

# Image Retrieval using Fuzzy Relevance Feedback and Validation with MPEG-7 Content Descriptors

M. Banerjee and M. K. Kundu

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

**Abstract.** Content-Based Image retrieval has emerged as one of the most active research directions in the past few years. In CBIR, selection of desired images from a collection is made by measuring similarities between the extracted features. It is hard to determine the suitable weighting factors of various features for optimal retrieval when multiple features are used. In this paper, we propose a relevance feedback framework, which evaluates the features, from fuzzy entropy based feature evaluation index (FEI) for optimal retrieval by considering both the relevant as well as irrelevant set of the retrieved images marked by the users. The results obtained using our algorithm have been compared with the agreed upon standards for visual content descriptors of MPEG-7 core experiments.

**Index Terms** Content-Based image retrieval, fuzzy feature evaluation index, invariant moments, MPEG-7 feature descriptors.

## 1 Introduction

Digital images are widely used in many application fields such as biomedicine, education, commerce crime prevention, world Wide Web searching etc. This necessitates finding of relevant images from a database, by measuring similarities between the visual contents (color, texture, shape etc.) of the query images and those stored in the database and commonly known as Content-based image retrieval (CBIR) [1],[2], [3], [4].

Due to the vast activities in CBIR research over the last few years, Moving Pictures Expert Group (MPEG) has started the standardization activity for "Multimedia Content description Interface" in MPEG-7 [5], to provide standardization of feature descriptions for audiovisual data.

---

\* *M. Banerjee is grateful to the Department of Science and Technology, New Delhi, India, for supporting the work under grant (SR/WOS-A/ET-111/2003)*

In a conventional CBIR, an image is usually represented by a set of features, where the feature vector is a point in the multidimensional feature space. However, images with high feature similarities may differ in terms of semantics. The discrepancy between the low level features like color, texture shape etc. and the high-level semantic concepts (such as sunset, flowers, outdoor scene etc.) is known as "semantic gap". To bridge this gap, users feedback may be used in an interactive manner which is popularly known as "relevance feedback" [6], [7], [8], [9]. The user rates the relevance of the retrieved images, from which the system dynamically learns the user's judgment to gradually generate better results. Owing to these facts, derivation and selection of optimal set of features, which can effectively model human perception subjectivity via relevance feedback, still remain a challenging issue.

Majority of the relevance feedback methods employ two approaches [10] 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 / relevant 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 remains 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., [11] have proposed a feature re-weighting technique by using both the relevant and the irrelevant information, to obtain more effective results.

Recently, relevance feedback is considered as a learning and classification process, using classifiers like Bayesian classifiers [12], neural network [13]. 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 [14].

A fuzzy entropy based feature evaluation mechanism is provided for relevance feedback, combining information from both relevant and irrelevant images. The effectiveness of the proposed method is compared, with the results of Schema(XM) which uses the feature descriptors of MPEG-7 core experiments. The remaining sections are organized as follows : The section 2 describes the mathematical formulations for relevance feedback frame work. The experimental results and conclusion are described in section 3

## **2 Estimation of relative importance of different features from relevance feedback**

Image retrieval using relevance feedback can be considered as a pattern classification problem. An effective relevance feedback system should be able to accumulate knowledge from small set of feedback images to form an optimal

query. Each type of visual feature tends to capture only one aspect of the image property. To evaluate the importance of individual components for a particular query  $I_{qr}$ , the information from relevant and irrelevant images may be combined as follows :

Let an image database  $S_d$  be composed of  $d$  distinct images,  $I=\{I_1, I_2, \dots, I_d\}$  where  $I \in S_d$ . The image  $I$  is 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 similarity function 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 variations 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. The information from relevant (intra-class) images  $I_r$  and irrelevant (inter-class) images  $I_{ir}$  are combined to compute fuzzy feature evaluation index (FEI) proposed by Pal et al., [15], [16] in pattern classification problems.

