Image retrieval using NN based pre-classification and fuzzy relevance feedback

Malay K. Kundu and Manish Chowdhury

Machine Intelligence Unit, Indian Statistical Institute, Kolkata - 700108, India

malay@isical.ac.in, manishchowdhury2005@gmail.com

***Abstract*— In this article, we have proposed an interactive image retrieval scheme using MPEG-7 visual features, Neural Network (NN)-based pre-classifier and fuzzy based feature evaluation scheme. The performance of the existing image retrieval systems is generally limited due to semantic gap, resulted due to the discrepancies between the computed low-level features and user's conception of an image. Partitioning the database by a NN-based pre-classifier, and using a fuzzy based feature evaluation scheme, the performance of the proposed scheme has been found to improved drastically by reducing the retrieval time and increasing the accuracy.**

***Keywords*— CBIR, MPEG-7, EHD, MLP, DB**

I. Introduction

Content-Based Image Retrieval (CBIR) techniques are aimed at retrieving relevant images from an image database by measuring similarity between the automatically derived low-level features of the query image and that of the images stored in the database. Relevance feedback mechanism has been used as an essential tool to provide significant performance boost in CBIR systems through continuous learning and interaction with end-users [1], [2]. The system provides initial retrieval results through query by example based on which the user judges the retrieved results as to whether and to what degree, they are relevant /irrelevant to the query.

This paper deals with the study of performance of a CBIR system using MPEG-7 Edge Histogram Descriptor (EHD), Multilayer Perceptron (MLP) and fuzzy relevance feedback mechanism. EHD feature are extracted using MPEG-7 visual feature and by using these feature MLP network is trained to classify a class of the image with labelled samples. Then classify the whole image database having particular number of subsets representing different classes. Time requirement for each query searching is proportional to total number of features in each database. To cut down this computation cost, the whole database (DB) is pre-classified by a NN. For every new query, the same network is used for classification. After classification, a particular relevance feedback block is used which access only portion of the DB, representing a particular class and also added to the DB if such data is not already present. This process cuts down the searching space, improving the computational cost and quality of the image retrieval system.

The rest of the article is organized as follows. Section 2. describe the proposed methodology. Experimental results and conclusion are discussed in section 3 and section 4 respectively.

II. Proposed Scheme

In the proposed method, 30% data are randomly selected from the database (DB) as labelled sample for different class of image present and their EHD features (MPEG-7 visual descriptor) are extracted for the training of MLP network. The exact configuration of the MLP is selected on trial and error basis based on the error rate of the output classification. The configuration having minimum error rate is selected for the proposed pre-classification block. Block diagram of the proposed scheme is shown in Figure 1.

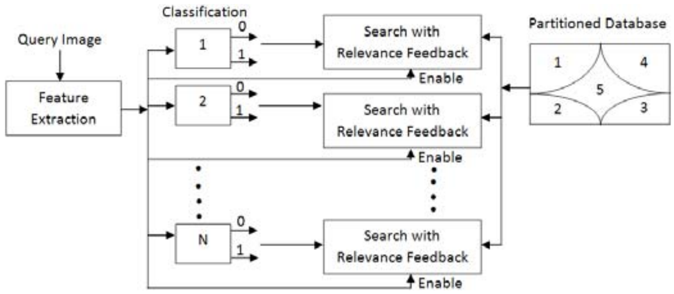

Fig. 1 Block Diagram of Image Retrieval System

The user enters the query image to the system. EHD features are extracted for this particular query image and by using these features; class identification of the query image is done by the proposed pre-classification block. If the class identification of the query image is correct, it will enable the respective fuzzy relevance feedback block and the system computes the similarities between the query image and all the images in the partitioned database of the particular class by using the Euclidean distance. The system retrieves the top-ranked 20 images from the partitioned database of the particular class and presents them to the user. The user marks the images returned to the search engine as *relevant* or *irrelevant* samples. A fuzzy based relevance feedback algorithm [2] uses this feedback information to select a set of better 20 images from the partitioned DB in the next iteration. This retrieval process finishes at a point when the user is satisfied with the retrieved result.

