# FEATURE SELECTION-BASED RELEVANCE FEEDBACK IN CONTENT-BASED RETRIEVAL SYSTEMS

*D. Sotiropoulos and G.A. Tsihrintzis*

Department of Informatics
University of Piraeus
Piraeus 185 34
GREECE

**ABSTRACT**

*We address the problem of modeling the* subjective *perception of similarity between two images that have been extracted from an image database with use of* objective *features. We propose the importation of user models in Content-Based Image Retrieval (CBIR) systems, which embody the ability of evolving and using different image similarity measures for different users. Specifically, a user-supplied relevance feedback procedure allows the system to determine which subset of a set of objective features approximates more efficiently the subjective image similarity of a specific user. Our implementation of the proposed system verifies our hypothesis and exhibits significant improvement in perceived image similarity.*

## 1. INTRODUCTION

Modern CBIR systems attempt to retrieve images from a database according to their similarity (objective or subjective, as analysed below) to the user's query [1]. The most common practice followed in several CBIR systems uses similarity measures that combine a fixed set of *objective* features (derived from the color, texture and shape content of an image) and produces a similarity value for two images [2]. Since the similarity value is produced from objective features, it applies universally (to all users) and, therefore, the procedure is invariant under differences in image similarity *perception* between different users. However, the proper use of similarity perception information may improve the accuracy and speed of CBIR systems, as recent studies have shown [3]. Specifically in [5], information about a specific user's image similarity perception is supplied to the CBIR system through an iterative procedure in which first the CBIR system retrieves images on the basis of objective features and then the user ranks the retrieved images through a *relevance feedback* procedure. The user-supplied ranking is fed into a learning algorithm which allows the CBIR system to retrieve images for the specific user with higher efficiency, i.e., the system returns a smaller number of images which are perceived by the specific user as more *similar*.

In this paper, we propose the importation of user models in CBIR systems, which embody the ability of evolving and using different image similarity measures for different users. Contrary to previous works, our approach investigates certain subsets in the objective feature set which are able to approximate more efficiently the subjective image similarity perception of a specific user. In particular, we utilize relevance feedback from the user in an incremental learning process in order to specify that feature subset and the corresponding similarity measure which exhibit the maximum possible accordance with the user's image similarity perception. Our implementation of the proposed system verifies our hypothesis and exhibits significant improvement in perceived image similarity.

Specifically, the paper is organized as follows: Section 2 describes in some detail the set of objective features that our system utilizes. Section 3 describes the overall architecture of our proposed system, while experimental results for its use are presented in Section 4. Finally, conclusions and suggestions for related future work are given in Section 5.

## 2. OBJECTIVE FEATURE SET

A similarity measure provides a quantitative representation of the degree of similarity between two images on the basis of a set of objective features that are assumed to sufficiently reflect the information contained in an image. In this paper, the feature set selected for image similarity classification corresponds to both image color content and the presence of certain textural characteristics like levels, edges, spots and ripples.

The features associated with color content refer to the different intensity level distributions of the red, green and blue color in a specific image (stored in a $m \times n \times 3$ matrix $M$ ), as they appear in the three corresponding constituent images:

$$RM[i,j] = M[i,j,1],\ 1 \le i \le m,\ 1 \le j \le n \tag{1}$$

$$GM[i,j] = M[i,j,2],\ 1 \le i \le m,\ 1 \le j \le n \tag{2}$$

$$BM[i,j] = M[i,j,3],\ 1 \le i \le m,\ 1 \le j \le n\ . \tag{3}$$

For the representation of the textural characteristics of an image, we used the 16 two-dimensional texture detection masks of Laws [1] that act convolutionary on the intensity (gray) level representation (stored in n $m \times n$ matrix *GrM*) of the original image:

$$GrM[i,j] = \frac{RM[i,j] + GM[i,j] + BM[i,j]}{3}\ . \tag{4}$$

The 16 convolutionary masks are computed as all the possible outer products of pairs of the following vectors:

$$L_5(level) = [1,4,6,4,1] \tag{5}$$

$$G_5(edge) = [-1,-2,0,2,1] \tag{6}$$

$$S_5(spot) = [-1,0,2,0,1] \tag{7}$$

$$R_5(ripple) = [1,-4,6,-4,1]\ . \tag{8}$$

In this way, we produce the following set of 16 $5 \times 5$ masks (matrix products) *M1, M2, …, M16*:

$$\begin{array}{lll} M1 = L_5^T L_5 & M7 = G_5^T S_5 & M13 = R_5^T L_5 \\ M2 = L_5^T G_5 & M8 = G_5^T R_5 & M14 = R_5^T G_5 \\ M3 = L_5^T S_5 & M9 = S_5^T L_5 & M15 = R_5^T S_5 \\ M4 = L_5^T R_5 & M10 = S_5^T G_5 & M16 = R_5^T R_5 \\ M5 = G_5^T L_5 & M11 = S_5^T S_5 & \\ M6 = G_5^T G_5 & M12 = S_5^T R_5 & \end{array}$$

To produce the 16 corresponding $m \times n$ feature matrices *LM1, LM2,…, LM16,* we compute the two-dimensional convolution of the matrix *GrM* with each of the previously computed masks, i.e.

$$LM_k = conv2(GrM, M_k),\ 1 \le k \le 16 \tag{9}$$

This procedure generates 19 feature matrices *RM, GM, BM, LM1,…, LM16,* each of the same dimensions as the original image matrix. To reduce the amount of information stored in these 19 feature matrices, we replace each matrix with a vector of 6 representative values, which are extracted as the *mean* $(\mu)$, *variance* $(\sigma^2)$, *skewness* $(\mu_3)$, *kurtosis* $(\mu_4)$, *energy* $(\mathrm{E})$ and *entropy* $(\mathrm{H})$. Specifically:

$$F = \{F_1, F_2, F_3, F_4, \ldots, F_{19}\} \tag{10}$$

where

$$F_1 = [\mu(RM), \sigma^2(RM), \mu_3(RM), E(RM), H(RM)]$$
$$F_2 = [\mu(GM), \sigma^2(GM), \mu_3(GM), E(GM), H(GM)]$$
$$F_3 = [\mu(BM), \sigma^2(BM), \mu_3(BM), E(BM), H(BM)]$$

and for $1 \le k \le 16$

$$F_{k+3} = [\mu(LM_k), \sigma^2(LM_k), \mu_3(LM_k), E(LM_k), H(LM_k)]$$

To generate the six representative statistical values of a $m \times n$ real matrix $M$ , we compute the value and the frequency of occurrence for each one of its elements. We have:

$$\mu(M) = \sum_{i=1}^{l} v_i \times p_i \tag{11}$$

$$\sigma^2(M) = \sum_{i=1}^{l} (v_i - \mu)^2 \times p_i \tag{12}$$

$$\mu_3(M) = \sigma^{-3} \sum_{i=1}^{l} (v_i - \mu)^3 \times p_i \tag{13}$$

$$\mu_4(M) = \sigma^{-4} \sum_{i=1}^{l} (v_i - \mu)^4 \times p_i \tag{14}$$

$$E(M) = \sum_{i=1}^{l} p_i^{\ 2} \tag{15}$$

$$H(M) = -\sum_{i=1}^{l} p_i \times \log_2 p_i\ , \tag{16}$$

where

$$p_i = \frac{h_i}{m \times n},\ 1 \le i \le l \tag{17}$$

are the relative frequencies of occurrence of the $l$ different values $v_1, v_2, ..., v_l$ appearing in a matrix M and $h_1, h_2, ..., h_l$ are the corresponding absolute frequencies [6].