**Feature evaluation index :** The fuzzy measure (FEI) is defined from inter-class and intra-class ambiguities and explained 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 is shown that fuzzy entropy ( $H_{qj}$ ) [17] gives a measure of 'intra-set ambiguity' along the  $q$ th co-ordinate axis in  $C_j$ . The entropy of a fuzzy set, having  $n_j$  points in  $C_j$  is computed as,

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

where the Shannon's function,  $(S_n \mu(f_{iqj})) = -\mu(f_{iqj}) \ln \mu(f_{iqj}) - \{1 - \mu(f_{iqj})\} \ln \{1 - \mu(f_{iqj})\}$

For computing  $H$  of  $C_j$  along  $q$ th component, a standard S-type function shown in Fig. 1 is considered. At  $b$  (cross over point),  $S(b; a, b, c) = 0.5$ . Similarly at  $c$  (shoulder point)  $S(c; a, b, c) = 1.0$  and at  $a$  (feet point)  $S(a; a, b, c) = 0.0$ .

The parameters are set as follows.  $b = (f_{qj})av$ ,  
 $c = b + \max\{|(f_{qj})av - (f_{qj})\max|, |(f_{qj})av - (f_{qj})\min|\}$ ,  $a = 2b - c$  where  $(f_{qj})av$ ,  $(f_{qj})\max$ ,  $(f_{qj})\min$  denote the mean, maximum and minimum values respectively computed along the  $q$ th co-ordinate axis over all the  $n_j$  samples in  $c_j$ .

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, where  $\mu(b) = \mu(f_{qj})av = 0.5$ , eqn. (2). Selecting  $b = (f_{qj})av$  indicates that, the cross over point is near to the query feature component. Higher value of  $H$ , indicates more samples having  $\mu(f)$  equal to 0.5. i.e., 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 to evaluate "inter-set ambiguity"  $H_{qjk}$ , 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 having differences between patterns of different classes [15]. 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. Considering the two types of ambiguities, the proposed Feature evaluation index (FEI) for the  $q$ th feature is,

$$(FEI_q) = \frac{H_{qjk}}{H_{qj} + H_{qk}} \quad (3)$$

Lower the value of  $FEI_q$ , higher is the quality of importance of the  $q$ th feature in recognizing and discriminating different classes. The user marks the relevant and irrelevant set from 20 returned images. To evaluate the importance of the  $q$ th feature, the  $q$ th component of the retrieved images is considered. i.e.,  $I^{(q)} = \{I_1^{(q)}, I_2^{(q)}, I_3^{(q)}, \dots, I_k^{(q)}\}$ . Value of  $H_{qj}$  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,  $I_{ir}^{(q)} = \{I_{ir1}^{(q)}, I_{ir2}^{(q)}, I_{ir3}^{(q)}, \dots, I_{irk}^{(q)}\}$ .  $H_{qjk}$  is computed combining both the sets.

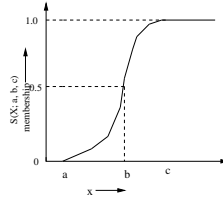


Fig. 1. S-type membership function

**Effect of sample size on evaluating importance :** To address the issue like, how to manage with limited number of returned images, we study the effect of image sample size on the FEI values. We combine two distinct categories of images. We mark category (1) as relevant and those from category (2) as irrelevant. We increase the sample size of the relevant and irrelevant images (double, triple, half, etc). We also test with other combinations. The FEI values computed from different combinations are shown in Table.1. As seen from the table, the order of the values obtained for  $(FEI)_q$  is same for a particular query category.

**Adjustment of weights :** The marking of relevant and irrelevant images is subjective. One may emphasize more on similarity between one feature (eg., color) than some other features ( eg., shape), when comparing two images. Accordingly the weights need to be adjusted. 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 .

$$I_r = \{I_j \in I_s : \frac{\delta f_q}{|F|} \leq \epsilon\} \quad (4)$$

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,

$$I_{ir} = \{I_j \in I_s : \frac{\delta f_{q1}}{|F|} \gg \epsilon \text{ and } \frac{\delta f_{q2}}{|F|} \ll \epsilon\} \quad (5)$$

After the features of the query and the stored images ( $S_d$ ) have been updated with the FEI values, the weighted components are expected to dominate over feature space such that, the rank of the irrelevant ones are pulled down. We have tested the results, from updating the weights with  $w_q = FEI_q^2, \frac{1}{FEI_q}^2$  and obtained better results from  $w_q = FEI_q^2$ , in majority of the cases. Intuitively this depends on how the combination of important features dominate over the others for a particular query. Importance of a features is decided from the decreasing order of (FEI) values. In order to further enhance the effect of the important features, we introduce another multiplying factor  $t_{qf}$  to make the component more dominating The weights of the individual features for successive iterations are expressed as follows ;  $w_{qf} = t_{qf} \times (FEI_{qf})^2$  where the value of  $t_{qf}$  are chosen as  $\{1.0, 0.1\}$ . to get better results.