A. NN for pre-classification

To increase the accuracy in the proposed scheme, we have used MLP neural network based pre-classifier with feedforward backpropagation [4] algorithm for learning the network which consists of three layers. There were 81 neurons in input and 5 neurons in output for our learning network. By using the rule of thumb, it is assumed that upper bound of the nodes in the hidden layer is $\sqrt{input\ nodes * output\ nodes}$ we have calculated the number of nodes in a single hidden layer. We have tested our network with the 10, 15, 20 nodes with single hid- den layer, it is found that 20 nodes with single hidden layer produce minimum error rate. We have tested our network with various number of iteration (500 to 3000) and different learning rate (0.1 to 0.7) and to obtain the best possible result. After going through the extensive experimentation learning rate (0.4), iteration (3000) and momentum factor (0.95) is selected for classification as error rate is mini- mum. The network with aforesaid configuration is used as a pre-classifier.

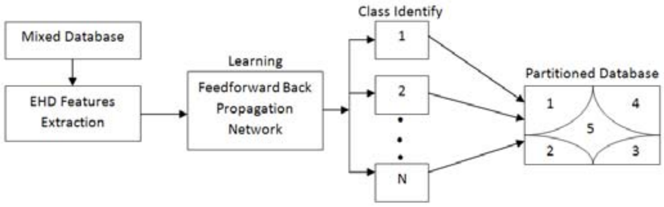

Fig. 2 Block Diagram of N-N pre-Classification

The block diagram of the pre-classification block is shown in Figure 2. EHD features are extracted from the mixed DB, from which 30% of the label data are randomly used to train the network. The feedforward backpropagation algorithm is used for learning the network. When the learning is completed successfully, network is tested on remaining 70% data randomly and their classes are identified to form different partitions of the DB based on the class label.

The feedforward backpropagation neural network is used. The first layer is the input layer, second layer is hid- den layer and has a log sigmoid activation function and the third layer, or output layer, has a linear activation function. All the neurons of one layer are fully interconnected with all the neurons of its just preceding and just succeeding layers. Weights measure the degree of correlation between the activity levels of neurons that they connect. The network is initialized with random weights and biases, and was then trained using the Levenberg-Marquardt algorithm (LM) [4]. Backpropagation is used to calculate the Jacobian *JX* of performance with respect to weight and bias variables X. Each variable is adjusted according to LM as shown in Eqn. (1, 2, 3), where *I* the identity unit matrix, *E* the error at the output and μ the learning parameter.

$$JJ = JX \times JX \quad (1)$$

$$Je = JX \times E \quad (2)$$

$$dX = - (JJ + \mu I)/Je \quad (3)$$

The learning function used in the proposed network is gradient descent with momentum weight/bias function as shown in Eqn. (4)

$$dW = MC * dWprev + (1 - MC)*LR*gW \quad (4)$$

where weight change *dW* for a given neuron was calculated from the neuron's input and error, the weight or bias *W*, learning rate (*LR*), and momentum constant (*MC*) according to the gradient descent with momentum. A momentum term could be added to increases the learning rate with stability. The *gW* define gradient with respect to performance. The performance of the network is measured by mean squared error (MSE), which can be quantitatively calculated. The smaller the MSE is, the better the network performs.

B. *Relevance Feedback Mechanism after Classification*

To preserve interactivity, the relevance feedback mechanism implemented in a search engine must operate in real time. A fuzzy based particular relevance feedback block will enable only when the respective class is identify. Relevance feedback retrieval systems prompt the user for feed- back on retrieval results and then use this feedback on sub- sequent retrievals with the goal of increasing retrieval performance. At each iteration user rates each returned result with respect to how useful the result is for his or her retrieval task at hand. Ratings may be simply *relevant* or *irrelevant*. A fuzzy based relevance feedback algorithm uses this feed- back information to modify the relative weight of different feature values to select another set of 20 images to retrieve for the user. The retrieval process, for a given query image, finishes at a point when the user is satisfied with the retrieved images as shown in Figure 3.

