

# Interactive Image Retrieval in a Fuzzy Framework

Malay K. Kundu, Minakshi Banerjee, and Priyank Bagrecha

Machine Intelligence Unit, Center for Soft Computing Research  
Indian Statistical Institute  
203, B. T. Road, Kolkata 700 108, India  
{minakshi\_r,malay,b.priyank.t}@isical.ac.in

**Abstract.** In this paper, an interactive image retrieval scheme using MPEG-7 visual descriptors is proposed. The performance of image retrieval systems is still limited due to semantic gap, which is created from the discrepancies between the computed low-level features (color, texture, shape, etc.) and user's conception of an image. As a result, more interest has been created towards development of efficient learning mechanism other than designing sophisticated low-level feature extraction algorithms. A simple relevance feedback mechanism is proposed, that learns user's interest and updates feature weights based on a fuzzy feature evaluation measure. This has an advantage of handling comparatively small number of samples over those using standard classifiers involving large number of training samples and having more complexity. Extensive experiments have been performed to test to what extent the performance of an image retrieval system can be enhanced further using MPEG-7 standard visual features at minimum cost.

Efficient image retrieval techniques from a large database have become an active field of research with the advent of the World-Wide Web. Content-Based Image Retrieval(CBIR) techniques are becoming more important with this basic requirement [1]. It is aimed at retrieving relevant images from an image database by measuring similarity between the automatically derived low-level features (color, texture, shape, etc. ) of the query image and the images stored in the database. Although different image characterization methods [2],[3] have been explored to represent images with basic low-level features but their usefulness is limited by the gap, between low-level features and high-level concepts known as semantic gap. Performance of CBIR is still far from user's expectations owing to semantic gap.

To facilitate effective use of audio, visual(color, texture, shape, etc.) and motion descriptors, ISO/IEC has launched MPEG-7 to address multimedia retrieval. It provides a collection of specific, standard descriptors [4] used as a benchmark for evaluation of new schemes for image retrieval [5]. Among various state-of-the-art techniques in narrowing down the semantic gap, relevance feedback mechanism has been identified as an essential tool to provide significant performance boost in CBIR systems [6], [7], [8], through continuous learning and interaction with

end-users. The system provides initial retrieval results through query-by-example, sketch, etc., based on which the user judges the retrieved results as to whether and to what degree, they are relevant (positive examples)/irrelevant (negative examples) to the query. Machine learning algorithm is then applied to learn the user's feedback and improve the results iteratively till user's satisfaction.

Most of the relevance feedback methods employ two approaches [6] namely, query vector moving technique and feature re-weighting technique to improve retrieval results. In the first approach, the query is reformulated by moving the vector towards positive examples and away from the negative examples, assuming that all positive examples will cluster in the feature space. Feature re-weighting method is used to enhance the importance of those components of a feature vector, that help in retrieving relevant images, while reducing the importance of the features that does not help. However in such cases, the selection of positive and negative examples, from a small number of samples having large number of features, still remain as a problem.

Relevance feedback techniques in CBIR, have mostly utilized information of the relevant images but have not made use of the information from irrelevant images. Zin et al., [9] have proposed a feature re-weighting technique by using both the relevant and the irrelevant information, to obtain more effective results.

Recently, relevance feedback has been considered as a learning and classification process, using classifiers like Bayesian classifiers [10], neural network [11], etc. However trained classifiers become less effective when the training samples are insufficient in number. To overcome such problems, active learning methods have been used in [12].

Our contributions in this paper deal with studying the performance of a CBIR system using MPEG-7 visual descriptors. We present a simple relevance feedback method which uses the concept of combining information from both relevant and irrelevant images, from a small number of retrieved images. The motivation is to test how far it can enhance the performance of a CBIR system already using MPEG-7 visual features by using a simple feedback mechanism requiring less computational time and managing small samples opposed to different classifiers [11], [10] used for relevance feedback. We also present a comparison of the system against a CBIR system which uses moments of significant points (corners and around) [13] of a color image involving 6 feature components only.