## 3. PROPOSED SYSTEM

In this section, we investigate the possibility of utilization of only certain feature subsets, rather than the full original feature set, namely those subsets that approximate more efficiently the subjective image similarity perception of a specific user. Our proposition is based on the hypothesis that each individual valuates differently the information contained in an image. This means that each of the 19 computed feature vectors is assigned a different degree of importance as different individuals assess image similarity via different features and potentially there might exist certain features that are unidentifiable by certain users. An extreme example of the latter may be color for color-blind users. Thus, we propose the importation of user models in CBIR systems, which embody the ability to evolve and use different similarity measures for different users in accordance with the different manners of valuating and combining image features.

Specifically, the proposed model is as shown in Fig. 1, where the target image corresponds to an existing image in the system database. The feature extractor conducts extracts the set of values for the complete feature set $F$. Afterwards, a predefined number of $M$ subsets from the original feature vectors set, $C_1, C_2, ..., C_M$, are assessed for their ability to approximate the subjective image similarity perception of a specific user. These subsets of feature vectors are fed into the corresponding neural networks to force them to realize $M$ different similarity measures.

Each neural network retrieves the most similar image according to the similarity measure which it realizes. In general, it is possible that all the retrieved images are different. The user valuates the set of the retrieved images and ranks the degree of similarity between the retrieved images and the target image according to his/her own perception. This information is subsequently used by the system in order to adjust the neural networks' parameters. This latter parameter refinement is conducted according to the second training procedure described in the previous paragraph and involves the adaptation of the parameter set of the entire network. This procedure is repeated for a preset number of times during which the network performance is recorded. In the end, we determine the neural network and the corresponding feature subset that exhibited the most effective performance in modeling the behaviour of the specific user.

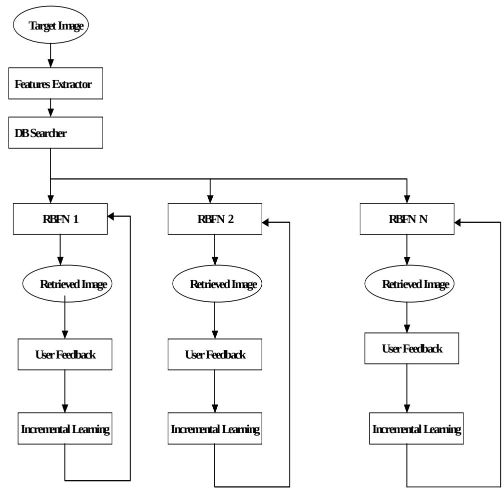


Figure 1

## 4. EXPERIMENTAL RESULTS

In this section, we describe one of the experiments that we have conducted in order to test our proposed system. More specifically, we have tested a set of 10 different neural networks, each corresponding to a different subset of the original feature vectors set $F$. The neural networks' parameters were refined by an incremental learning procedure which was completed in 10 stages (iterations). In each stage, the internal parameters of the neural networks were modified according to the user feedback on the basis of the absolute difference between the user supplied similarity values and the similarity values returned from the neural networks. After the end of each training stage, the neural networks were ranked according to the absolute difference between the desired and the estimated similarity value so that a lower ranking represented a better performance. At the end of the incremental learning procedure, we computed the mean ranking of each neural network.

