# Statistical Analysis of Feature Vector Relevance in CBIR System

Goran J. Zajić[1], Nenad S. Kojić[1], Nikola B. Reljin[1], Branimir D. Reljin[2]

*Abstract* – **An influence of feature vector (FV) components on retrieving accuracy in CBIR system with relevance feedback is considered. System uses FVs with only 24 components describing color, line directions and texture. The reduction of FV dimension is based on the statistics of global image features. The proposed system was tested over Corel 1K dataset. The statistics of FV influence was described by the number of images belonging or not belonging to the query class (in-class vs. out-class images) under different conditions.**

*Keywords* – **Low-level image features, image retrieval, relevance feedback, feature vector reduction**.

## I. Introduction

The explosive growth of powerful but cheap technologies in last decades has lead to mass production and usage of different multimedia devices. Professional and personal users are allowed to create, upload and download different data. As a result, a measureless amount of all-genres files are stored in memory devices and/or circulates through the Internet. To avoid an information collapse, many systems for indexing, searching, browsing and retrieving of multimedia content were created. First, and still very popular, solutions were completely text-oriented. Appropriate keywords are associated to files (mainly, but not only, to images) and the searching/retrieving is based on the text similarity [1]. Unfortunately, since image data contains very rich information, it is very difficult to capture the content of an image using only a few keywords. Also, the manual annotation process is quite subjective, ambiguous, incomplete, and time-consuming.

One promising way for overcome drawbacks recognized in text-base approach was the content-based image retrieval (CBIR) technique. In CBIR systems the low-level image features (color, texture, shape, etc.) are used as objective descriptors of images or their parts [2-4]. From those features an appropriate feature vector (FV) was created for each image from dataset. Then the retrieving procedure is based on relatively simple proximity measure between FVs to quantitatively evaluate the closeness (i.e., the similarity) between a query (key image or a user supplied sketch) and images from database. A number of CBIR systems are reported [5-8]. Although the CBIR techniques produce very good results in retrieving, the so-called "semantic gap", between the low-level objective features (content), used by computers, and the high-level semantics (context), recognized by humans, is a hard limiting factor. A very efficient way of resolving this drawback introduces a user in the searching/retrieving process. Such an approach is known as the (user's) *relevance feedback* (RF). First step of retrieving in RF system is of the standard CBIR form: for a given query, system calculates the distances between FVs and selects images from database which are (objectively) more close to a query, and presents them to a user, for evaluation. The user annotates subjectively best-matched samples. From these samples weights of pre-extracted 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. In this way the semantic gap may be bridged efficiently, as reported [9-10].

In all CBIR systems at least two problems exist, provoking researchers to find as best as possible solutions. One problem relates to the difference between objective features and subjective image content. It is necessary to find low-level image features that describe as best as possible the human visual perception. A variety of features are suggested and even standardized, for instance in MPEG-7 Standard [11]. The second problem is addressed to the number of feature vector components. Intuitively thinking, it is expected that high-dimensional feature vector gives better information about the image content and leads to better accuracy in retrieving. But, except the computational complexity, this expectation is not verified in machine learning, due to the "curse of dimensionality" [12]. Many non-dominant low-level features may produce a masking effect and even false decision. To overcome this problem, different methods for reduction of FV dimension, to eliminate redundancy among low-level features, are suggested [13-19].

In this paper the analysis of the relevance of particular components of the FV on the image retrieving, from the statistical point of view, was performed. We analyzed the CBIR RF system proposed in [20]. This system, which uses only 24 components describing color, line directions and texture, will be briefly described in Section II. Section III. refers to the statistical analysis of the relevance of feature vector components to the retrieving, while Section IV is addressed to some concluding remarks.

