Research Article

Accelerating of Image Retrieval in CBIR System with Relevance Feedback

Goran Zajić,[1] Nenad Kojić,[1] Vladan Radosavljević,[2] Maja Rudinac,[1] Stevan Rudinac,[3] Nikola Reljin,[1] Irini Reljin,[1, 3] and Branimir Reljin[3]

[1] *College of Information and Communication Technologies, Belgrade, Serbia*

[2] *Computer and Information Sciences Department, Information Science and Technology Center, Temple University, Philadelphia, PA 19122, USA*

[3] *Digital Image Processing, Telemedicine and Multimedia Laboratory, Faculty of Electrical Engineering, University of Belgrade, Bulevar Kralja Aleksandra 73, 11000 Belgrade, Serbia*

Recommended by Ebroul Izquierdo

Content-based image retrieval (CBIR) system with relevance feedback, which uses the algorithm for feature-vector (FV) dimension reduction, is described. Feature-vector reduction (FVR) exploits the clustering of FV components for a given query. Clustering is based on the comparison of magnitudes of FV components of a query. Instead of all FV components describing color, line directions, and texture, only their representative members describing FV clusters are used for retrieval. In this way, the "curse of dimensionality" is bypassed since redundant components of a query FV are rejected. It was shown that about one tenth of total FV components (i.e., the reduction of 90%) is sufficient for retrieval, without significant degradation of accuracy. Consequently, the retrieving process is accelerated. Moreover, even better balancing between color and line/texture features is obtained. The efficiency of FVR CBIR system was tested over TRECVid 2006 and Corel 60 K datasets.

1. INTRODUCTION

The end of the last millennium was characterized by an explosive growth of digital technologies leading to widespread, cheap, but powerful, devices for audio-video data acquisition, processing, storing, and displaying. These new technologies, known as multimedia, have enabled the creation of huge digital multimedia libraries for personal entertainment, professional, and commercial use. Today, all aspects of human life are covered by appropriate digital record. Moreover, global networking through Internet permits us to be a part of a "global village" reaching any point over the Globe and using any available information. New technologies have a strong impact on our daily life and we changed our way of living, working, thinking, and learning. But, surprisingly, a growth of available information produces an opposite effect: more files less benefits. How to find relevant information into the ocean of available data? The effective data storage and management become highly important. There are constant and urgent needs for efficient indexing, searching, browsing, and retrieving of required data. Searching for textual data is more or less well suited. It is based on the similarity between key words and is applied in well-known and powerful browsing systems like Google, Yahoo, and so forth, which are commonly and permanently used from billions of consumers. Oppositely, searching for audio and video materials is not so easy due to perceptual limitations. Machine browsing and searching are based on the audio/image content (described by some objective measures like loudness, brightness, pitch, etc. for audio, and color, texture, shape, etc., for images) but there is a strong difference between an objective measure and subjective human perception [1].

Early work on image retrieval dated back to the 1980s and first systems were textual-oriented: appropriate annotations are manually associated to images describing as best as possible their visual content, enabling later the searching and browsing by using appropriate key words [2, 3]. Although such a technique may be very efficient and can be even automated, it suffers from several major drawbacks, especially when working with large databases. First, the process of manual annotating of database is extremely time-consuming. Although many public multimedia databases allow users to

freely annotate the image content (when the procedure of annotating may be significantly accelerated), there is at least one important negative consequence: different users may annotate images in different ways. Since the annotation is not unified, retrieval results are often unsatisfactory. Caused mostly by rapid development of the entertainment industry, systems for automatic annotating have been developed as well. Yet, procedures are still complicated and often unreliable. One of the significant drawbacks is caused by linguistic limitations. How to explain the content of particular image? Moreover, there is a need for precise description when annotating, and finding right combination of keywords when retrieving. Very often text descriptors are incomplete causing hard mismatches between user's needs and retrieving results.

To overcome drawbacks recognized in text-based approach, content-based image retrieval (CBIR) techniques are proposed. These techniques extract low-level image features such as color, texture, shape [4–8], from individual images and arrange them in some predetermined way forming an appropriate feature vector (FV). Retrieving procedure is based on relatively simple proximity measure between FVs to quantitatively evaluate the closeness (i.e., the similarity) between a query and images from database. These low-level content-based indexing techniques can be even automated to a high degree of accuracy, but in practice they still exhibit hard drawback usually reported as a "semantic gap" between the capabilities of low-level objective features and the users' subjective needs. A number of CBIR systems are reported [9–12].

The retrieving procedure can be significantly improved by introducing the user as a part of the retrieval loop. Starting with a query image, the system selects initial set of images from database, objectively more close to a query, and presents them to a user though appropriate graphic user interface (GUI). The user selects subjectively the best-matched samples and annotates them in appropriate way. From these samples, weights of preextracted features are updated, according to subjective perception of visual content. An active learning strategy exploits both positive and negative examples to gain feedback from user. Such a procedure, usually called *relevance feedback* (RF) [13, 14], is a way to effectively bridge the gap between the low-level image features and the high-level human perception. Typical architecture of a CBIR system with RF is depicted in Figure 1.

In all CBIR systems, we are faced with the problem of producing low-level image features that accurately describe human visual perception. Additional problem is related to computational complexity. Intuitively thinking, it is expected that high-dimensional feature vector gives better information about the image content. Yet, except the computational complexity when working with high-dimensional vectors, this expectation is not verified in machine learning due to the "curse of dimensionality" [15]. Many nondominant low-level features may produce a masking effect and false decision. To overcome this problem, several methods for dimension reduction are reported. These methods can be classified into two general categories: linear dimension reduction (LDR) and nonlinear dimension reduction (NLDR).
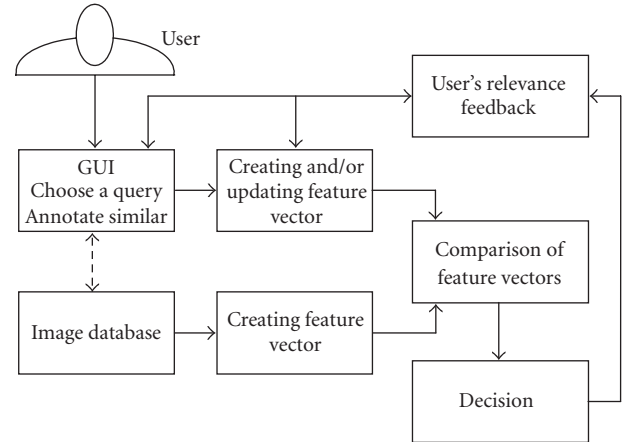

FIGURE 1: Typical architecture of CBIR system with relevance feedback.

Typical examples for LDR are principal component analysis (PCA) and singular value decomposition (SVD), which find the low-dimensional subspace of eigenvalues that capture the most variance of original dataset. The LDR works well with linear correlated datasets, but may be inadequate in processing of inherently nonlinear phenomena, such as most of the natural signals. For nonlinear phenomena, better results are expected by using NLDR, for instance, nonlinear PCA [16] or some other nonlinear methods embedded into the neural network approach [17].

In this paper, a CBIR system with relevance feedback, which exploits feature vector reduction (FVR), is described. Our method for data reduction is very simple but effective. It is based only on the comparison of magnitudes of adjacent FV components. All images from database are indexed by numerals, and for each image, an FV describing the image content (color, texture, edge direction, and cooccurrence matrix) is performed, as usual. Initially, FVs are high dimensional (having 556 components in our case), but the searching procedure uses significantly reduced number of components, enabling faster and even more reliable search. Reduction is based on clustering of FV components. When loading a query image, its FV components are calculated and compared with their neighbors. Components having the magnitude within the prescribed limits are declared as components belonging to the same cluster. Then, each cluster is described by its representative element. Instead of all FV components only cluster representatives are used in the searching procedure. From intensive simulations, we verified that the FV reduction of about 90% is possible without significant degradation of accuracy, while the searching process is accelerated and even better balancing between color, and line/texture features is obtained.