## 1 Evaluating Importance of a Feature from Relevance Rating

In conventional CBIR approaches an image  $I$  is usually represented by a set of features,  $F = \{f_q\}_{q=1}^N$ , where  $f_q$  is the  $q$ th feature component in the  $N$  dimensional feature space. The commonly used decision function for measuring similarity between the query image  $I_{qr}$  and other images  $I$ , is represented as,

$$D_{is}(I, I_{qr}) = \sum_{q=1}^N w_q \|f_q(I) - f_q(I_{qr})\| \quad (1)$$

where  $\|f_q(I) - f_q(I_{qr})\|$  is the Euclidean distance between the  $q$ th component and  $w_q$  is the weight assigned to the  $q$ th feature component. The weights should be adjusted such that, the features have small variation over the relevant images and large variation over the irrelevant images. Let  $k$  similar images  $I_s = \{I_1, I_2, \dots, I_k\}$  where,  $I_k \in I_s$ , are returned to the user. Let  $I_r$  be the set of relevant images and  $I_{ir}$  be the set of irrelevant images as marked by the user.  $I_r = \{I_j \mid I_j \text{ relevant, for } I_j \in I_s\}$  and  $I_{ir} = \{I_j \mid I_j \text{ irrelevant, for } I_j \in I_s\}$ . The information from  $I_r$  and  $I_{ir}$  are combined to compute the relative importance of the individual features, from fuzzy feature evaluation index (FEI) [14] in pattern classification problems.

The (FEI) is defined from interclass and intraclass ambiguities as follows : Let  $C_1, C_2, \dots, C_j \dots C_m$  be the  $m$  pattern classes in an  $N$  dimensional  $(f_1, f_2, f_q, \dots, f_N)$  feature space where class  $C_j$  contains,  $n_j$  number of samples. It can be shown that entropy of a fuzzy set [15] gives a measure of 'intrasets ambiguity' along the  $q$ th co-ordinate axis in  $C_j$  is computed as,

$$H(A) = \left(\frac{1}{n_j \ln 2}\right) \sum_i S_n(\mu(f_{iqj})); i = 1, 2 \dots n_j \tag{2}$$

where the Shannon's function,  $S_n(\mu(f_{iqj})) = -\mu(f_{iqj}) \ln \mu(f_{iqj}) - \{1 - \mu(f_{iqj})\} \ln \{1 - \mu(f_{iqj})\}$  Entropy is dependent on the absolute values of membership ( $\mu$ )  $H_{min} = 0$  for  $\mu = 0$  or  $1$ ,  $H_{max} = 1$  for  $\mu = 0.5$  Entropy (H) of  $C_j$  along  $q$ th component can be computed using a standard S-type membership function as shown in (3).

$$\begin{aligned} S(x; a, b, c) &= 0 && x \leq a \\ &= 2 \times \left\{ \frac{(x-a)}{(c-a)} \right\}^2 && a \leq x \leq b \\ &= 1 - 2 \times \left\{ \frac{(x-c)}{(c-a)} \right\}^2 && b \leq x \leq c \\ &= 1 && x \geq c \end{aligned} \tag{3}$$

where,

$$\begin{aligned} b &= (f_{qj})av \\ c &= b + \max\{|(f_{qj})av - (f_{qj})max|, |(f_{qj})av - (f_{qj})min|\} \\ a &= 2b - c \end{aligned}$$

$(f_{qj})av$ ,  $(f_{qj})max$ ,  $(f_{qj})min$  denote the mean, maximum and minimum values respectively computed along the  $q$ th co-ordinate axis over all the  $n_j$  samples in  $c_j$ . Since  $\mu(b) = \mu(f_{qj})av = 0.5$ , the values of  $H$  are 1.0 at  $b = (f_{qj})av$  and would tend to zero when moved away from  $b$  towards either  $c$  or  $a$  of the S function. Selecting  $b = (f_{qj})av$  indicates that, the cross over point is near the query feature component. Higher value of  $H$ , indicates more number of samples having  $\mu(f)$  equal to 0.5. with a tendency of the samples to cluster around the mean value, resulting in less internal scatter within the class.