## II. The CBIR RF System with Reduced FV

The CBIR RF system proposed in [20] uses only 24 low-level FV components describing color, line directions and texture. Reduced FVs are derived from full-length FVs inspired by MPEG-7 descriptors [11], as already applied in CBIR RF system reported in [21]. Before FV reduction we started with the FVs containing 310 components, in total. The color (COL)

[1] Goran J. Zajić, Nenad Z. Kojić and Nikola B. Reljin are with the ICT College, Zdravka Čelara 16, 11000 Belgrade, Serbia, E-mails: gzajic@gmail.com, {nenad.kojic;nikola.reljin}@ict.edu.yu

[2] Branimir D. Reljin is with the Faculty of Electrical Engineering, University of Belgrade, Bulevar Kralja Aleksandra 73, 11000 Belgrade, Serbia, E-mail: reljinb@etf.bg.ac.yu

is described by 162 components of the histogram in HSV space (coded as 18x3x3). Line (LIN) directions, for 5 degree steps, are described by corresponding histogram (72 components), Gabor (GAB) wavelet coefficients uses 60 components (6 directions with 5 scales, each described by its mean and standard deviation), and from gray-level co-occurrence (COO) matrix 16 components are used (four directions: 0, 45, 90 and 135 degrees, each described by four descriptors: energy, entropy, contrast and inverse differential moment). From such four-group 310-component FVs, we created reduced four-group 24-component FVs, based on global statistic of image features.

From 162 components of the HSV color histogram first three dominant components, normalized to their total number (162), are used for the new FV. From these components, denoted as $DC1$, $DC2$ and $DC3$, the next three components are calculated, according to (1)

$$\begin{aligned} RC1 &= \frac{1}{8-1}\sqrt{(DC1-DC2)^2+\ldots+(DC1-DC8)^2} \\ RC2 &= \frac{1}{8-2}\sqrt{(DC2-DC3)^2+\ldots+(DC2-DC8)^2} \\ RC3 &= \frac{1}{8-3}\sqrt{(DC3-DC4)^2+\ldots+(DC3-DC8)^2} \end{aligned} \tag{1}$$

Relations (1) describe the relevance of first three dominant components ($DC1,2,3$) within the rest of first eight dominant components ($DC1$ to $DC8$).

The same reduction procedure is applied to the histogram of line directions. First 3 dominant components, say $DL1$, $DL2$ and $DL3$, are included directly, and the next three components, denoted as $RL1$, $RL2$ and $RL3$, are calculated according to relation (1), now applied to line histogram.

From initial set of 60 Gabor features, we calculated 4 new values: from 30 means (M) and 30 standard deviations (S) of GAB wavelet coefficients we calculated their means and standard deviations: GMM (Gabor mean of means), GSM (standard deviation of means), GMS (mean of standard deviations), and GSS (standard deviation of standard deviations). Finally, from co-occurrence matrix we derived 8 components: mean and standard deviation for each of 4 components (energy, entropy, contrast, and inverse moments) related to four directions. In this way reduced feature vectors consist of only 24 components describing global statistics of color (6 components), line direction (6), and texture (4+8=12).

The retrieving procedure is performed using the CBIR RF system as in [21]. Query may be loaded externally or internally, from database. Before searching user can select feature group(s) (COL, LIN, GAB, and/or COO) which will be used in retrieving process, and define their tolerances $\Delta_J$. First retrieving step is pure objective, based on the similarity between FV components of a query ($FVQ$) and images from database ($FVD$). The Euclidean distance was used as a similarity metric. FV components from selected groups: COL, LIN, GAB, and/or COO, are compared separately, and after each comparison a set of $B$ images, objectively best-matched to a query, is presented to user, for evaluation. Images selected as relevant ($R$) are used in the relevance feedback procedure, described in [21]. As a performance measure we used the precision, $P_B$, defined as the ratio of the number of $R$ images versus the top $B$ images

$$P_B = \frac{R}{B}\times 100\,. \tag{2}$$

The initial searching step may be a bottleneck, since a query has to be compared to a whole dataset. For accelerating this step, before calculating Euclidean distance, components of feature vectors of images from database, $FVD_{i,j}$, are compared with corresponding components of a query, $FVQ_j$

$$\varepsilon_{i,j} = \left|FVQ_j - FVD_{i,j}\right| \,. \tag{3}$$

In (3) $i = 1,2,\ldots,I$ denotes the image from database, and $j = 1,2,\ldots,J$ refers to the $j$th component of the FV. Images from database are then preselected according to the following decision rule

$$\varepsilon_{i,j} - \Delta_j = \begin{cases} >0, & \text{skip this image} \\ \text{otherwise, take this image} & \end{cases} \tag{4}$$

where the quantity $\Delta_j$ is given (predetermined) tolerance. Preselection was performed separately for each group of features. For COL and LIN features only their first three components ($DC1,2,3$, and $DL1,2,3$) are used for testing (4).

