

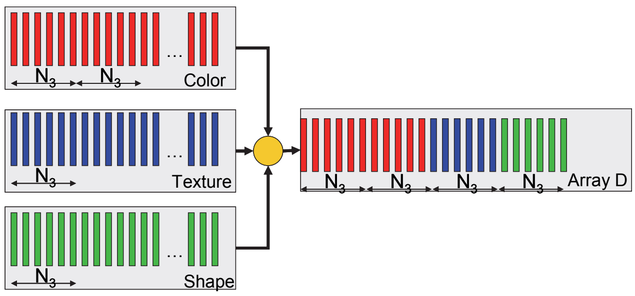

Fig. 3. Merging resulting arrays of image IDs after ranking in group of features.

Selection of relevant images is achieved by the use of the query image feature vector. All images IDs in series D are ranked based on the Euclidean distance between a feature vector of images and a feature vector of the query one. The closest N4 = 5 images IDs with the highest ranks from series D make a new set of relevant images REL, and the remaining images are members of a new set of irrelevant images NEREL. The number of images in NEREL series is volatile because of the union of sets COLOR, TEXTURE and SHAPE. The union of these sets has a volatile number of images because of an appearance of identical images in some sets. The efficiency of CBIR system searching procedure depends on the numbers of selected images N3 and N4 in the process of integration of the results in the initial search by individual image features.

C. Phase Three

After the initial selection of relevant images in the third phase of CBIR system searching procedure intelligent logic is implemented for correction of searching results. Intelligent logic is based on RBF-type neural networks [10]. In this phase the creation of a new feature vector is achieved by a neural network, based on the old one with correction displacement. The direction and orientation of displacement is determined by feature vectors of relevant and irrelevant images. The algorithm of the neural network of this type is based on changing the query feature vector and the width of Gaussian curve. After the first application of intelligent logic, N5 = 30 images are shown to the user. In this step of searching procedure the user's task is to select a set of relevant images for the next iteration based on his subjective measuring of similarity. Thus, the user can change the flow and quality of the searching procedure. A sequential series of modifications of the query feature vector is obtained by repeating this process. A new position of the query feature vector is closer to the cluster of relevant images, which is regulated by different weighting coefficients for the relevant and irrelevant images in the process of changing the query feature vector. The width of Gaussian curve is determined by the standard deviation of clusters of similar images. Modifying this parameter the sensitivity of algorithm is changed in order to gain influence of similar images in the clusters with a small standard deviation. An objective measure of similarity used in this algorithm is the Euclidean distance.

III. Image Description

In this paper the content of the image is described by low-level features for color, texture and shape. Matrices of individual features are formed based on a series of feature values extracted from the images by corresponding algorithms. The CBIR system that is presented in this work uses the individual feature matrices for initial search. In further steps of searching procedure the feature vector matrix is used, which is obtained by concatenation of individual feature matrices. The application of the intelligent logic in phase three and the final step of ranking for the selection of relevant images in phase two are based on processing the image feature vectors.

Features are extracted from each image and stored in ten feature matrices with different dimensions 1000xKj, j = 1,..., 10 where K_j is the length of individual feature, and j is the number of extracted features.

In the searching procedure of CBIR system presented in this work the following low-level features are used, HSV (Hue Saturation Value) color histogram, color moment descriptor, color layout descriptor, scalable color descriptor and color correlogram for color, the homogeneous texture descriptor, radial co-occurrence matrix descriptor, edge histogram descriptor and wavelet texture grid descriptor. The global edge histogram descriptor is used as a characteristic feature of the shape. A brief description of features is given below, and a method for feature calculation is explained in the literature in detail [13].

A. Color Histogram Descriptor

A color histogram descriptor is extracted from images in HSV color space with adequate quantization per channels (H:S:V=18:3:3). The dimension of descriptor is 162 coordinates.

B. Color Moments Descriptor

Color moments have been proved to be efficient and effective in representing color distributions of images. The first (*mean*), second (*variance*) and the third (*skewness*) color moments are used. Images are divided into 3x3 blocks, and color moments are extracted from each image block for each RGB color channel. A color moments feature consists of 81 components.