After combining the classes  $C_j$  and  $C_k$  the mean, maximum and minimum values  $(f_{qkj})av$ ,  $(f_{qkj})max$ ,  $(f_{qkj})min$  respectively of  $q$ th dimension over the samples  $(n_j + n_k)$  are computed similarly, where  $n_k$  are the samples in class  $C_k$ . The criteria of a good feature is that, it should be nearly invariant within class,

while emphasizing differences between patterns of different classes [14]. The value of  $H$  would therefore decrease, after combining  $C_j$  and  $C_k$  as the goodness of the  $q$ th feature in discriminating pattern classes  $C_j$  and  $C_k$  increases. The measure denoted as  $H_{qjk}$  is called "inter-set ambiguity" along  $q^{th}$  dimension between classes  $C_j$  and  $C_k$ . Considering the two types of ambiguities, the proposed Feature evaluation index (FEI) for the  $q$ th feature is,

$$(FEI_q) = \frac{H_{qjk}}{H_{qj} + H_{qk}} \quad (4)$$

Lower value of  $FEI_q$ , indicates better quality of importance of the  $q$ th feature in recognizing and discriminating different classes. The precision of retrieval can be improved with these values.

In the proposed algorithm, the number of classes are two. The relevant/positive images constitute the (intra-class) and the irrelevant/negative images constitute the (inter-class) image features. To evaluate the importance of the  $q$ th feature, the  $q$ th component of the retrieved images is considered. i.e.,  $I^{(q)} = \{I_1^{(q)}, I_2^{(q)}, I_3^{(q)}, \dots, I_k^{(q)}\}$

$H_{qj}$  is computed from  $I_r^{(q)} = \{I_{r1}^{(q)}, I_{r2}^{(q)}, I_{r3}^{(q)}, \dots, I_{rk}^{(q)}\}$ . Similarly  $H_{qk}$  is computed from the set of images,  $I_{ir}^{(q)} = \{I_{ir1}^{(q)}, I_{ir2}^{(q)}, I_{ir3}^{(q)}, \dots, I_{irk}^{(q)}\}$ .  $H_{qkj}$  is computed combining both the sets. Images are ranked according to Euclidean distance. The user marks the relevant and irrelevant set from 20 returned images, for automatic evaluation of (FEI). The weight  $w_q$  is a function of the evaluated ( $FEI_q$ ) as shown in eqn. 5

$$w_q = F_q(FEI_q) \quad (5)$$

Now the problem is, what would be weight updation function for the automatically evaluated important features, when all feature elements are merged into a big overall feature vector. Owing to such situations, different feature updation functions like,  $[w_q = FEI_q^2, \frac{1}{FEI_q^2}, exp(FEI_q)]$  could be tested. Whichever is the best strategy can be decided by selecting a better performing  $w_q$  for majority of the queries in the database. For the query feature vector  $F$ , the individual components of relevant images are expected to vary within a smaller range say ( $\epsilon$ ) and may be represented as.

$$I_r = \{I_j \in I_s : \frac{\delta f_q}{|F|} \leq \epsilon\} \quad (6)$$

In the first pass, all features are considered to be equally important. Hence  $w_1 = w_2, \dots = w_q = 1$ . The feature spaces of the relevant images are therefore altered in a similar fashion after updating the components with  $w_q$ . As a result, the ranks of the relevant images are not affected much. For irrelevant images, one feature component may be very close to the query, whereas other feature component may be far away from the query feature. But the magnitude of the similarity vector may be close to the relevant ones. These images may be characterized as,

$$I_{ir} = \{I_j \in I_s : \frac{\delta f_{ql}}{|F|} \gg \epsilon \text{ and } \frac{\delta f_{qm}}{|F|} \ll \epsilon\} \quad (7)$$

where,  $l \in 1, 2, \dots, N$  and  $m=l^c$ . For example, multiplying by  $FEI_q^2$  will in effect decrease the component feature separation such that, the relevant component i.e., the term  $\frac{\delta f_{qm}}{|F|} \ll \epsilon$  of (6) will be more closer to the query than the irrelevant component and due to combined distance within the similarity metric, the relevant images may come up. Multiplying by  $\frac{1}{FEI_q^2}$  increases the feature separation between the irrelevant component, such that due to the combined effect the irrelevant image may be pulled down.