The paper is organized as follows. Section 2 briefly reviews the related work on feature vector reduction. Section 3 presents the proposed FVR CBIR system. Experimental results performed over images from TRECVid 2006 and Corel

60 K datasets are given in Section 4, and obtained results are compared with those known from literature. Section 5 consists of concluding remarks.

2. RELATED WORK

In any CBIR system, some preprocessing of images from database is necessary. It includes the determination of relevant low-level features (such as color, texture, shape) describing as best as possible the content of each image i, $i = 1, 2, \ldots, I$. Features are expressed by corresponding numerical values, and are grouped into appropriate feature vector $\mathbf{F}_i = [F_i(1), F_i(2), \ldots, F_i(j), \ldots, F_i(J)]$ of the length J. Each coordinate $j = 1, 2, \ldots, J$ of a vector $\mathbf{F}_i$ corresponds to particular feature component. Feature vectors were stored in appropriate *feature matrix*, $\mathbf{F} = \{\mathbf{F}_i\}$, of dimension $I \times J$. Then, the retrieving procedure is based on relatively simple proximity measure $d_i = d(\mathbf{F}_q, \mathbf{F}_i)$, $i = 1, 2, \ldots, I$, (e.g., Euclidean distance, Mahalanobis, or similar) between a query feature vector $\mathbf{F}_q$ and feature vectors $\mathbf{F}_i$, $i = 1, 2, \ldots, I$, associated with images from database. Image i with the smallest distance d_i is objectively the closest (i.e., more similar) to a query. After initial search, which is based on objective measure, the retrieving procedure may be improved by using user's relevance feedback [13, 14, 18–26].

Intuitively, as many feature components J are used, better accuracy in first retrieving step is expected. Yet, conversely, retrieving process then becomes slower and even useless, in case of huge databases, because a query has to be compared with all images from database. Additional problem is known as a "curse of dimensionality" when a number of redundant FV components may degrade the retrieving procedure. It is necessary to apply some dimension reduction technique to eliminate redundancy among low-level features. Several dimension reduction methods are suggested for CBIR systems. These methods are based mainly on the principal component analysis (PCA) [27–31] and on the linear discriminant analysis (LDA) [32–37]. The PCA finds the low-dimensional subspace that captures the most variance of original dataset, that is, this method extracts the *most descriptive* features. The objective of LDA is to perform dimensionality reduction while preserving as much of the class discriminatory information as possible. This way, the LDA constructs *most discriminative* features. The LDA was successfully used in face recognition [38]. Several improvements are further embedded into the LDA, for instance, biased discrimination analysis (BDA) [39] and direct kernel BDA (DKBDA) [40]. The nonlinear dimension reduction method, which is better suited for nonlinear nature of data features, is proposed for handling feature vectors for music data [41]. Furthermore, in CBIR systems with relevance feedback [13, 14], the number of positive and negative examples annotated by a user is relatively small (20 to 30, caused by a limited space on the screen), dictating the choice of the learning method which has to be embedded into the system. One very efficient learning method working with small sample dataset is the support vector machine (SVM) method [42], which is exploited in CBIR RF systems [43, 44].

3. FVR CBIR SYSTEM DESCRIPTION

3.1. Preliminary considerations

By closer inspection of feature vector $\mathbf{F}_i$ for a given image, it can be concluded that its components $F_i(j)$, $j = 1, 2, \ldots, J$, may have significantly different magnitudes, since they are calculated in different ways. Components with higher values will be dominant in determining an objective distance d_i between a query and images from database and may produce unfair competition and even a masking effect. To avoid the dominance of such components and permit fair influence of all FV patterns, each term $F_i(j)$ in a feature matrix is columnwise rescaled as described in [13]. Typical examples of rescaled FV obtained for a real image from Corel 60 K dataset [45], with 556 components describing color, edges and texture, are depicted in Figure 2. But even after rescaling, as we can see from Figure 2, feature vector components still have significantly different values. Some of the components are dominant, having values close to unity, while a number of components have very small value or are even zero-valued. Consequently, not all of the feature vector components have the same influence on the objective similarity measure. Moreover, a number of nondominant components can produce the masking effect, inhibiting the influence of dominant components, leading thus to the false decision. These facts are taken into account in our CBIR system with a feature vector reduction.

3.2. Feature vector reduction

In our CBIR system, we started with 556 components describing low-level image features: color, line directions, and texture. Feature vector components are ordered as follows: 32 coordinates for dominant colors in HSV (hue-saturation-value) space, 32 coordinates for dominant colors in YCbCr (luminance Y and two chrominance components: Cb = Y−B, Cr = Y−R) space, 164 coordinates for HSV histogram (coded as $18 \times 3 \times 3 = 162$, while two more components are their mean and standard deviations, SD), 177 coordinates for YCbCr histogram ($7 \times 5 \times 5 + \text{mean} + \text{SD}$), 73 coordinates describing histogram of line directions (72 directions with 5-degree step, while the last coordinate corresponds to nonclassified directions), 62 coordinates describing texture by Gabor transform coefficients, and 16 coordinates from gray-level cooccurrence matrix. Feature vector components are columnwise rescaled as described in [13] and have typical form as in Figure 2.

In proposed CBIR system, feature vector reduction is based on clustering of FV components of the given query. Block scheme of our FVR CBIR system is depicted in Figure 3. Before starting the clustering process, a user defines the tolerance T (in percents) characterizing the elements belonging to the same cluster. The clustering process is performed as follows. For a given query image, its feature vector with 556 elements is created, exactly in the same way as for images from database. The first component of the query FV, $F_q(1)$ is assumed as the first component of the first cluster, C_1, and is compared with the next component $F_q(2)$. If
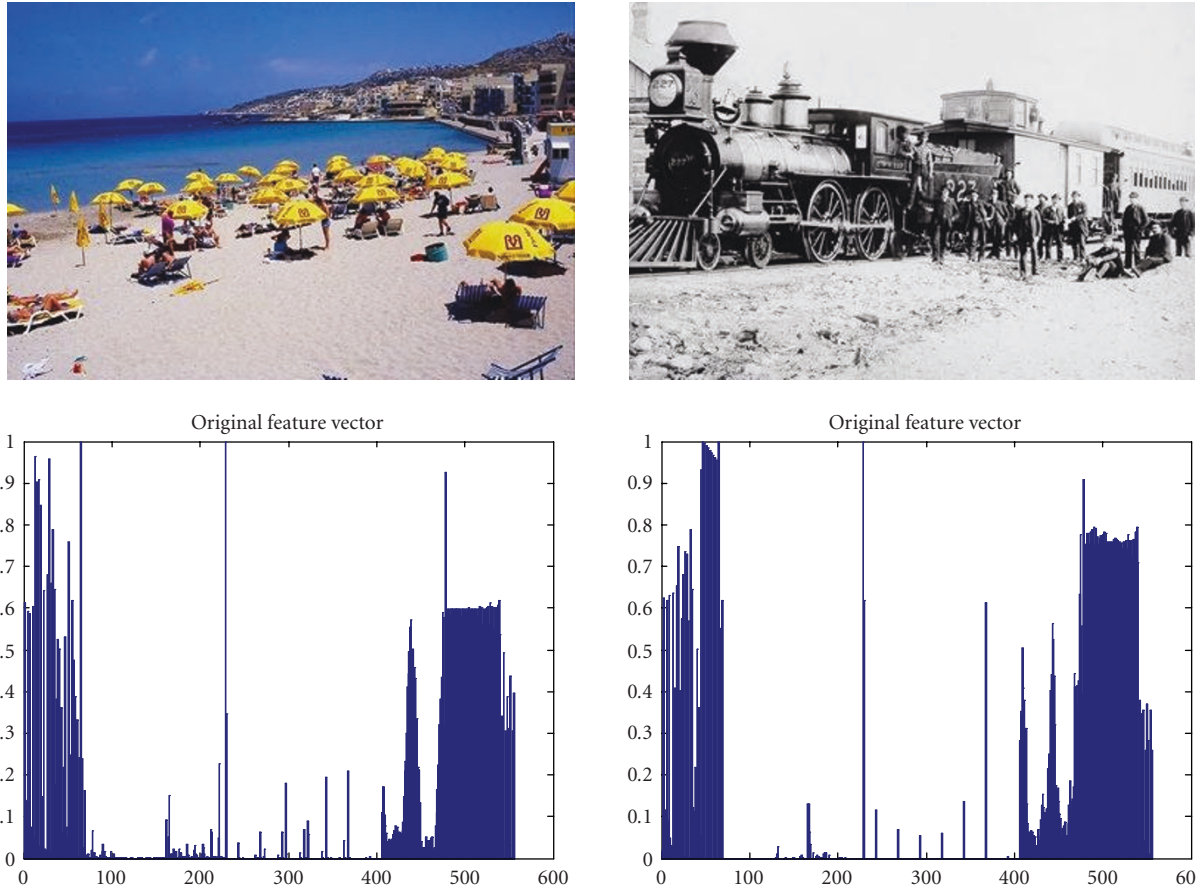
Figure 2: "Beach" scene (left) and "Train" image (right) from Corel 60 K dataset and their feature vectors with 556 components describing color (first 405 components), line directions, and texture (last 151 components).
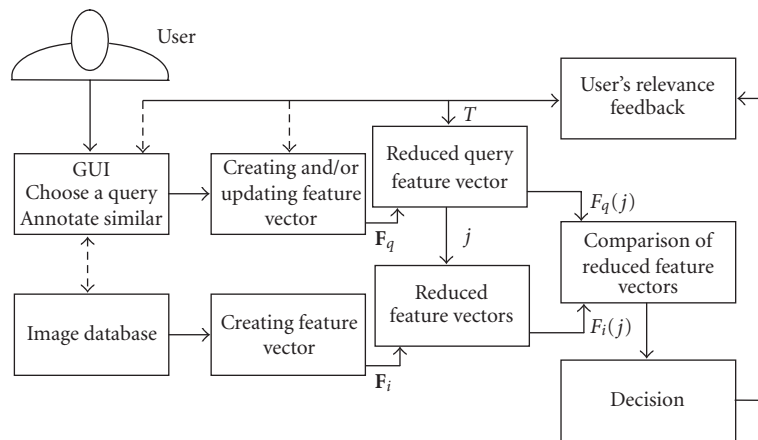