Before testing, initial tolerances $\Delta_{\text{init}}$ and a maximal number of images (say, $T$) satisfying the test (4), was defined. We used $\Delta_{\text{init}}$=0.005 for COL features, $\Delta_{\text{init}}$=0.001 for other features, and $T$=50.

Image preselecting was performed in an iterative way, assuming three or two conditions depending on the COL/LIN or GAB/COO features, as will be described, briefly.

### *A. Preselecting based on COL and LIN Components*

First testing (4) is applied to all three dominant colors ($DC1,2,3$), jointly. If the number of images, say $COL_1$, satisfying the condition (4) for all three $DC1,2,3$ components, is less than $T$, the condition (4) is applied to $DC1$ and $DC2$, and selected images, $COL_2$, are added to images $COL_1$. If the sum $COL_1$ and $COL_2$ is less than $T$, the same procedure is applied only to $DC1$, and selected images, $COL_3$, are added to $COL_1$ and $COL_2$. If the sum $COL_1$, $COL_2$ and $COL_3$, is still less than $T$, the tolerance threshold $\Delta_{\text{init}}$ is extended, and the procedure is repeated with $DC1,2,3$, $DC1,2$, or only $DC1$, etc., until the number of selected images reaches $T$.

The same procedure is applied to line directions, but now initial tolerance is $\Delta_{\text{init}}$=0.001, and components $DL1$, $DL2$, and $DL3$ are used.

### *B. Preselecting based on GAB and COO Components*

For Gabor features we started with initial tolerance $\Delta_{\text{init}}$=0.001, and all four components (GMM, GSM, GMS and GSS). If the number selected images, $GAB_1$, according to (4) is less than $T$, the testing (4) is applied only to mean values, GMM and GMS. The procedure is repeated with higher value of tolerance threshold, until the number of selected images reaches $T$.

Similarly, for the co-occurrence features, initial selection is performed with $\Delta_{\text{init}}$=0.001 and using all 8 components: means and standard deviations for each of 4 components (energy, entropy, contrast, and inverse moments) related to four directions. If the number of selected images satisfying condition (4) is less than $T$, in the second pass only means are considered. If necessary, the procedure continues with higher tolerance until the number of selected images reaches $T$.

In this way from a whole dataset a group of $4T$ (200 in our case) images, satisfying the condition (4) is created. Then, the Euclidean distances between a query and this set of images are calculated and ordered images are presented to a user for evaluation and the RF procedure is started, as in classical RF system. Note that for RF we always use all 24 FV components (6 COL, 6 LIN, 4 GAB and 8 COO components). In preselection, maximal number of components is 18, if all four groups of features are used for testing given by (4).

Described CBIR RF system is tested over Corel 1K [22] dataset. A Corel 1K dataset contains 1000 images sorted in ten classes, with 100 images each, labeled as: Africa (code numbers 0-99), beaches (100-199), monuments (200-299), busses (300-399), dinosaurs (400-499), elephants (500-599), flowers (600-699), horses (700-799), mountains (800-899), and cookies/food (900-999). Corel 1K dataset is very homogeneous: images within the same class are quite similar (except several cases), while classes significantly differ. So, for evaluating the searching method, we can test its ability to retrieve images from the same class as a query (irrespective of, possible, subjective mismatching).

TABLE I: PRECISION $P_{20}$ OBTAINED USING FULL-LENGTH FVS AND FVR1 [21], FVR2 [23], AND PROPOSED (NEW) FV REDUCTION [20].