**Features used for characterizing an image :** The features are computed from a set of moments invariant to rotation, translation and scaling. The moments  $m_{pq}$  of order p and q of a function  $f(x, y)$  for discrete images, are usually approximated as,

$$m_{pq} = \sum_x \sum_y x^p y^q f(x, y) \quad (6)$$

The centralized moments are expressed as,

$$\mu_{pq} = \sum_x \sum_y (x - \bar{x})(y - \bar{y})f(x, y)$$

$$\text{where } \bar{x} = \frac{m_{10}}{m_{00}}, \bar{y} = \frac{m_{01}}{m_{00}}$$

The normalized central moments are computed as,

$$\eta_{pq} = \frac{\mu_{pq}}{\mu_{00}^\gamma} \text{ where } \gamma = \frac{p+q}{2} + 1 \text{ for } p+q = 2, 3, \dots$$

A set of seven moments invariant to translation, rotation and scale can be computed from  $\eta_{pq}$ , from which  $\theta = \eta_{02} + \eta_{20}$  is considered. Let  $I(x, y)_R, I(x, y)_G, I(x, y)_B$  represent the R,G,B component planes of the image matrix  $I(x, y)$  and  $I_s(x, y)_R, I_s(x, y)_G, I_s(x, y)_B$  represent the component planes of the representative locations [18] shown in Fig.2. Six values of invariant moments ( $\theta$ ) are computed from the described component planes, to represent the feature vector.

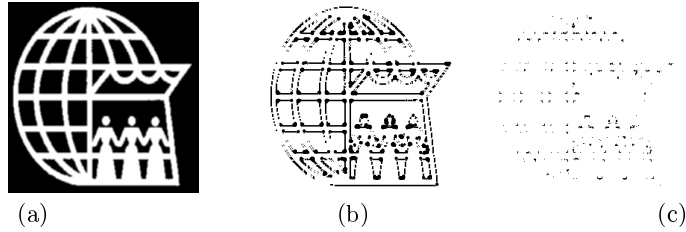
### 3 Experimentation

The performance of image retrieval system is tested upon SIMPLIcity images which consists of 1000 images from 10 different categories. The average precision value from each category (randomly chosen queries), after retrieving (10,20,40) images are shown in Table 2. Although the performance mostly depends upon the choice of features, it is observed that, the results almost converged after two iterations. A retrieval result is shown in Fig.3.

*Complexity* The time complexity  $T(n)$  for matching and sorting our results is represented as,  $T(n) = O(ND) + O(N \log N)$ . The complexity involved in computing Euclidean distance is  $O(ND)$  where  $D$  is the number of components within the feature vector. Sorting of  $(N)$  images with quick sort of  $O(N \log N)$ . As  $(D=6)$ . Hence  $D \ll \log N$ . Therefore  $T(n) \simeq O(N \log N)$ .