C. Color Correlogram Descriptor

A color correlogram characterizes the color distributions of pixels and the spatial correlation of pairs of colors. In this paper, the three-dimensional histogram of correlogram consists of colors ($c1$, $c2$) of any pixel pair, while the third coordinate is their spatial distance (d). A simplified version of the feature ($c1$,$c1$,d), called the *color autocorrelogram*, is used in the presented work. The color autocorrelogram only captures the spatial correlation between identical colors and thus reduces the dimension of feature to 648 coordinates.

D. Scalable Color Descriptor

A scalable color descriptor is a color histogram extracted in *HSV* color space, and encoded for storage efficiency. After quantization per color channels (H:S:V=16:4:4) and scaling by 1D Haar transformation a set of 32 coordinates is derived which consists of 16 low-pass coefficients and 16 highpass coefficients.

E. Color Layout Descriptor

A color layout descriptor compactly characterizes the spatial distribution of colors within image blocks sized 8x8 pixels. The color layout uses an array of representative colors for the image blocks, in the *YCbCr* color space as the starting point for the descriptor definition. The size of this color descriptor is 12 coordinates.

F. Radial Co-occurrence Texture Descriptor

A radial co-occurrence texture descriptor is a global texture descriptor. Four types of features *Entropy*, *Energy*, *Contrast* and *inverse difference moment* for 24 different directions are extracted from the 256 level gray images, The dimension of this descriptor is 96 coordinates.

G. Wavelet Texture Grid Descriptor

A wavelet texture grid descriptor is realized by dividing an image into 4x4=16 regions and application of the fourth-level Haar wavelet transform. The variances of the high frequency sub-bands (12 sub-bands) of each region are obtained. The dimension of this descriptor is 192 coordinates.

H. Edge Histogram Texture Descriptor

An edge histogram descriptor is obtained by dividing the image into 4x4=16 regions. An edge histogram is created for a region by applying five filters for five main directions as it is proposed in the MPEG-7 standard. The dimension of the descriptor is 80 coordinates.

I. Homogenous Texture Descriptor

Homogenous Texture Descriptors are describing *directionality*, *coarseness*, and *regularity* of patterns in the images. This descriptor is created by applying tuned Gabor filters with 6 orientations and 5 scales tuned filters. The dimension of this descriptor is 62 coordinates.

J. Shape Descriptor

The feature that describes a shape [12], [14] is obtained from the feature of global edges (edge histogram) with the application of the Fourier transform. A total length of this descriptor is 30 coordinates.

IV. Results

The performances of the presented CBIR systems have been tested on the Corel 1K [15] image database, which contains 1000 images classified and sorted into 10 classes according to the topic. The topics are Africa (class 1), Beach (class 2), Buildings (class 3), Buses (class 4), Dinosaurs (class 5), Flowers (class 6), Elephants (class 7), Horses (class 8), Mountains (class 9) and Food (class 10).

The system is observed by measuring two types of efficiency of the searching process. One is measured after the initial search and the other after the first iteration of the neural network. The first efficiency is calculated as the ratio of the number of images that belong to the class of the query one and the number of images in REL series. The second one is calculated as the ratio of the number of images that belong to the class of query image and the number of images that are presented to the user (N5 = 30 frames). In the calculation of both efficiencies the query image is treated as an image found in the search. The query images are randomly generated for the experiment, so that each sequence consists of one image from each class. Three random sequences are generated and each one contains ten images.

The efficiency of CBIR systems depends on the numbers of selected images (N3, N4) in the initial search process, and it is shown in Fig. 4 – Fig. 9.

The query images in all three random sequences are grouped into classes for easier comparison of efficiency. As can be seen in Fig. 4, the efficiency of searching process is high when the numbers of selected images (N3 and N4) in the initial search are small. By increasing the number of images that are selected in a preprocessing phase the number of images that are not relevant but similar to the query one in some features are higher. The results of efficiency after the first iteration of the neural network are shown in Fig. 5. The values are very high, except in the case of class 3 and class 8. It is obvious that the selection of a small number of relevant images can be improved with the user assistance and the application of intelligent logic. In this way the user performs the correction of searching process according to his requests. Even in the case of class 3 and class 9. In these cases the number of relevant images is increased and satisfactory, so in next iteration of relevance feedback the user can improve the searching results.