Figure 3: Block scheme of proposed FVR CBIR system with relevance feedback.

the relative absolute difference (RAD), described by (1), of the first cluster component and the next query FV component is within the prescribed tolerance T, elements $F_q(1)$ and $F_q(2)$ are assumed belonging to the same (first) cluster, and the next component $F_q(3)$ is compared with C_1, and so forth.

If for some jth element of the query FV, the RAD is greater than prescribed tolerance T, this element $F_q(j)$ is declared as the first element of the next (second) cluster, and is denoted as C_2. The previous cluster then is closed and the new cluster is created in the same way. The procedure is repeated for

TABLE 1: Results of FVR for "Beach" scene and" Train" image from Corel 60 K dataset with prescribed tolerance of $T = 80\%$.

Feature	Number of components before FVR	Number of components after FVR	
		Beach	Train
Dominant colors in HSV space	32	0	2
Dominant colors in YCbCr space	32	1	9
Color histogram in HSV space	164	10	3
Color histogram in YCbCr space	177	14	6
Histogram of line directions	73	9	11
Dominant texture features	62	10	10
Grey-level cooccurrence matrix	16	12	16
Sum	556	56	57

all FV components of a query, and as a result K clusters are created.

The relative absolute difference (in percents) for the kth cluster is calculated as

$$\mathrm{RAD}_k = \frac{\mathrm{abs}(C_k - F_q(j))}{C_k} \times 100, \quad k = 1, 2, \ldots, K, \quad (1)$$

where C_k is the first element of the kth cluster, $F_q(j)$ is the jth component of the query FV. The described algorithm is applied to all FV elements, but separately on color (first 405 components) and line/texture features (last 151 components) of a query image. Also, clusters are formed after two scans of FV components: from left to right (LR set of clusters), that is, from coordinates $j = 1$ to $j = 556$, and vice versa (RL set), that is, from $j = 556$ to $j = 1$, and final clusters are obtained as an intersection of two obtained sets: LR ∩ RL. After forming clusters, each of them is represented by only one element. From intensive simulations, we found that the query FV component with the highest magnitude within a cluster is the best cluster representative. The position j of this component and its magnitude $F_q(j)$ are temporarily stored. At the retrieving process, for images from database, only their FV components $F_i(j)$ from the same positions j corresponding to cluster representatives are used, as indicated in Figure 3. In this way, since the number of clusters K can be significantly lower than the number J of all feature vector components, the retrieving process will be accelerated accordingly. The rest of our system is of the structure that we already used in CBIR systems without feature vector reduction [25, 26]. As a similarity measure, we used Mahalanobis distance while updating of the query feature vector is performed with assistance of radial basis function neural network.

Characteristic results after applying proposed FV reduction method are presented in Figure 4, where reduced FVs of images from Figure 2 are depicted. Tolerance of $T = 80\%$ is assumed. The first row consists of reduced FVs with exact positions j of components within the whole FV with 556 coordinates. The second row consists of temporarily stored components describing only FV cluster representatives. As we can infer, the reduction of about 10 times is obtained:

the number of elements in reduced FVs (i.e., the number of clusters) now equals $K = 56$ for a "Beach" scene (left) and $K = 57$ for a "Train" image (right), instead of initial number of 556 components. Note that two images from Figure 2 perceptually are quite different having also different FVs. Qualitative description of these two images requires different features. Colorful "Beach" image requires more color histogram features (components between coordinates $j = 65$ to $j = 405$ in initial FV) for qualitative description, while gray image "Train" requires more dominant color features: components with coordinates $j = 1$ to $j = 64$. In both cases, line and texture features (151 components in total, from coordinates $j = 406$ to $j = 556$) are very important and without reduction they would be probably masked by larger number of color features. Our FV reduction method eliminates redundant components into the query FV and produces better balancing between color and line/texture features. Also, the method is case sensitive depending on the content of particular query feature vector, as illustrated in Figure 4 and Table 1, where a comparison between reduced FVs for the same example is performed. The second column in Table 1 contains the number of particular feature components in FV prior to reduction. The third and fourth columns are related to reduced FVs for the images "Beach" and "Train," respectively. As we can infer, from initial ratio 405 : 151 (meaning color versus line/texture features) reduced FVs have the ratio of 25 : 31 for "Beach" image and even 20 : 37 for a "Train" image. After FV reducing, the influence of color on the image retrieval becomes less expressed.