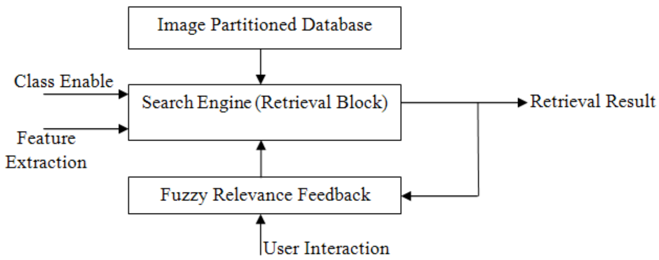

Fig. 3 Block Diagram of Relevance Feedback

In conventional CBIR approaches an image *I* is usually represented by a set of features, $F = \{f_q\}^N_{q=1}$, where f_q is the q^{th} feature component in the *N* dimensional feature space. Euclidean distance is used with weight w_q for measuring the similarity between query image I_{qr} and other images *I*. The weights should be adjusted such that, the features have small variation over the relevant images and large variation over the irrelevant images. The information from relevant images (I_r) and irrelevant images (I_{ir}) are combined to compute the relative importance of the individual features, from fuzzy feature evaluation index (FEI) [5] in pattern classification problems.

The FEI is defined from interclass and intraclass ambiguities as follows: Let C_1, C_2,.. C_j ... C_m be the m pattern classes in an N dimensional ($f_1, f_2, f_q, \ldots : f_N$) feature space where class C_j contains, n_j number of samples. The entropy of a fuzzy set gives a measure of intraset ambiguity along the q^{th} co-ordinate axis in C_j is computed in Eqn. (5),

$$H(A) = \left(\frac{1}{n_j \ln 2}\right) \sum_{i} S_n\left(\mu\left(f_{iqj}\right)\right) ;\; i = 1,2, \ldots n_j \tag{5}$$

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 (μ). $H_{min} = 0$ for μ=0 or 1, $H_{max} = 1$ for μ=0.5. Entropy (H) of C_j along q^{th} component can be computed using a standard S type membership function [2], [5].

The criterion of a good feature is that, it should be nearly invariant within class, while emphasizing differences between patterns of different classes [5]. The value of H would therefore decrease, after combining the class C_j and C_k as the goodness of the q^{th} feature in discriminating pat- tern classes C_j and C_k increases. The measure denoted as H_{qjk} is called "*interset 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 given in Eqn. (6),

$$\left(FEI_q\right) = \frac{H_{qjk}}{H_{qj} + H_{qk}} \tag{6}$$

Lower value of FEI_q, indicates better quality of importance of the q^{th} feature in recognizing and discriminating different classes [2]. The precision of retrieval can be improved with these values. The relevant images constitute the (intraclass) and the irrelevant images constitute the (interclass) image features. To evaluate the importance of the q^{th} feature, the q^{th} component of the retrieved images is considered. H_{qj} and H_{qk} is computed from $I_r^{(q)}$ and $I_{ir}^{(q)}$ respectively. H_{qjk} is computed by 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. (7),

$$w_q = F_q(FEI_q) \tag{7}$$

In the first pass, all features are considered to be equally important. The feature spaces of the relevant images are there- fore 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. Multiplying by w_q increases the feature separation between the irrelevant components, such that due to the combined effect the irrelevant image may be pulled down.

III. –EXPERIMENT RESULTS AND DISCUSSION

To prove the effectiveness of the proposed system, extensive experiments have been performed on MPEG-7 EHD features upon 500 images of Simplicity databases [6] of 5 different classes (tribal people, ocean, building, bus and dinosaurs). The results are compared with the EHD and M-band wavelet without classification [2], [3]. The weight updating formula $w_q = FEI_q^2$ is used in each iteration as it generate better results in majority of the cases. EHD feature are chosen to evaluate the overall similarity between images and produced better results where spatial distribution of edges and semantic significance is more important [7]. The experiment was performed on Dell Precision T7400 with 4GB RAM machine and was implemented by using MATLAB R2008a.