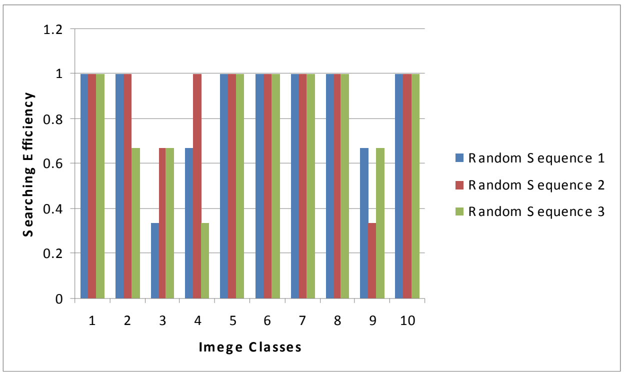

Fig. 4. Efficiency of searching procedure after initial search for three random image sequences, N_3=3 and N_4=3.

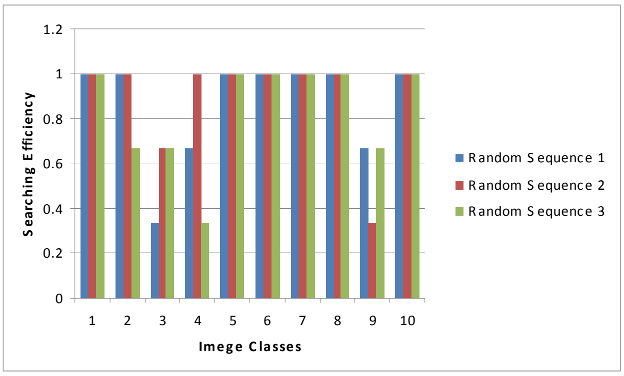

Fig. 5. Efficiency of searching procedure after the first iteration of neural network for three random image sequences, N_3=3, N_4=3 and N_5=30.

The efficiency of the searching process with the increase in the number of selected images in preprocessing is shown in Fig. 6 – Fig. 9. The trend is such that the value of efficiency decreased with the increase of the number of selected images from the group of features. Reason for this is the increase of the number of images that do not belong to the class of query one but in some of the features are very similar to it. After the first iteration of the neural network, results of efficiency do not change drastically with the increase of the number of relevant images that are forwarded to the intelligent logic. This happens because apart from relevant images, the irrelevant ones affect the correction of searching results. Some images that do not belong to the class of query image are positioned in the area of very similar images to the query one. These images do not affect the convergence of system into an area with a larger number of relevant ones.

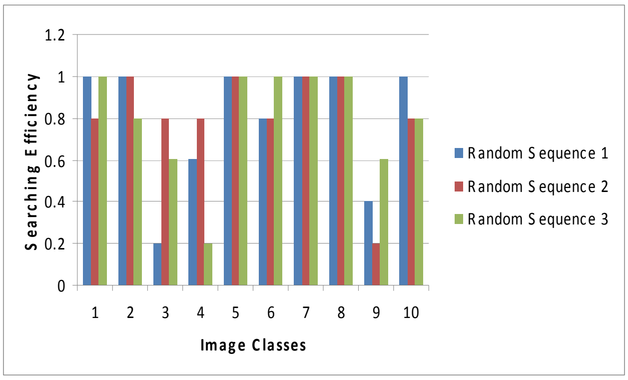

Fig. 6. Efficiency of searching procedure after initial search for three random image sequences, N_3=3 and N_4=5.

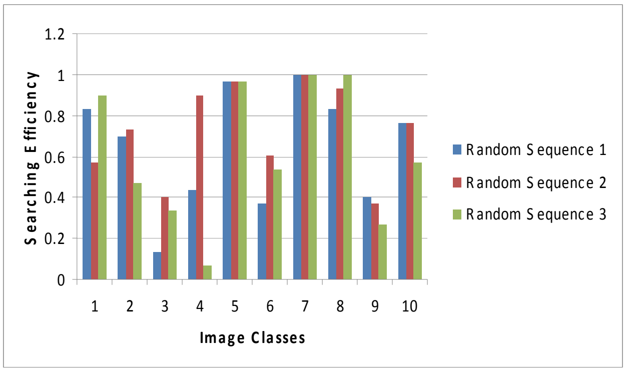

Fig. 7. Efficiency of searching procedure after the first iteration of neural network for three random image sequences, N_3=3, N_4=5 and N_5=30.