Since it is more likely that images from database, which are similar to a query, have similar values of corresponding FV components, it is expected that such clustering of FV components will not produce significant degradation of retrieving. This assumption is illustrated in Figure 5, which represents a set of ten images from MIT database [46] closest to a query (far left image labeled by 1), after the first retrieval pass (only objective measure, without RF). The first row consists of retrieved images without FV reduction (tolerance $T = 0\%$). Reduction of about two times ($T = 20\%$), second row, produces no influence in ordering of first ten closest images. Moreover, further reductions of 3.56 : 1 and
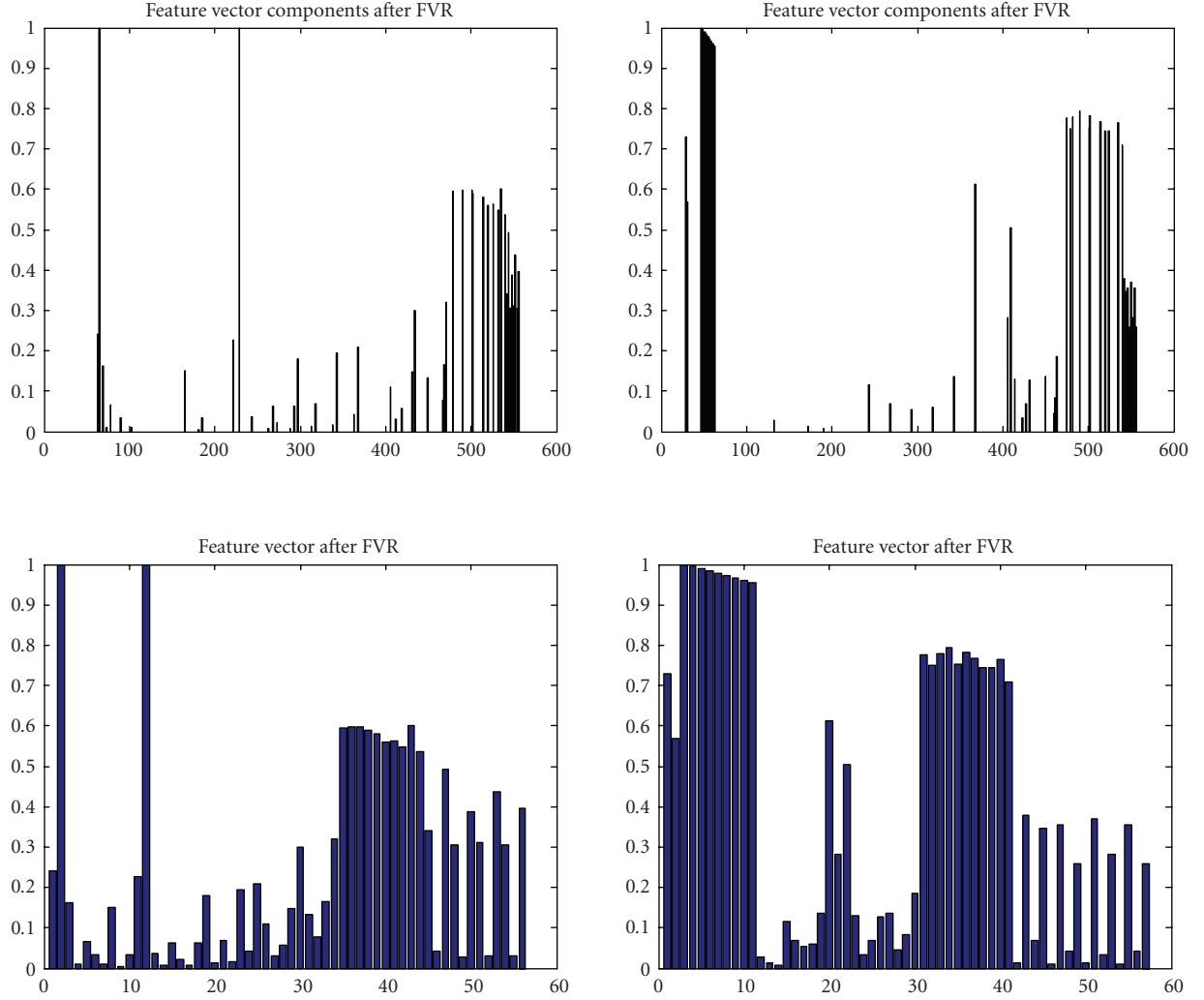
FIGURE 4: Reduced feature vectors of Corel images "Beach" and "Train," as in Figure 2, if tolerance is $T = 80\%$.

even 10.3 : 1 ($T = 40\%$ and $T = 80\%$, resp.) have not any influence on the ordering of first 6 closest images.

4. TESTING OF FVR CBIR SYSTEM

A feature vector reduction described in Section 3 is embedded into our CBIR relevance feedback module reported in [25, 26]. A system is tested by using images from unclassified TRECVid 2006 dataset and Corel 60 K dataset [45]. TRECVid dataset is composed of 146 587 keyframes extracted from 259 video clips from TV news. Only referent keyframes (79 484 in total) are used for testing. Corel 60 K dataset consists of 60,000 images from 600 semantic classes ("dogs," "horses," "forests," "beach," "buses," etc.) each having 100 images. Note that folders' names are not quite adequate because many images with similar contents are not in the same folder and some quite different images are in the same semantic folder. Yet, since our system is purely content-based, we performed retrieving process without referring to image folders.

For all images from datasets, full-length feature vectors (with 556 components) are created. In the searching procedure, at the first step the user is asked to determine desired tolerance T necessary for creating reduced FV, and the number B of first best-matched images which is presented to a user from appropriated GUI as in Figure 7. The first step in retrieving process was purely objective, based only on the distances (Mahalanobis) between the FVs of a query and images from datasets. Next steps include the RF performed by assistance of RBF neural network as in [25, 26]. Three different retrieving scenarios are performed as follows.

(1) Search with full-lengths feature vectors (all 556 components), without FVR.

(2) Search with reduced FVs ($T = 80\%$, reduction of about 90%) only in the first retrieving step, while the RF is performed over the full-length FVs.

(3) Search by using reduced FVs ($T = 80\%$, reduction of about 90%) in all retrieving steps.

TABLE 2: Simulation results P_{20} for all three scenarios over Corel 60 K and TRECVid 2006 test sets.

Scenario	Without FVR			FVR in the first step			FVR in both steps		
Step	1st	2nd	3rd	1st	2nd	3rd	1st	2nd	3rd
Dataset	Precision P_{20} (%)								
TRECVid 2006	60.25	64.75	69.25	50.5	60.75	67.75	50.5	56.5	60.5
Corel 60 K	33.5	56	66.5	28.75	51	60.25	28.75	45	53.75

Without reduction, 556 elements

Tolerance 20%, 272 elements, reduction 2.04 : 1

Tolerance 40%, 156 elements, reduction 3.56 : 1

Tolerance 80%, 54 elements, reduction 10.3 : 1

FIGURE 5: First 10 images objectively closest to a query (left) after applying feature vector reduction of prescribed tolerance.

As a query, we used 20 randomly selected images from both datasets and for each scenario. Three retrieving steps (one objective and two with RF) are performed in each experiment so the total number of steps was $2 \times 20 \times 3 \times 3 = 360$. (Note that in our research we counted the objective retrieval (without RF) as a first retrieving step, instead of many authors who labeled this step as a zero step and counted only RF steps, e.g., in [40].) The quality of retrieving processes was evaluated by using the precision as a performance measure,

$$P_B = \frac{R}{B} \times 100. \qquad (2)$$

The precision P_B is defined as the ratio of the number of subjectively relevant images (R) versus the top of B best-matched images presented to a user. Three independent users evaluated the retrieval process. The values of P_{20} (for $B = 20$ displayed images) are presented in Table 2 and Figure 6.

Retrieving results over TRECVid 2006 test set are slightly better, particularly at the first retrieving step, see Figure 6(a): the precision P_{20} of 50%–60% is obtained versus 30% for Corel images, although in the first dataset the total number of images of the same class, for some queries, was less than the number of displayed images ($B = 20$). As we can see from Table 2 and Figure 6, the best averaged results are obtained by applying the first scenario, without FVR. The reason for that is, probably, that for randomly selected images used for evaluation, a larger feature vector gives most detailed information about image. However, note that for some cases, results when applying FVR may be the same or even better than without
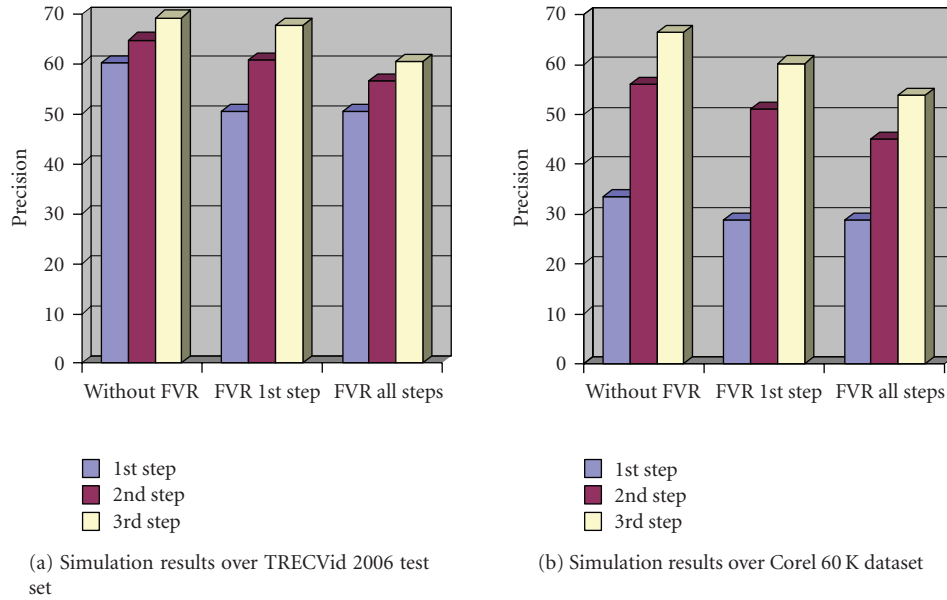
(a) Simulation results over TRECVid 2006 test set

(b) Simulation results over Corel 60 K dataset

FIGURE 6: Retrieving precision P_{20} (averaged) for three retrieval steps (first is objective, others are with RF), for TRECVid 2006 and Corel 60 K datasets.