## 2 Experimentation

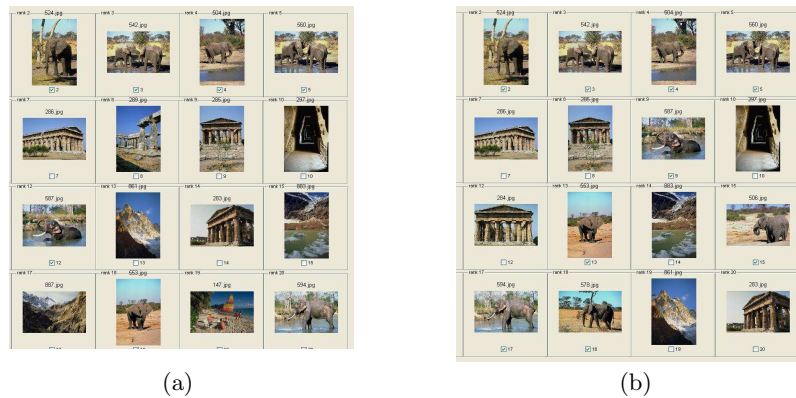
To prove the effectiveness of the proposed relevance feedback mechanism, extensive experiments have been performed on MPEG-7 standard visual features downloaded from <http://standards.iso.org/ittf/licence.html>, upon two databases namely, (A) SIMPLIcity images consisting of 1000 images from 10 different categories (B) Corel 10000 miscellaneous database which is labeled into 79 semantic categories and downloaded from <http://wang.ist.psu.edu/IMAGE>. The updation formula  $w_q = FEI_q^2$  is used in each iteration as it generated better results in majority of the cases.

Among the different representation schemes used as the MPEG-7 core features, CSD(Color Structure Descriptor) and EHD (Edge Histogram Descriptor) [5] are chosen as color and texture features, to evaluate overall similarity between images. Shape could be more important if region based properties were extracted. *The color structure descriptor* histogram aims at identifying localized color distribution using a small sliding structuring window, which is constructed in the hue-min-max-difference (HMMD). CSD is defined by non linear quantization of the color space, and represented by a 184 bin histogram. A 184 bin of CSD can be mapped to lower number of bins (120, 64, or 32) by coarsely quantizing the color space and L2 (Euclidean) distance is used for similarity evaluation. In *Edge Histogram Descriptor* the original image is divided into 16 subimages, each subimage is divided into a fixed number of blocks. Each image block is partitioned into 2x2 block of pixels. The edge detector operators are then applied to 2X2 blocks, treating each block as a pixel and the average intensity as the corresponding block intensity value. A total of 80 bins, 3 bits/bin, are used for representing the edge histogram while L2 distance is used as the metric.

As representation schemes associated with the traditional color histograms, suffer from problems associated with high-dimensional indexing, the performance is also evaluated against a CBIR system, using color moments at significant spatial locations (corners and around) with a feature vector size of six components only [13].

MPEG-7 CSD feature extraction approximately takes 1-2 secs, whereas EHD takes approximately 500ms. Such a difference can be accounted to the use of sliding structuring window depending upon the image size in the case of CSD descriptors. The proposed relevance feedback has been tested independently using the features (CSD, EHD and moment based [13]) and retrieval performance measured in terms of precision as shown in eqn. (8).