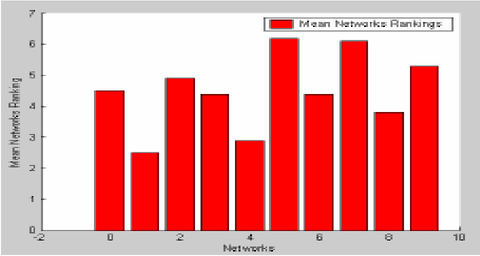

Figure 2

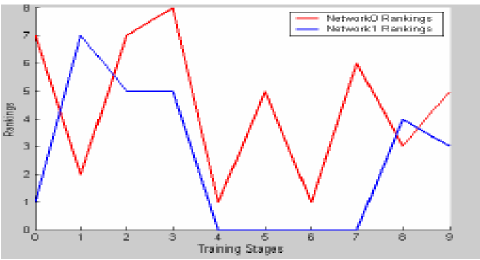

Figure 3

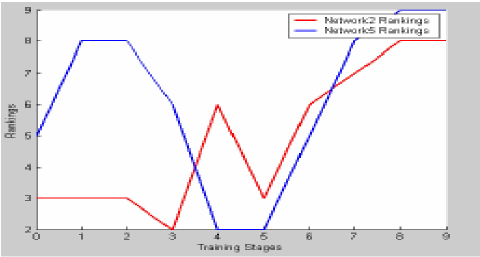

Figure 4

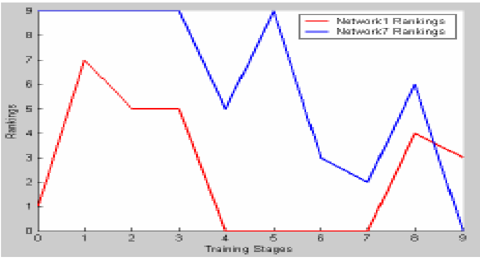

Figure 5

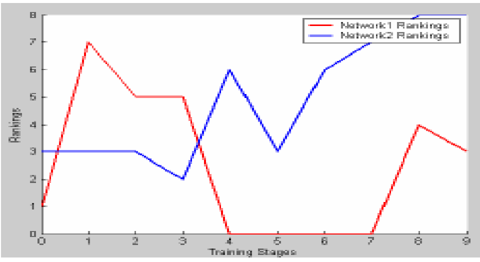

Figure 6

In Figs. 2-6, we see that the neural networks indexed by the values 1 and 4 had the best overall performance among all neural networks. Consequently, the corresponding feature subsets appear to be more efficient in modelling that specific user's similarity perception and can be considered as more appropriate similarity measures for the specific user.

## 5. CONCLUSIONS AND FUTURE WORK

In this paper, we propose the importation of user models in Content-Based Image Retrieval (CBIR) systems, which embody the ability of evolving and using different image similarity measures for different users. Specifically, a user-supplied relevance feedback procedure allows the system to determine which subset of a set of objective features approximates more efficiently the subjective image similarity of a specific user. Our implementation of the proposed system verifies our hypothesis and exhibits significant improvement in perceived image similarity. Of course, this fact provides very strong evidence that confirms our initial hypothesis that relates *objective* image features to *subjective* image similarity perception.

In the future, we will improve and refine our system and enhance it with a greater number of objective image features. We are also testing a similar approach to the content-based retrieval of music files. The results of this and related work will be announced at a future occasion.

## 6. REFERENCES

[1] L.G. Shapiro and G.C. Stockman, *Computer Vision*, Prentice Hall, 2001.

[2] Ingemar J. Cox, L. Miller, Stephen M. Omohundro, Peter N. Yianilos , "PicHunter : Bayesian Relevance Feedback for Image Retrieval," *Int. Conf. on Pattern Recognition*, Vienna, Austria, 1996.

[3] Ingemar J. Cox, Matt L. Miller, Thomas P. Minka, Thomas V. Papathomas, and Peter N. Yianilos, "The Bayesian Image Retrieval System, PicHunter : Theory, Implementation and Psychophysical Experiments," *IEEE Transactions on Image Processing*, 2000.

[4] S.Haykin, *Neural Networks*, Second Edition, Prentice Hall, New Jersey, 1999.

[5] Hyoung K. Lee and Suk I. Yoo, "A neural network based image retrieval using nonlinear combination of heterogeneous features", In *Proceedings of the 2000 Congress on Evolutionary Computation CEC00*, 2000.

[6] Andrej Materka and Michal Strezelecki, "Texture Analysis Methods – A Review", Technical Univercity of Lodz. Institute of Electrononics, COST B11 report, Brussels 1998