| Class | Full-FV | FVR1 | FVR2 | New FVR |
|---|---|---|---|---|
| 0 - 99 | 71.5 / 82.5 | 70.2 / 82.7 | 65.0 / 85.5 | 63.2 / 82.5 |
| 100 - 199 | 34.0 / 56.0 | 46.2 / 61.8 | 56.5 / 64.5 | 45.3 / 62.2 |
| 200 - 299 | 40.0 / 63.5 | 39.5 / 63.7 | 35.0 / 59.0 | 41.2 / 68.5 |
| 300 - 399 | 67.5 / 88.0 | 69.2 / 91.2 | 69.5 / 92.5 | 69.8 / 91.3 |
| 400 - 499 | 100 / 100 | 99.3 / 100 | 96.5 / 100 | 98.2 / 100 |
| 500 - 599 | 54.0 / 77.5 | 53.5 / 76.7 | 49.0 / 75.0 | 53.4 / 80.3 |
| 600 - 699 | 67.0 / 99.5 | 62.0 / 87.1 | 60.5 / 82.0 | 61.7 / 77.4 |
| 700 - 799 | 78.5 / 86.0 | 77.6 / 89.2 | 78.5 / 94.0 | 77.8 / 88.3 |
| 800 - 899 | 30.0 / 54.0 | 34.4 / 59.8 | 37.0 / 65.5 | 38.2 / 67.4 |
| 900 - 999 | 56.0 / 74.5 | 52.3 / 81.7 | 48.5 / 73.0 | 49.3 / 75.8 |
| **Total** | **59.9 / 78.2** | **60.4 / 79.4** | **59.6 / 79.1** | **59,8 / 79.4** |

Simulations are performed using standard Pentium machine (2GHz clock, 2GB DDR). The efficiency of proposed system with feature vector reduction (FVR) was compared to our previous results obtained with a system with full-length FVs and systems with FVRs as in [21] and [23], denoted respectively as FVR1 and FVR2. Statistics of retrieving efficiency, described by precision $P_{20}$, are given in Table I. In columns 2-5 first number relates to the first (objective) retrieving step and the second one corresponds to first relevance feedback. Note that in some cases the FVR may produce even better retrieving result after the first step (shadowed cells in Table 1), due to better balancing between color and texture components. Averaged execution time for one retrieving step was about 70, 25, 15, and 12 milliseconds, respectively for full-FV, FVR1, FVR2, and proposed FVR. Note that execution times are given only as a comparative measure of different methods, since in our experiments no optimizations are involved in computer programs.

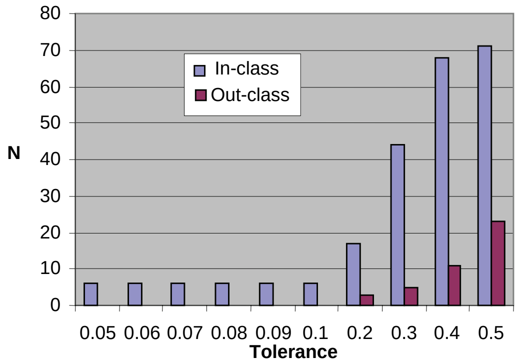
## III. STATISTICAL ANALYSIS OF FV RELEVANCE

Our research was targeted to finding possible relation between objective low-level descriptors and subjectively classified image classes. As a first step in such research, the statistical analysis of FV components in CBIR system with relevance feedback, as described in Section 2, was performed, and obtained results are presented in this paper. Statistical analysis of FV components was derived considering only color ($DC1,2,3$) and texture (GMM, GSM, GMS, GSS) descriptors. A part of results is given in Table II and Fig. 1 and in Fig. 2. Among images satisfying condition (4) we denote those belonging to a class of given query as „In“ (meaning, in-class). Other images, satisfying condition (4), but not belonging to a query class, are denoted as „Out“ images. In Table II. rows C-In/Out correspond to color test, and those denoted as G-In/Out correspond to Gabor test.

As expected, condition (4) applied to joint coordinates (all three dominant colors, or all four GAB coordinates) was very hard and only small number of images satisfies this condition: about 6 images for COL test and less than 11 for GAB test, if tolerance $\Delta$ is less than to 0.1. Also, GAB features are not so precise for selecting image classes: the number of out-class images is greater than in-class ones. By increasing $\Delta$ the number of in-class images increases (very fast for color) but also the number of out-class images (particularly for GAB test.