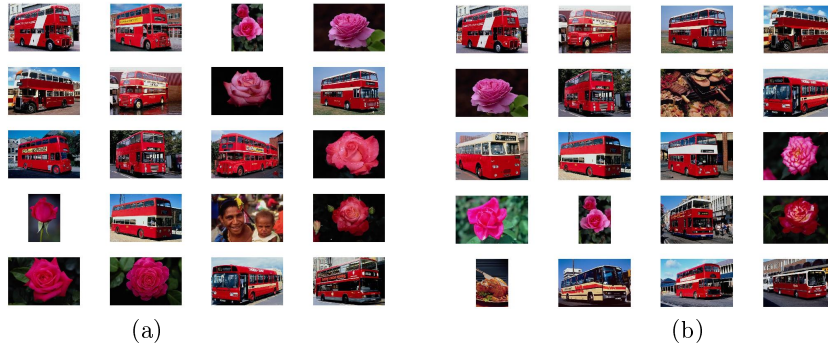
**Table 1.** feature evaluation index

Images(Intra, Inter)	$(FEI_1)^2$	$(FEI_2)^2$	$(FEI_3)^2$	$(FEI_4)^2$	$(FEI_5)^2$	$(FEI_6)^2$
(10,10)	0.207	0.115	0.199	0.351	0.351	0.351
(20,20)	0.207	0.115	0.199	0.351	0.351	0.351
(30,30)	0.220	0.121	0.195	0.350	0.350	0.350
(5,5)	0.309	0.151	0.247	0.349	0.349	0.349
(5,10)	0.289	0.164	0.265	0.306	0.306	0.306
(10,5)	0.433	0.213	0.401	0.534	0.534	0.534



**Fig. 2.** (a) Original image (b) Edge signature on which the high curvature region marked as (\*) (c) corner signature

**Comparison with MPEG-7** MPEG-7 [19] is an ISO/IEC standard, which provides a collection of specific, agreed upon standard (audio, visual) descriptors. MPEG-7 experimentation Model (XM) software [5] is the frame work of all the reference codes, and make use of MPEG-7 visual descriptors. These can serve as a test bed for evaluation and comparison of features in CBIR context. The features of XM commonly used as standard visual content descriptors for still images are listed in Table. 3.



**Fig. 3.** Retrieved results using set(A) (a) first set of candidates (b) After feature evaluation with (FEI), (iteration 1),  $t_i * FEI_{i=1}^N = \{1 \times 0.22, 1 \times 0.32, 1 \times 0.39, 0.1 \times 0.36, 0.1 \times 0.32, 0.1 \times 0.37\}$ . The top left image is the query image.

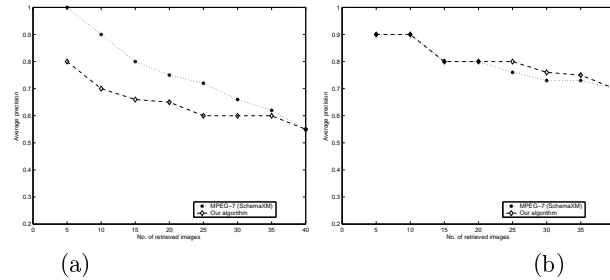
**Table 2.** Average precision % from our algorithm

Category	unweighted features			weighted features (iteration 1)		
	10 images	20 img.	40 img.	10 img.	20 img.	40 img.
Africa	68.20	60.25	55.03	70.50	61.00	58.00
Beach	70.00	55.23	50.60	71.48	56.23	52.58
Building	70.26	60.50	54.46	72.40	63.67	56.00
Bus	80.00	70.59	60.67	81.45	72.77	62.67
Dinosaur	100.0	95.0	90.7	100.0	95.0	92.0
Elephant	83.5	75.5	65.8	85.0	77.0	66.0
Flower	90.0	80.5	70.6	92.0	83.0	71.2
Horses	100.0	90.0	80.5	100.0	95.0	83.0
Mountains	70.5	65.8	60.9	72.0	68.0	62.0
Food	60.8	55.8	53.40	62.0	57.0	55.20

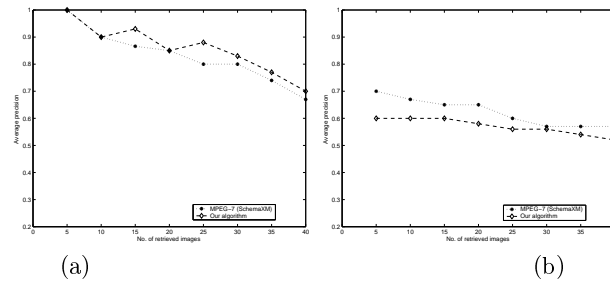
**Table 3.** Standard Visual content descriptors of MPEG-7

Color Descriptors	Texture Descriptors	Shape Descriptors based
Dominant Colors	Edge Histogram	Region Based Shape
Scalable Color	Homogeneous Texture	Contour Based Shape
Color Layout		
Color Structure		