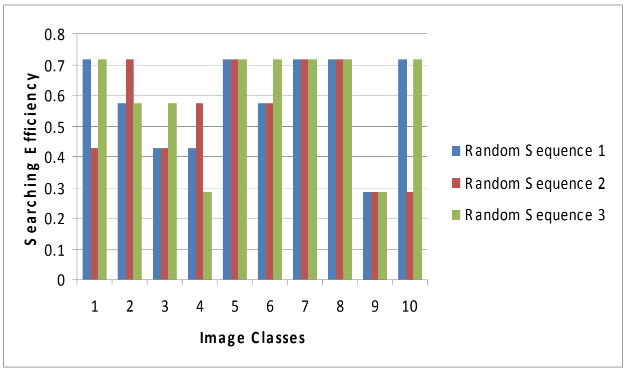

Fig. 8. Efficiency of searching procedure after initial search for three random image sequences, N_3=5 and N_4=7.

The change in efficiency in all three observed cases of the selection of images for the worst cases of class 3 and class 9 is shown in Fig. 10 - Fig. 11. The query images from these classes have generated the smallest number of relevant images. It can be seen that the increase of N4 does not in general affect the increase of the number of relevant images. Optimal values for N3 and N4 are 3 and 5 respectively, for which the largest number of relevant images are achieved after the ranking. The ratio of the actual number of relevant images in all three observed cases after the initial search and after the application of intelligent logic is shown in Fig. 12 – Fig. 14. The results show that intelligent logic can increase the quality of searching process even in cases where only one picture is declared as relevant. As said above, with the increase of N4, searching efficiency decreases but the number of the relevant images, which will be presented to the user, rises. This is very important because the higher number of images presented to the user, which satisfies the users requirements, more precisely directs the further searching process.

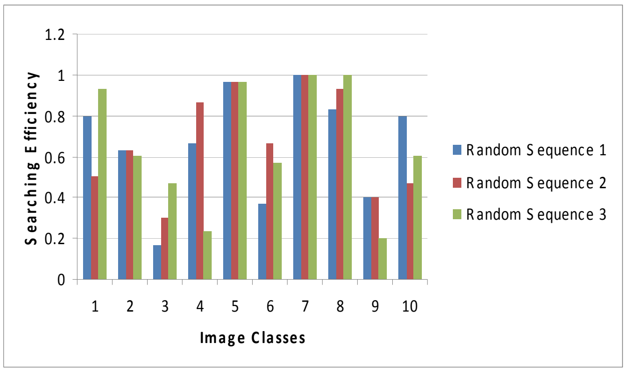

Fig. 9. Efficiency of searching procedure after the first iteration of neural network for three random image sequences, N_3=5, N_4=7 and N_5=30.

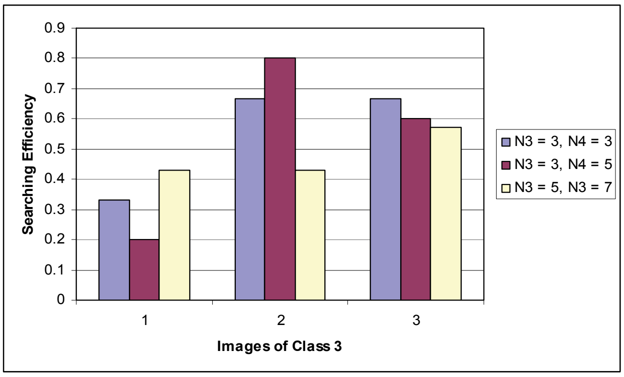

Fig. 10. Efficiency of searching procedure for images from class 3 for all three cases of initial search.