The proposed relevance feedback performance after class identification is measured in terms of precision as de- scribed in [2]. 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 20 retrieved set, in each round of feedback iteration and the remaining are negative examples for up- dating the weight parameters and revising the features. One such example may be the case of a tribal people face where the precision is from 50% to 70% without doing the class identification. But the rate of precision increases drastically by doing the class identification before using the relevance feedback mechanism i.e. from 80% to 85% as shown in Fig. 4, Fig.5 and Fig.6. The average precision is obtained from the set of same query after different iterations. From the graph of Fig. 7, it can be shown that the accuracy achieved in the system with classifier is better than our earlier work [2, 3] for the same query after 3rd or 4th iteration.

Fig. 4 Retrieval result obtain without any iteration (Upper left hand side is a query image)

Fig.5 After first iteration

Fig. 6 After fourth iteration

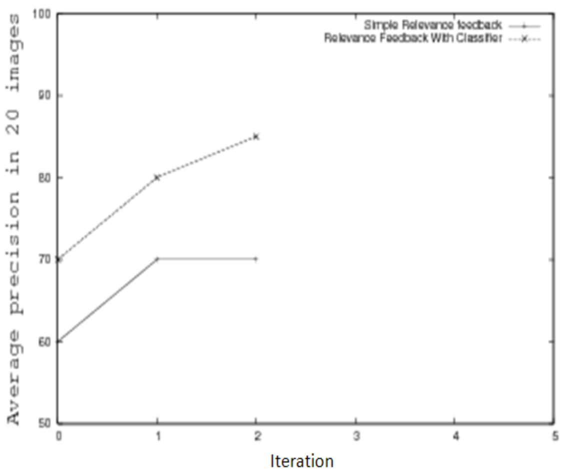

Fig. 7 Comparative studies of relevance feedback with and without classifier

IV. CONCLUSION

Experimental result shows that the proposed image retrieval system with classifier based on MPEG-7 EHD features is able to improve the accuracy of the retrieval performance. If the pre-classification is wrong then our whole retrieval result will be wrong. To overcome this problem we are trying to implement fuzzy ranking membership function. This proposed mechanism could be tested for video retrieval as future scope of research.

REFERENCES

[1] P. Y. Yin, and B. Bhanu, and K. C. Chang and A. Dong, "Integrating Relevance Feedback Techniques for Image Retrieval Using Reinforcement Learning", IEEE Transactions on Pattern Analysis and Machine Intelligence , Vol 27, No. 10, pp. 1536-1551, 2005..

[2] M. K. Kundu, and M. Banerjee, and P. Bagrecha, "Interactive Image Retrieval in a Fuzzy Framework", Proceedings 8th International Workshop on fuzzy logic and Application , Palermo, Italy, pp. 246-253, 2009.

[3] M. K. Kundu, and P. Bagrecha, "Color Image Retrieval Using M-Band Wavelet Transform Based Color-Texture Feature ", Proceedings of 7th International Conf. on Advances in Pattern Recognition (ICAPR 2009) , Kolkata, India, pp. 117-120, 2009..

[4] A. L. Betker, and T. Szturm, and Z. Moussavi, "Application of feedforward backpropagation neural network to center of mass estimation for use in a clinical environment", Proceedings of the 25th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Vol 3, pp. 2714-2717, 2003.

[5] S. K. Pal, and B. Chakraborty, "Intraclass and interclass ambiguities (fuzziness) in feature evaluation", Pattern Recognition Letters, Vol 2, pp. 275-279, 1984.

[6] Z. Wang and J. Li and G. Wiederhold , "Simplicity: Semantics sensitive integratedmatching for picture libraries", IEEE Transactions on Pattern Analysis and Machine Intelligence , Vol 23, No. 9, pp. 947-963, 2001

[7] B. S. Manjunath and J. R. Ohm and V. V. Vasudevan and A. Yamada, "Color and Texture Descriptors", IEEE Transactions on Circuits and Systems for Video Technology , Vol 11, No. 6, pp. 703-715, 2001