As low-level features are not always powerful in representing the semantic concepts, the images similar in semantic contents are selected as positive examples among the first twenty retrieved set, in each round of feedback iteration and the remaining are negative examples for updating the weight parameters and revising the features. In most of the cases, CSD has been able to produce better results than EHD in terms of color distribution. EHD produced better results where spatial distribution of edges, is more important. However, with CSD descriptors the retrieval accuracy is limited when semantic significance is more important than color distribution only. One such example may be the case of an elephant where the improvement in precision is from 50 % - 75% at third iteration as shown in Figs. 1 -2. The weight updating approach is tested in terms of Average precision (with n=20 images) which is the average value obtained using eqn. 8 considering all randomly selected queries of the databases. The average precision obtained from the set of same query after different iterations are shown in Fig. 3(a). As seen from Fig. 3(a) that EHD and moment based methods generated almost similar results whereas CSD features performed better in terms of average retrieval precision. The experiments have been implemented on a Dell (T7400, 4GB, RAM) machine using MATLAB R2008a package. The cputime taken for each iteration is approximately 500ms. for database (A) and 3 secs for database (B).

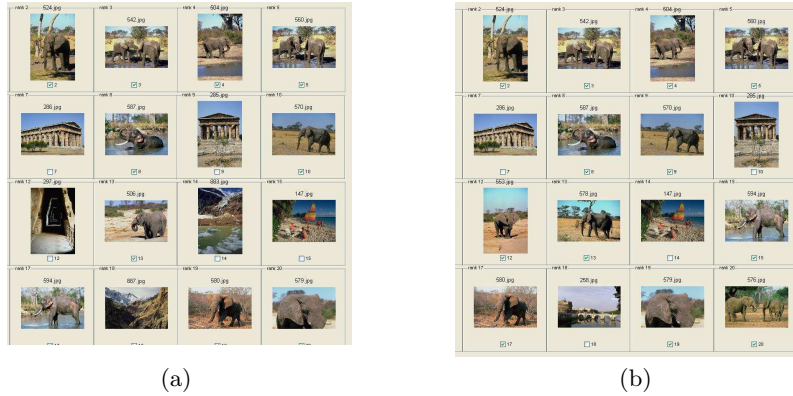


**Fig. 1.** The proposed relevance feedback scheme on MPEG-7, CSD visual descriptors, (a)First pass (b) iteration1

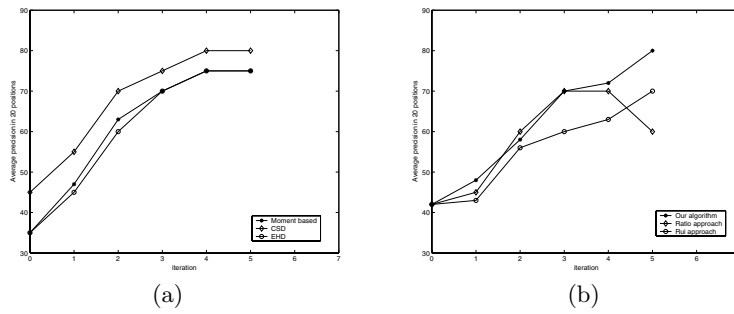
The results for different queries are defined in terms of Precision as,

$$P(n) = \frac{\text{target images within } n \text{ positions}}{n} \times 100 \quad (8)$$

The average precision as obtained from other two methods namely, (a) Ratio approach [16] in which the relative ratios of standard deviations of the component features of positive and negative examples are considered (b) Rui's method [6] in



**Fig. 2.** The proposed relevance feedback scheme on MPEG-7, CSD visual descriptors, (a) iteration2 (b) iteration3



**Fig. 3.** Comparative studies (a) using different feature extraction schemes (b) using different weight updation schemes

which a feature’s weight is assigned in inverse proportion to the features variance across the images marked relevant is shown in Fig. 3(b). Experimental results show that, the proposed relevance feedback mechanism enhances the results using the MPEG-7 feature descriptors better than Ratio and Rui’s method in most of the cases. As the computed FEI is the measure of heterogeneity in the relevant and irrelevant set of the component feature space it is able to improve the results consistently, with the *S*-type function in the interval (0, +1).