These features have been rigorously tested in the standardization process. We have seen the query results for all the classes available in Schema ( Beaches, buildings, horses, cars, flowers etc. ) which utilize the MPEG-7 visual descriptors. The performance is compared for the case of evaluating overall similarity between images from precision rate defined as,  $P_r = (1/n_1) \sum r/n$  where, the system retrieves  $r$  images that belongs to the same class  $C1$  from  $n$  retrieved images.  $n_1$  is the number images queried from category  $C1$ . The images from other categories are taken as outliers. The results are shown in Figs. 4 and 5. Our results can be fairly compared with MPEG-7 visual descriptors, as seen from Figs. 4 and 5. Although the initial precision is found better for SchemaXM as shown in Fig. 4(a) and Fig.5(b) but our algorithm fairly catches the results.



**Fig. 4.** Comparison with MPEG-7 visual descriptors (a) Category beaches (b) Flowers



**Fig. 5.** Comparison with MPEG-7 visual descriptors (a) Category horse (b) Vehicles

**Conclusion :** In the current work, a fuzzy entropy based relevance feedback, framework for image retrieval is proposed. We intend to test the effectiveness of the system, using simple features , invariant moments in the present case. We plan to improve our scheme by associating text and other features, for better compatibility with MPEG-7 and better semantic modeling.



## 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)** (2000) 1349–1380
2. Kunttu, I., Lepisto, L., Rauhamaa, J., Visa, A.: Multiscale fourier descriptors for defect image retrieval. *Pattern Recognition Letters* **27(2)** (2006) 123–132
3. Gevers, T., Smeulders, A.W.M.: Combining color and shape invariant features for image retrieval. *Image and Vision computing*. **17(7)** (1999) 475–488
4. Chen, Y., Wang, J.Z., Krovetz, R.: Clue: Cluster-based retrieval of images by unsupervised learning. *IEEE Transactions on Image Processing* **14(8)** (2005) 1187–1201
5. : MPEG-7 Multimedia Description Schemes (XM). ISO/IEC JTC1/SC29/WG11 N3914 (2001)
6. Han, J., Ngan, K.N., Li, M., Zhang, H.J.: A memory learning framework for effective image retrieval. *IEEE Transactions on Image Processing* **14(4)** (2005) 521–524
7. Chang, F.C., Hang, H.M.: A relevance feedback image retrieval scheme using multi-instance and pseudo image concepts. *IEICE Transactions on Information and Systems* **D(5)** (2006) 1720–1731
8. Yin, P., 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)** (2005) 1536–1551
9. Lim, J.H., Jin, J.S.: Combining intra-image and inter-class semantics for consumer image retrieval. *Pattern Recognition* **38(6)** (2005) 847–864
10. Rui, Y., Huang, T.S., Mehrotra, S.: Content-based image retrieval with relevance feedback in MARS. In *Proceedings of the IEEE International Conference on Image Processing* (1997) 815–818
11. Jin, Z., King, I., Li, X.Q.: Content-based retrieval by relevance feedback. In *Lecture Notes in Computer Science* **1929** (2000) 521–529
12. 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)** (2006) 1622–1632
13. Qian, F., Zhang, B., Lin, F.: Constructive learning algorithm-based rbf network for relevance feedback in image retrieval. In *Lecture Notes in Computer Science* **2728** (2003) 352–361
14. He, X., King, O., Ma, W., M.Li, H.J.Zhang: 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)
15. Pal, S.K., Chakraborty, B.: Intra-class and inter-class ambiguities (fuzziness) in feature evaluation. *Pattern Recognition Letters* **2** (1984) 275–279
16. Pal, S.K., Chakraborty, B.: Fuzzy set theoretic measures for automatic feature evaluation. *IEEE Transactions on Systems, Man and Cybernetics* **16(5)** (1986) 754–760
17. Pal, S.K., Majumder, D.D.: *Fuzzy mathematical Approach to Pattern Recognition*. Willey Eastern Limited, New York (1985)
18. Banerjee, M., Kundu, M.K.: Content based image retrieval with multiresolution salient points. *Fourth Indian Conference Computer Vision, Graphics and Image Processing, ICVGIP 2004, India* (2004) 399–404
19. Manjunath, B.S., Salembier, P., (editors), T.S.: *Introduction to MPEG-7: Multimedia Content description Interface*. John Wiley and Sons, Inc., USA (2002)