FVR—compare Figures 7(a), 7(b), 8(a), and 8(b), where two examples from TRECVid 2006 and Corel 60 K datasets, are presented. The precision P_{10} of 70% (TRECVid 2006) and 40% (Corel) is obtained after the first pass of FVR, compared to 70% and 50%, respectively, if no FVR is applied. This assumption is in accordance to results in [47], where authors also founded that reduction of 90% can lead to even better retrieving than the use of full-dimensional vector. Note that they considered only color features and testing is performed over Corel dataset with 6 192 images and TRECVid 2003 with 32 318 keyframes.

The second scenario, which deploys feature vector reduction only in the first step, is second ranked in average, while the worst results are achieved under the third scenario (reduced FVs in all steps), but even then the results are quite satisfactory: after the first pass the precision P_{20} of 50% (TRECVid) and 30% (Corel) is obtained, and of 60% or 52% after the third step (second RF step). When using larger number of iterations (through RF module), results for all three scenarios converge to the same limit of about 90% or more, for precision P_{20} as a performance figure of merit. These results are comparable to those recently reported in the paper of Tao et al. [40]. They considered Corel dataset with 10 800 images and the feature vector with 521 components (393 for color and 128 for texture) prior to reduction, and used direct kernel biased discriminant analysis and SVM. Under the same conditions as we used (objective retrieval and two RF iterations), they gained the precision P_{20} of about 56% and of about 95% after $1 + 9$ iterations ([40, Figure 4]).

In our approach, feature vector reduction of about 90% decreases computational time for about 15% to 25%, compared to the case without FVR. Using Pentium machine with AMD Athlon 64 processor 32000+, with 2.01 GHz, and the

memory DIMM 2 GB DDR/400 MHz Kingston, the execution time for one retrieving step without FVR was about 47 seconds for TRECVid 2006 (processing all of 80 000 images) and about 23 seconds for Corel 60 K dataset (60 000 images). When applying FVR of 90%, the execution time reduces to 40-41 seconds (TRECVid) and to 17-18 seconds (Corel). As we can conclude, execution time is not in linear dependence with the dataset dimension. We also tested our system over small datasets of only several thousands of images when the execution time was less than 0.03 second. Note also that in our experiments, none of optimizations are applied to computer programs. It is expectable that the retrieving procedure will be accelerated after appropriate optimization of computer programs.

5. CONCLUSION

The paper considers the feature vector reduction in CBIR system. Our system uses standard feature vector describing color, line directions, and texture, having 556 components without reduction. Here we propose the FV reduction based on clustering of FV components of given query. Components of a query FV with similar magnitudes are grouped into clusters and each cluster is described by its representative element: by its position j in a full-length FV and corresponding value $F_q(j)$. In this way, the method rejects redundant components of a query FV and produces better balancing between color and line/texture features, as well. Then components of FVs of images from database are temporarily selected in the same way: for images $i = 1, 2, \ldots, I$ from database only their components $F_i(j)$ corresponding to positions j of cluster representatives are used in searching procedure, instead of all FV elements. In this way, the retrieving process is accelerated for about 20% compared to retrieving
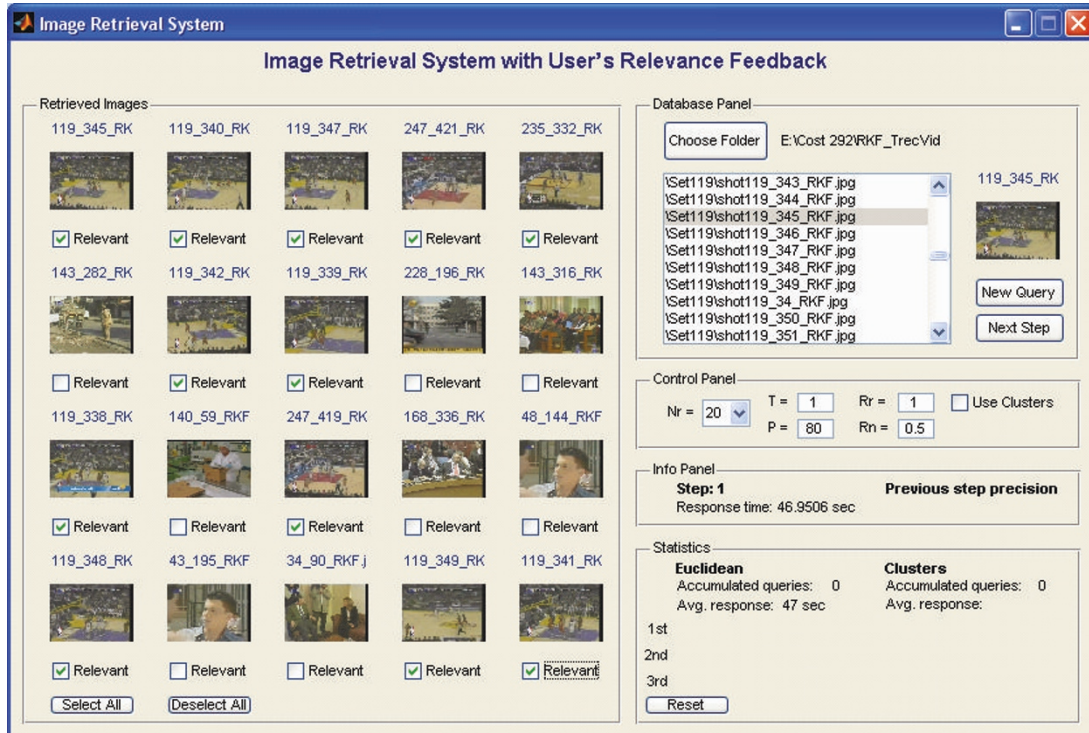
(a) TRECVid 2006 image 119_345_RK.jpg. First step (only objective retrieval) without feature vector reduction. Execution time is about 47 seconds. Precision $P_{10} = 70\%$, $P_{20} = 60\%$
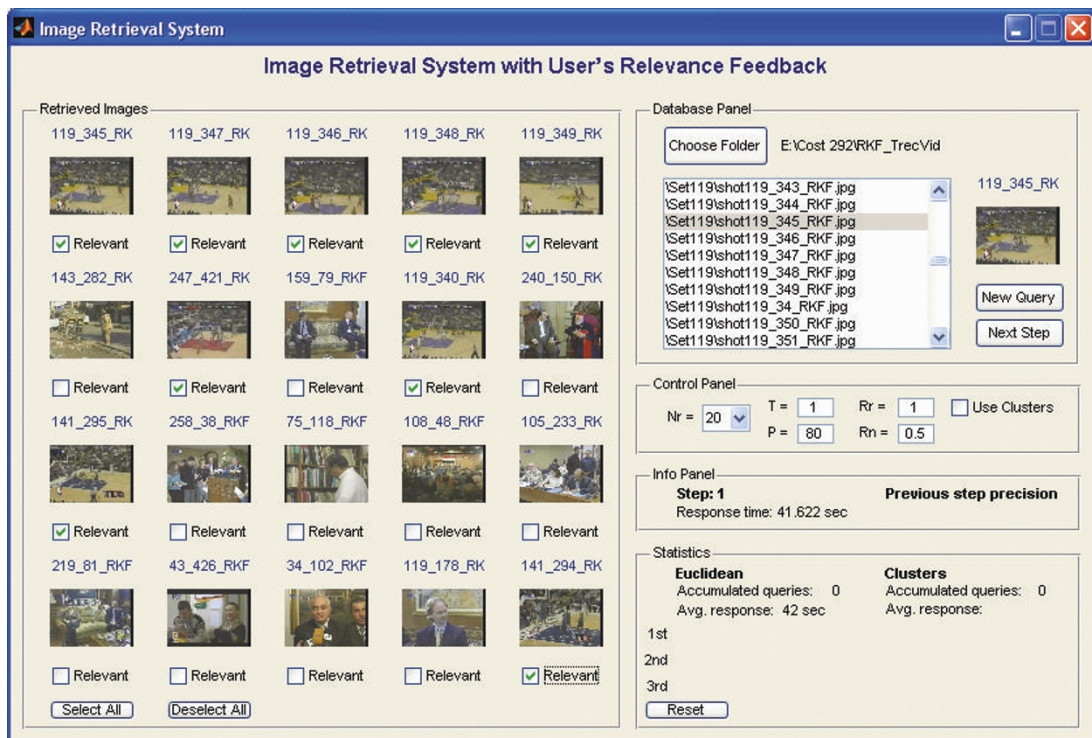