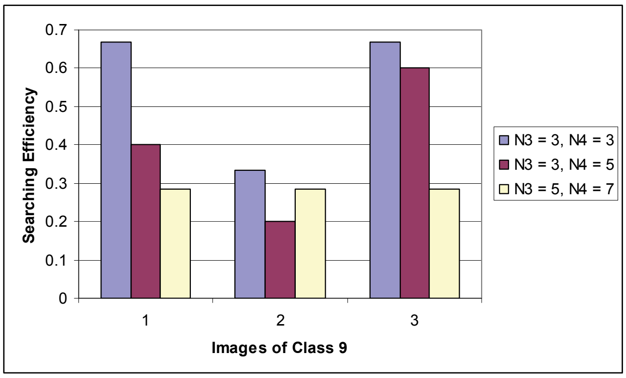

Fig. 11. Efficiency of searching procedure for images from class 9 for all three cases of initial search.

With the implementation of this system a set of relevant images is presented to the user. They are obtained by the automatic initial search and the application of intelligent logic. Using this system the user is not involved in the initial search of CBIR systems, but has the ability to increase the quality and efficiency of the searching process with the selection of relevant images in the next iteration. The processing of individual image feature leads to the reduction of dimensionality of the searching system. The user can select the individual image feature, or group of features, and starts the searching process. This selection does not affect ranking in the preprocessing phase.

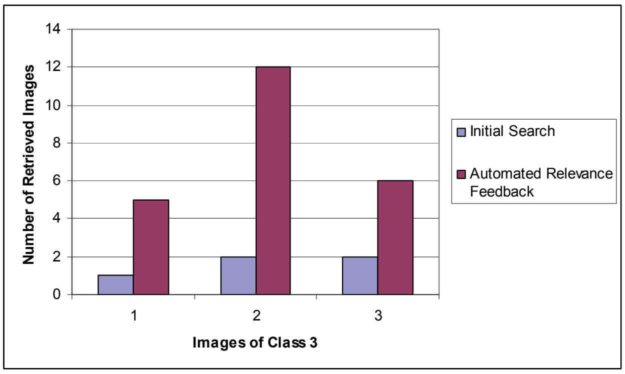

Fig. 12. Comparison of number of relevant images selected after initial search and after the first iteration of neural network for class 3 N_3=3, N_4=3 and N_5=30.

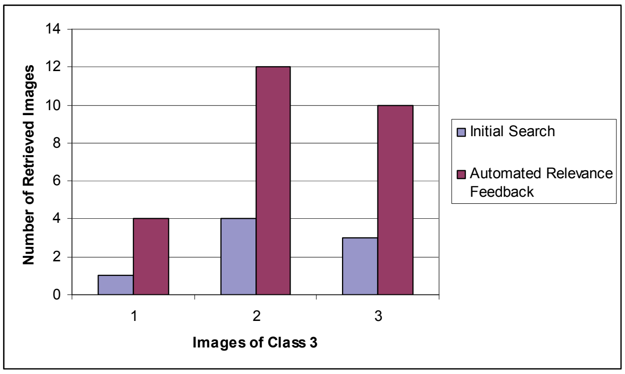

Fig. 13. Comparison of number of relevant images selected after initial search and after the first iteration of neural network for class 3 N_3=3, N_4=5 and N_5=30.

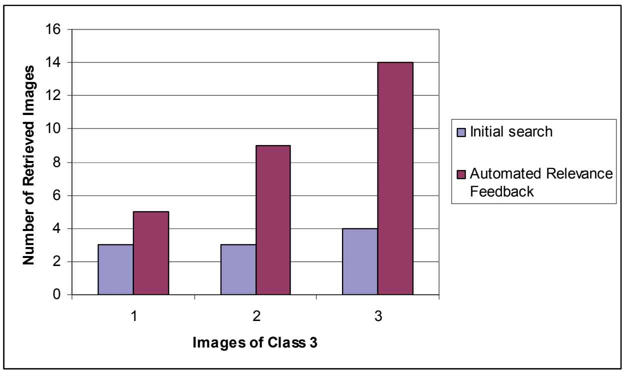

Fig. 14. Comparison of number of relevant images selected after initial search and after the first iteration of neural network for class 3 N_3=5, N_4=7 and N_5=30.

V. Conclusion

A CBIR system with relevance feedback and automated selection of relevant images in the initial search process is presented in this paper. Images in the database are described by three types of image features: color, texture and shape. Automated selection of relevant images is realized by ranking. Artificial intelligence is used for the correction of searching process based on selected relevant and irrelevant images. The main objective of this paper is realization of the system with selective processing of individual features. This kind of processing provides reduction of dimensionality of the image database based on user selection of relevant features with high efficiency in the searching process. The efficiency of searching process of the proposed CBIR system is satisfying for a small number of selected images in a preprocessing phase. Further work will focus on reducing the number of features, the speed of convergence of the algorithm and implementation of web-oriented structure.