TABLE II: STATISTICS FOR IN/OUT-CLASS IMAGES FOR DIFFERENT TOLERANCES $\Delta$.

| $\Delta$ | 0.05 | 0.06 | 0.07 | 0.08 | 0.09 | 0.1 | 0.2 | 0.3 | 0.4 | 0.5 |
|---|---|---|---|---|---|---|---|---|---|---|
| C-In | 6 | 6 | 6 | 6 | 6 | 6 | 17 | 44 | 68 | 71 |
| C-Out | 0 | 0 | 0 | 0 | 0 | 0 | 3 | 5 | 11 | 23 |
| G-In | 3 | 4 | 5 | 6 | 9 | 11 | 35 | 67 | 77 | 86 |
| G-Out | 4 | 10 | 13 | 16 | 23 | 29 | 94 | 218 | 403 | 548 |



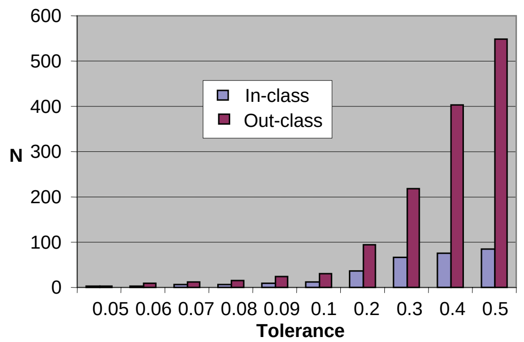




Fig. 1. Statistics for COL and GAB features. The number of In/Out-class images after the first (joint) condition (4) for different $\Delta$.

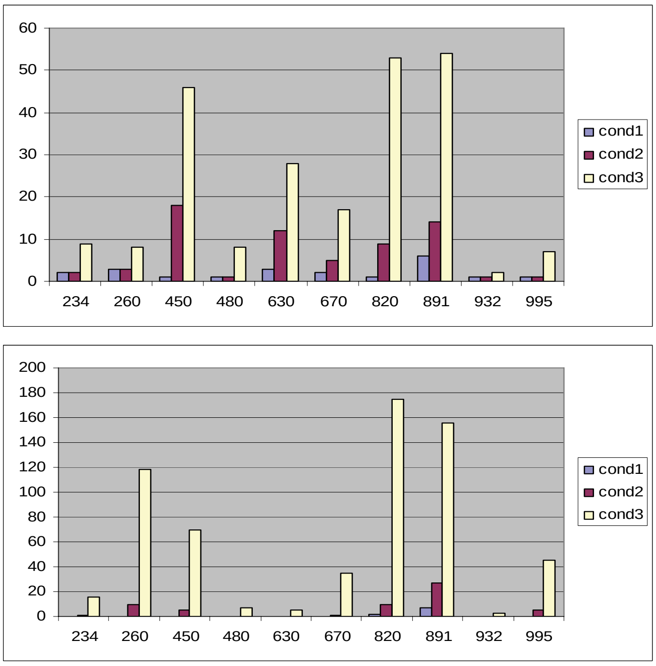



Fig. 2. Statistics for COL features with tolerance Δ=0.05, for 5 Corel classes. The number of In-class images (up) and out-class images (down) after applying all three conditions (4): cond1 (joint *DC*1,2,3), cond2 (joint *DC*1,2), cond3 (only *DC*1).

Fig. 2 presents the statistics for COL features for 5 Corel classes The number of in-class images increases with less restrictive condition, but also the out-class images, being several times greater than in-class ones.

## IV. Conclusions

In this paper we investigated the influence of particular FV components to accuracy of image retrieval. We used the CBIR system with short FVs of only 24 components describing color, lines and texture [20]. Before retrieving, the image dataset was preprocessed, neglecting images with FV components out of prescribed tolerance. Image classification based on only color feature gives satisfactory good results. The system is tested over Corel 1K dataset. In further research we will consider the efficiency of our system over other datasets.

## Acknowledgements

This work is a part of activities of the Laboratory for Digital Image Processing, Telemedicine and Multimedia (IPTM) from the University of Belgrade, Serbia, related to COST Action 292 "Semantic multimodal analysis of digital media."