(b) TRECVid 2006 image 119_345_RK.jpg. First step (only objective retrieval) with feature vector reduction of about 90%. Execution time is about 42 seconds. Precision $P_{10} = 70\%$, $P_{20} = 45\%$

Figure 7: Retrieving results after the first step for TRECVid 2006 image 119_345_RK.jpg: (a) without feature vector reduction; (b) feature vector reduction of 90%.
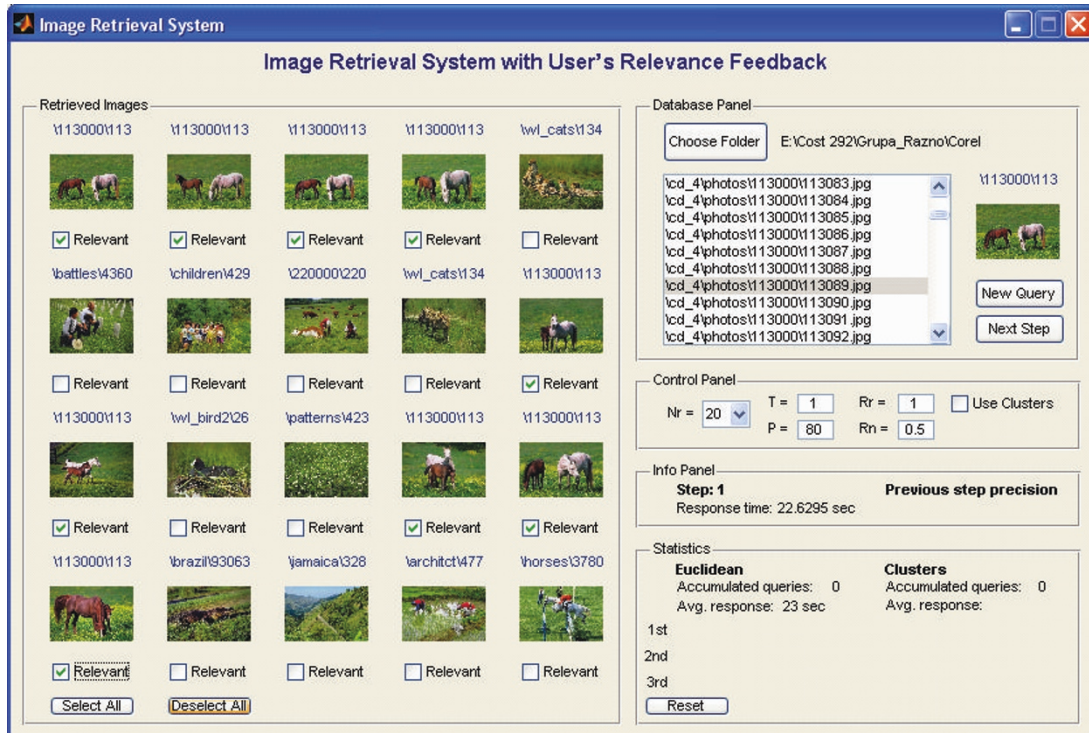
(a) Corel 60 K image 13089.jpg. First step without feature vector reduction. Execution time is about 23 seconds. Precision $P_{10} = 50\%$, $P_{20} = 45\%$
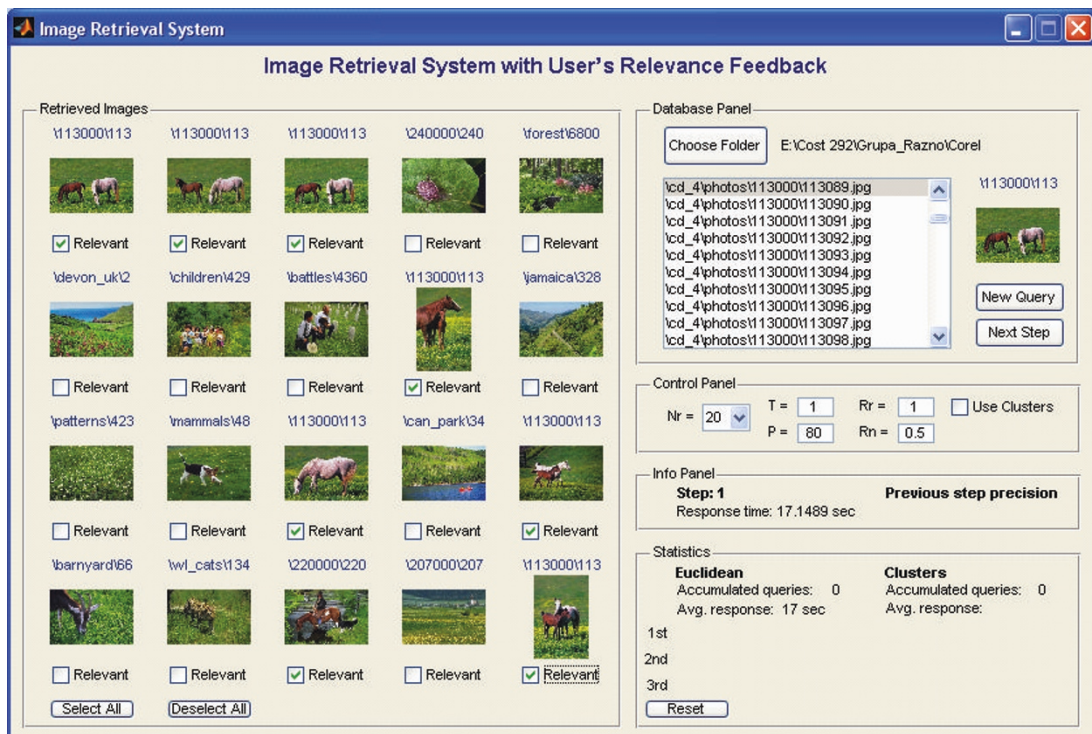
(b) Corel 60 K image 13089.jpg. First step, feature vector reduction of 90%. Execution time is about 17 seconds. Precision $P_{10} = 40\%$, $P_{20} = 40\%$

FIGURE 8: Retrieving results after the first step for Corel 60 K image 13089.jpg: (a) without feature vector reduction; (b) feature vector reduction of 90%.

with full-length FVs, without significant degradation of accuracy. Moreover, since FV reduction is performed for a given query, clustering process is adaptive to a content of observed image, that is, the method is case sensitive, depending on particular query. Proposed algorithm for dimension reduction is simple and consequently fast and well suited for cases when existing database should be updated. Adaptability and efficiency of proposed FVR algorithm was tested over TRECVid 2006 dataset of about 80 000 key frames, and Corel image database with 60 000 images. Our results are comparable to those recently reported in [40, 47]. In our future work we will investigate the possibility of combining hierarchical search and our FVR algorithm, expecting further improvements in the retrieving procedure.