References

[1] M. Flickner, H. Sawhney, W. Niblack, J. Ashley, Q. Huang and B. Dom, "Query by Image and Video Content: The QBIC System," *IEEE Computer*, vol. 28, no. 9, pp. 23–32, 1995.

[2] W. Y. Ma and B. S. Manjunath, "NETRA: A Toolbox for Navigating Large Image Databases," in *Proc. IEEE Int. Conf. on Image Processing (ICIP '97)*, vol. 1, pp. 568–571, Santa Barbara, CA, USA, October 1997.

[3] Y. Rui, T. Huang and S. Mehrotra, "Content-based Image Retrieval with Relevance Feedback in MARS," in *Proc. IEEE Int. Conf. on Image Processing*, pp. 815–818, 1997.

[4] J. Calic, N. Campbell, A. Calway, M. Mirmehdi, T. Burghardt, S. Hannuna, C. Kong, S. Porter, N. Canagarajah and D. Bull, "Towards Intelligent Content based Retrieval of Wildlife Videos," in *Proc. 6th Int. Workshop WIAMIS-05*, EFPL, Montreux, Switzerland, April 2005.

[5] Y. Alemu, J. Koh, M. Ikram, and D. Kim, "Image retrieval in multimedia databases: A survey," *Proc. of 5th International Conference on Intelligent Information Hiding and Multimedia Signal Processing*, 2009.

[6] S. Feng, D. Xu, and X. Yang, "Attention-driven salient edge(s) and region(s) extraction with application to CBIR," *Signal Processing*, vol. 90, no. 1, pp. 1-15, January, 2010.

[7] D. Liu, K. A. Hua, K. Vu, and N. Yu, "Fast query point movement techniques for large cbir systems," *IEEE Trans. on Knowl. and Data Eng*., 21(5):729–743, 2009.

[8] V. Radosavljević, N. Kojić, S. Čabarkapa, G. Zajić, I. Reljin, and B. Reljin, "An image retrieval system with user's relevance feedback," in *Proc. 7th Conf. WIAMIS-2006*, pp. 9-12, Seoul, Korea, 19-21, April 2006.

[9] P. Muneesawang, and L. Guan, "An interactive approach for CBIR using a network of radial basis functions," *IEEE Trans. Multimedia*, vol. 6, pp. 703-716, Oct. 2004.

[10] G. Zajić, N. Kojić, V. Radosavljević, M. Rudinac, S. Rudinac, Nikola Reljin, I. Reljin, and B. Reljin, "Accelerating of Image Retrieval in CBIR System with Relevance Feedback," *EURASIP Journal of Signal Processing, Spec. Issue on Knowledge Assisted Media Analysis for Interactive Multimedia Applications*, Volume 2007 (2007), Article ID 62678, 13 pages.

[11] G. Zajić, N. Kojić, N.Reljin, and B. Reljin, "Experiment with Reduced Feature Vector in CBIR System with Relevance Feedback," *The 5th International Conference on Visual Information Engineering*, VIE'08, July 29- August 1, Xi'an, China, 2008.

[12] G. Zajić and B. Reljin, "CBIR System With Automated Selection of Relevant Images", in *Proc. Conf. Scientific and Professional Symposium INFOTEH-2011*, Jahorina, 16-18. March, 2011.

[13] N. Kojić, G. Zajić, S. Čabarkapa, M. Pavlović, V. Radosavljević, and B. Reljin, "The Influence of the Feature Vector Content on the CBIR System Efficiency," TELSIKS 2009, *9th International Conference on Telecommunications in Modern Satellite, Cable and Broadcasting Services*, Nis-Serbia, 7.-9. October 2009.

[14] M. Yang, K. Kpalma, and J. Ronsin, "A Survey of Shape Feature Extraction Techniques," *Pattern Recognition Techniques, Technology and Applications*, Peng-Yeng Yin (Ed.) (2008) 43-90.

[15] Corel Gallery Magic 65000 (1999), www.corel.com