### 3 Conclusion

Experimental results show that the proposed image retrieval system based on MPEG-7 features is able to improve the retrieval performance more than 20% in most of the cases, within two iterations of relevance feedback. Since, MPEG-7 Visual descriptors, describe features for effective image or video retrieval, effectiveness of the proposed feedback mechanism could be tested for video retrieval as future scope of research.

## References

1. Smeulders, A.W.M., Worring, M., Santini, S., Gupta, A., Jain, R.: Content-based image retrieval at the end of the early years. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 22(12), 1349–1380 (2000)
2. Cheng, K.O., Law, N.F., Siu, W.C.: Multiscale directional filter bank with applications to structured and random texture retrieval. *Pattern Recognition* 40(4), 610–621 (2007)
3. Martínez, J.C., Medina, J.M., Barranco, C.D., Perales, G., Hidalgo, J.M.S.: Retrieving images in fuzzy object-relational databases using dominant color descriptors. *Fuzzy Sets and Systems* 158(3), 312–324 (2007)
4. Manjunath, B.S., Salembier, P., Sikora, T.: Introduction to MPEG-7: Multimedia Content description Interface. John Wiley and Sons, Inc., Chichester (2002)
5. Manjunath, B.S., Ohm, J.R., Vasudevan, V.V., Yamada, A.: Color and texture descriptors. *IEEE Transactions on Circuits and Systems for Video Technology* 11(6), 703–715 (2001)
6. Rui, Y., Huang, T.S., Mehrotra, S.: Relevance feedback: a power tool for interactive content-based image retrieval. *IEEE transactions on Circuits and Systems for Video technology* 8(5), 644–655 (1998)
7. Yin, P.Y., Bhanu, B., Chang, K.C., Dong, A.: Integrating relevance feedback techniques for image retrieval using reinforcement learning. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 27(10), 1536–1551 (2005)
8. Lim, J.H., Jin, J.S.: Combining intra-image and inter-class semantics for consumer image retrieval. *Pattern Recognition* 38(6), 847–864 (2005)
9. Jin, Z., King, I., Li, X.Q.: Content-based retrieval by relevance feedback. In: Laurini, R. (ed.) *VISUAL 2000*. LNCS, vol. 1929, pp. 521–529. Springer, Heidelberg (2000)
10. Ves, E.D., Domingo, J., Ayala, G., Zuccarello, P.: A novel bayesian framework for relevance feedback in image content-based retrieval systems. *Pattern Recognition* 39(9), 1622–1632 (2006)
11. Qian, F., Zhang, B., Lin, F.: Constructive learning algorithm-based rbf network for relevance feedback in image retrieval. In: Bakker, E.M., Lew, M., Huang, T.S., Sebe, N., Zhou, X.S. (eds.) *CIVR 2003*. LNCS, vol. 2728, pp. 352–361. Springer, Heidelberg (2003)
12. He, X., King, O., Ma, W., Li, M., Zhang, H.J.: Learning a semantic space from user's relevance feedback for image retrieval. *IEEE transactions on Circuits and Systems for Video technology* 2003 13(1) (2003)
13. Banerjee, M., Kundu, M.K., Das, P.K.: Image Retrieval with Visually Prominent Features using Fuzzy Set Theoretic Evaluation. In: *IEE International Conference on Visual Information Engineering VIE 2006*, India, pp. 298–303 (2006)
14. Pal, S.K., Chakraborty, B.: Intra-class and inter-class ambiguities (fuzziness) in feature evaluation. *Pattern Recognition Letters* 2, 275–279 (1984)
15. Pal, S.K., Majumder, D.D.: *Fuzzy mathematical Approach to Pattern Recognition*. Willey Eastern Limited, New York (1985)
16. King, I., Jin, Z.: Integrated probability function and its application to content-based image retrieval by relevance feedback. *Pattern Recognition* 36, 2177–2186 (2003)