ACKNOWLEDGMENTS

REFERENCES

[1] D. Feng, W. C. Siu, and H. J. Zhang, Eds., *Multimedia Information Retrieval and Management*, Springer, New York, NY, USA, 2003.

[2] N.-S. Chang and K.-S. Fu, "Query by pictorial example," *IEEE Transactions on Software Engineering*, vol. 6, no. 6, pp. 519–524, 1980.

[3] S.-K. Chang and T. L. Kunii, "Pictorial data-base systems," *IEEE Computer Magazine*, vol. 14, no. 11, pp. 13–21, 1981.

[4] M. J. Swain and D. H. Ballard, "Color indexing," *International Journal of Computer Vision*, vol. 7, no. 1, pp. 11–32, 1991.

[5] C. W. Niblack, R. Barber, W. Equitz, et al., "QBIC project: querying images by content, using color, texture, and shape," in *Storage and Retrieval for Image and Video Databases*, vol. 1908 of *Proceedings of SPIE*, pp. 173–187, San Jose, Calif, USA, February 1993.

[6] H. Tamura, S. Mori, and T. Yamawaki, "Textural features corresponding to visual perception," *IEEE Transactions on Systems, Man and Cybernetics*, vol. 8, no. 6, pp. 460–473, 1978.

[7] B. S. Manjunath and W. Y. Ma, "Texture features for browsing and retrieval of image data," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 18, no. 8, pp. 837–842, 1996.

[8] A. K. Jain and A. Vailaya, "Shape-based retrieval: a case study with trademark image databases," *Pattern Recognition*, vol. 31, no. 9, pp. 1369–1390, 1998.

[9] A. P. Pentland, R. W. Picard, and S. Scarloff, "Photobook: tools for content-based manipulation of image databases," in *Storage and Retrieval for Image and Video Databases II*, vol. 2185 of *Proceedings of SPIE*, pp. 34–47, San Jose, Calif, USA, February 1994.

[10] M. Flickner, H. Sawhney, W. Niblack, et al., "Query by image and video content: the QBIC system," *Computer*, vol. 28, no. 9, pp. 23–32, 1995.

[11] J. R. Bach, C. Fuller, A. Gupta, et al., "Virage image search engine: an open framework for image management," in *Storage and Retrieval for Still Image and Video Databases IV*, vol. 2670 of *Proceedings of SPIE*, pp. 76–87, San Jose, Calif, USA, February 1996.

[12] W. Y. Ma and B. S. Manjunath, "NeTra: a toolbox for navigating large image databases," in *Proceedings of IEEE International Conference on Image Processing (ICIP '97)*, vol. 1, pp. 568–571, Santa Barbara, Calif, USA, October 1997.

[13] Y. Rui, T. S. Huang, and S. Mehrotra, "Content-based image retrieval with relevance feedback in MARS," in *Proceedings of IEEE International Conference on Image Processing (ICIP '97)*, vol. 2, pp. 815–818, Santa Barbara, Calif, USA, October 1997.

[14] Y. Rui, T. S. Huang, M. Ortega, and S. Mehrotra, "Relevance feedback: a power tool for interactive content-based image retrieval," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 8, no. 5, pp. 644–655, 1998.

[15] T. M. Mitchell, *Machine Learning*, McGraw-Hill, New York, NY, USA, 1997.

[16] J. Karhunen and J. Joutsensalo, "Representation and separation of signals using nonlinear PCA type learning," *Neural Networks*, vol. 7, no. 1, pp. 113–127, 1994.

[17] S. Haykin, *Neural Networks: A Comprehensive Foundation*, John Wiley & Sons, New York, NY, USA, 1999.

[18] J. Peng, B. Bhanu, and S. Qing, "Probabilistic feature relevance learning for content-based image retrieval," *Computer Vision and Image Understanding*, vol. 75, no. 1, pp. 150–164, 1999.

[19] G. Aggarwal, T. V. Ashwin, and S. Ghosal, "An image retrieval system with automatic query modification," *IEEE Transactions on Multimedia*, vol. 4, no. 2, pp. 201–214, 2002.

[20] G. Lu, "Techniques and data structures for efficient multimedia retrieval based on similarity," *IEEE Transactions on Multimedia*, vol. 4, no. 3, pp. 372–384, 2002.

[21] P. Muneesawang and L. Guan, "An interactive approach for CBIR using a network of radial basis functions," *IEEE Transactions on Multimedia*, vol. 6, no. 5, pp. 703–716, 2004.

[22] K.-M. Lee and W. N. Street, "Cluster-driven refinement for content-based digital image retrieval," *IEEE Transactions on Multimedia*, vol. 6, no. 6, pp. 817–827, 2004.

[23] B. Ko and H. Byun, "FRIP: a region-based image retrieval tool using automatic image segmentation and stepwise Boolean AND matching," *IEEE Transactions on Multimedia*, vol. 7, no. 1, pp. 105–113, 2005.

[24] J. Calic, N. Campbell, A. Calway, et al., "Towards intelligent content based retrieval of wildlife videos," in *Proceedings of the 6th International Workshop on Image Analysis for Multimedia Interactive Services (WIAMIS '05)*, EFPL, Montreux, Switzerland, April 2005.

[25] S. Čabarkapa, N. Kojić, V. Radosavljević, G. Zajić, and B. Reljin, "Adaptive content-based image retrieval with relevance feedback," in *Proceedings of the International Conference on Computer as a Tool (EUROCON '05)*, vol. 1, pp. 147–150, Belgrade, Serbia, November 2005.

[26] V. Radosavljević, N. Kojić, S. Čabarkapa, G. Zajić, I. Reljin, and B. Reljin, "An image retrieval system with user's relevance feedback," in *Proceedings of the 7th International Workshop on Image Analysis for Multimedia Interactive Services (WIAMIS '06)*, pp. 9–12, Seul, Korea, April 2006.

[27] J. B. Kruskal and M. Wish, *Multidimensional Scaling*, Sage, Beverly Hills, Calif, USA, 1977.

[28] I. T. Jolliffe, *Principal Component Analysis*, Springer, New York, NY, USA, 2nd edition, 2002.

[29] K. I. Diamantaras and S. Y. Kung, *Principal Component Neural Networks*, John Wiley & Sons, New York, NY, USA, 1996.

[30] M. Turk and A. P. Pentland, "Eigenfaces for recognition," *Journal of Cognitive Neuroscience*, vol. 3, no. 1, pp. 71–86, 1991.

[31] T. Serre, B. Heisele, S. Mukherjee, and T. Poggio, "Feature Selection for Face Detection," MIT A.I. Memo no. 1697, September, 2000.

[32] R. A. Fisher, "The use of multiple measurements in taxonomic problems," *Annals of Eugenics*, vol. 7, pp. 179–188, 1936.

[33] R. A. Fisher, "The statistical utilization of multiple measurements," *Annals of Eugenics*, vol. 8, pp. 376–386, 1938.

[34] P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman, "Eigenfaces vs. Fisherfaces: recognition using class specific linear projection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19, no. 7, pp. 711–720, 1997.

[35] K. Etemad and R. Chellappa, "Discriminant analysis for recognition of human face images," *Journal of the Optical Society of America A*, vol. 14, no. 8, pp. 1724–1733, 1997.

[36] Y. Wu, Q. Tian, and T. S. Huang, "Discriminant-EM algorithm with application to image retrieval," in *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '00)*, vol. 1, pp. 222–227, Hilton Head Island, SC, USA, June 2000.

[37] K. Fukunaga, *Statistical Pattern Recognition*, Academic Press, New York, NY, USA, 2nd edition, 1990.