## References

[1] S. K. Chang, T. L. Kunii, "Pictorial database systems", *IEEE Trans Software Eng.*, vol. 6, no. 6, pp. 519-524, Nov. 1980.
[2] M. Swan, D. Ballard, "Color indexing", *Int. J. Comput. Vision*, vol. 7, no. 1, pp. 11-32, 1991.
[3] W. Niblack, R. Barber, W. Equitz, M. Flickner, E. Glasman, D. Petkovic, P. Yanker, C. Faloutsos, G. Taubino, "The QBIC project: Querying images by content using color, texture, and shape", in *Proc. SPIE: Storage and Retrieval of Images and Video Databases*, pp. 173-181, Feb. 1993.
[4] B. Manjunath, W. Ma, "Texture features for browsing and retrieval of image data", *IEEE Trans. Pattern Anal. Machine Intelligence*, vol. 18, no. 8, pp. 837-842, 1996.
[5] A. Pentland, R. Picard and S. Sclaroff, "Photobook: Tools for content-based manipulation of image databases", in *Proc SPIE Conf. Storage and Retrieval of Still Image and Video Databases II*, pp 34-47, San Jose, CA, USA, 1994.
[6] 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, 1995.
[7] J. R. Back, C. Fuller, A. Gupta, A. Hampapur, B. Horowitz, R. Jain and C. F. Shu, "The virage image search engine: An open framework for image management", in *Proc. SPIE Conf. Storage and Retrieval for Still Image and Video Databases IV*, pp. 76-87, San Jose, CA, USA, February 1996.
[8] W. Y. Ma and B. S. Manjunath, "NETRA: A toolbox for navigating large image databases", in *Proc. IEEE Int. Conf. on Image Processing*, 1997.
[9] 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.
[10] Y. Rui and T. Huang, "Relevance feedback: A power tool for interactive content-based image retrieval", *IEEE Trans. on Circ. Syst. for Video Technology*, vol. 8, pp 644–655, 1998.
[11] Philippe Salembier, „Overview of the MPEG-7 standard and of future challenges for visual information analysis", *EURASIP Journal on Applied Signal Processing*, Volume 2002, Issue 4, pp. 343-353, 2002.
[12] T. Mitchell, *Machine Learning*, McGraw-Hill, NY, 1997.
[13] J. B. Kruskal, M. Wish, *Multidimensional Scaling*, Thousand Oaks, 1977.
[14] I. T. Jolliffe, *Principal Component Analysis*. 2nd edition, New-York: Springer-Verlag, 2002.
[15] Y. Wu, Q. Tian, and T. S. Huang, "Discriminant EM algorithm with application to image retrieval," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, South Carolina, June 13-15, 2000.
[16] H. Yu, J. Yang, "A direct LDA algorithm for high-dimensional data with application to face recognition", *Pattern Recognition*, vol. 34, pp. 2067-2070, 2001.
[17] X. Zhou, T. Huang, "Small sample learning during multimedia retrieving using biasmap", in *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, vol. 1, pp.11-17, 2001.
[18] D. Tao, X. Tang, X. Li, Y. Rui, "Direct kernel biased discriminant analysis: A new content-based image retrieval relevance feedback algorithm", *IEEE Trans. Multimedia*, vol. 8, no. 4, pp. 716-727, Aug. 2006.
[19] L. Zhang, F. Lin, B. Zhang, "Support vector machine learning for image retrieval", in *Proc. IEEE Int. Conf. Image Processing* (ICIP), pp. 721-724, 2001.
[20] G. Zajić, N. Kojić, N. Reljin, B. Reljin, "Experiment with reduced feature vector in CBIR system with relevance feedback", in *5th IET Visual Information Engineering 2008 Conference* (ViE 2008), Xi'an, China, July 29 – Aug. 01, 2008
[21] G. Zajić, N. Kojić, V. Radosavljević, M. Rudinac, S. Rudinac, N. Reljin, I. Reljin, B. Reljin, "Accelerating of image retrieval in CBIR system with relevance feedback", *EURASIP Journal of Advances in Signal Processing*, Vol. 2007, 1-13, 2007.
[22] Corel Gallery Magic 65000 (1999), www.corel.com
[23] S. Rudinac, G. Zajić, M. Ušćumlić, M. Rudinac, B. Reljin, „Comparison of CBIR systems with different number of feature vector components", in *Proc. Workshop SMAP 2007*, London, UK, Dec. 17-18, 2007.