[38] H. Yu and J. Yang, "A direct LDA algorithm for high-dimensional data—with application to face recognition," *Pattern Recognition*, vol. 34, no. 10, pp. 2067–2070, 2001.

[39] X. S. Zhou and T. S. Huang, "Small sample learning during multimedia retrieval using BiasMap," in *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '01)*, vol. 1, pp. 11–17, Kauai, Hawaii, USA, December 2001.

[40] D. Tao, X. Tang, X. Li, and Y. Rui, "Direct kernel biased discriminant analysis: a new content-based image retrieval relevance feedback algorithm," *IEEE Transactions on Multimedia*, vol. 8, no. 4, pp. 716–727, 2006.

[41] J. Shen, J. Shepherd, and A. H. H. Ngu, "Towards effective content-based music retrieval with multiple acoustic feature combination," *IEEE Transactions on Multimedia*, vol. 8, no. 6, pp. 1179–1189, 2006.

[42] V. N. Vapnik, *The Nature of Statistical Learning Theory*, Springer, New York, NY, USA, 1995.

[43] L. Zhang, F. Lin, and B. Zhang, "Support vector machine learning for image retrieval," in *Proceedings of IEEE International Conference on Image Processing (ICIP '01)*, vol. 2, pp. 721–724, Thessaloniki, Greece, October 2001.

[44] G.-D. Guo, A. K. Jain, W.-Y. Ma, and H.-J. Zhang, "Learning similarity measure for natural image retrieval with relevance feedback," *IEEE Transactions on Neural Networks*, vol. 13, no. 4, pp. 811–820, 2002.

[45] Corel Gallery Magic 65000 (1999), http://www.corel.com/.

[46] http://vismod.media.mit.edu/pub/VisTex/.

[47] P. Howarth and S. Rüger, "Trading precision for speed: localised similarity functions," in *Proceedings of the 4th International Conference on Image and Video Retrieval (CIVR '05)*, vol. 3568 of *Lecture Notes in Computer Science*, pp. 415–424, Springer, Singapore, July 2005.

Goran Zajić is a Teaching and Research Assistant at the ICT College in Belgrade, Serbia. He received the Dipl.Ing. degree (5-year university degree) in electrical engineering from the Faculty of Electrical Engineering, University of Belgrade, Serbia. From 2005, he is a member of group for Digital Image Processing, Telemedicine and Multimedia Laboratory (IPTM Group), at the Faculty of Electrical Engineering in Belgrade. As Member of the IPTM group, he participated in COST 292 Action "Semantic multimodal analysis of digital media," Working Group 5. His research interests include neural network, signal processing, and multimedia analysis. G. Zajić is coauthor of 2 journal papers and 10 conference papers.

Nenad Kojić received the Dipl.Ing. degree (5-year university degree) in 2003, and M.S. degree in 2006, from the Faculty of Electrical Engineering, University of Belgrade, Serbia. Currently he is a Ph.D. student at the Faculty of Electrical Engineering, University of Belgrade. He is a College Professor at the ICT College, Belgrade. His research interests include neural network, routing algorithms, heterogeneous wireless networks, image processing, and multimedia. He is a coauthor of 1 journal paper and 15 conference papers. He is involved in the European Project COST 292 Action "Semantic multimodal analysis of digital media," Working Group 5.

Vladan Radosavljević received the Dipl.Ing. degree (5-year university degree) in electrical engineering in 2003, from the Faculty of Electrical Engineering, University of Belgrade, Serbia. Currently, he is working towards the Ph.D. degree in computer science at the Temple University, Philadelphia, USA. His research interests include spatial-temporal data mining, content-based image retrieval, and signal processing. He is a coauthor of 1 journal paper and 9 conference papers. He is involved in the European project COST 292 Action "Semantic multimodal analysis of digital media," Working Group 5.

Maja Rudinac received the Dipl.Ing. degree in electrical engineering (5-year university degree) in 2006, from the Faculty of Electrical Engineering, University of Belgrade, Serbia. She is currently employed by College of Information and Communication Technology, Belgrade, as Teaching and Research Assistant. Her research interests include digital image processing, multimedia content analysis, content-based image and video retrieval, and medical signal processing. As member of IPTM Group, group for digital Image Processing, Telemedicine and Multimedia on Faculty of Electrical Engineering in Belgrade, she participates in the COST 292 Action "Semantic multimodal analysis of digital media," Working Group 5. Maja Rudinac is a coauthor of 1 journal paper and 7 conference papers.

Stevan Rudinac received the Dipl.Ing. degree in electrical engineering (5-year university degree) in 2006, from the Faculty of Electrical Engineering, University of Belgrade. He is currently working as a Research Assistant in Digital Image Processing, Telemedicine and Multimedia Laboratory (IPTM) at the Faculty of Electrical Engineering, University of Belgrade. His research interests cover broad areas of multimedia content analysis, multimedia information retrieval, digital signal processing, and digital image processing with focus on content-based image and video retrieval. He is involved in the

COST 292 Action "Semantic multimodal analysis of digital media," Working Group 5. Stevan Rudinac is a coauthor of 1 journal paper and 7 conference papers.

Nikola Reljin received the Dipl.Ing. degree in electrical engineering (5-year university degree) in 2006, from the Faculty of Electrical Engineering, University of Belgrade, Serbia. He has working experience in web programming, projecting, maintenance, and installation of TV equipment. Currently, he is a Teaching and Research Assistant at the ICT College in Belgrade, Serbia. From 2005, he is a member of group for Digital Image Processing, Telemedicine and Multimedia (IPTM Group), at the Faculty of Electrical Engineering in Belgrade. He is involved in the COST 292 Action "Semantic multimodal analysis of digital media," Working Group 5. His research interests include web programming, signal processing and multimedia analysis, and management. Nikola Reljin is a coauthor of 1 journal paper and 6 conference papers.

Irini Reljin received the degree (5-year university degree), M.S., and the Ph.D. degrees in electrical engineering, all from the Faculty of Electrical Engineering (FEE) University of Belgrade, Serbia. Since 1983, she is with the ICT College in Belgrade, working as a College Professor. Since 2001 she joined the FEE, University of Belgrade, as an Assistant Professor, teaching the mulimedia and video technologies at undergraduate studies, as well as neural networks applications at graduate studies. She has published over 20 journal papers and over 150 conference presentations, as well as several book chapters, and has given a number of invited lectures on different aspects of communications, signal and image processing, fractal and multifractal analyses, content-based indexing, and retrieval. She has participated in a number of scientific and research projects in the areas of telecommunications, multimedia, and telemedicine, and currently she participated in COST 292 Action "Semantic multimodal analysis of digital media." Her research interests are in video and multimedia analyses, digital image processing, neural networks, statistical signal analysis, fractal and multifractal analyses. She is a Member of the IEEE, SMPTE (Society of Motion Pictures and Television Engineers), BSUAE (Trans Black Sea Union of Applied Electromagnetism), Gender Team, as well as several national societies.

Branimir Reljin received the Dipl.Ing. degree (5-years university degree), the M.S. and the Ph.D. degrees in electrical engineering, all from the Faculty of Electrical Engineering (FEE) University of Belgrade, Serbia. Since 1974 he is joined at the FEE, University of Belgrade passing all teaching positions. He has given a number of invited lectures in institutions and universities in Serbia and Montenegro, as well as in Europe and in the USA. He has published more than 350 papers in technical journals and conferences, four books and several book chapters. Branimir Reljin was a Project Leader for many national and international projects, and currently he is a Coordinator of Working Group 5 in COST Action 292 "Semantic multimodal analysis of digital media." He is a member of several scientific and professional societies, and has a Senior Member Grade of the IEEE Society.

Also, he is an IEEE Serbia and Montenegro CAS&SP Chair, and a member of the editorial boards of several journals and a number of conferences. He is a General Chair of the IEEE cosponsored symposia on Neural Network Applications in Electrical Engineering (NEUREL), and a guest editor of the special issue on "Neural Network Applications in Electrical Engineering," in the Neurocomputing Journal, Elsevier, 2007.