Relevance Feedback Optimization in Content Based Image Retrieval Via Enhanced Radial Basis Function Network

Shahrooz Nematipour, Jamshid Shanbehzadeh, Reza Askari Moghadam

***Abstract*— this paper proposes a novel adaptive method to improve relevance feedback procedure in content based image retrieval. First, we transform low-level features to high-level ones by means of a multilayer neural network and these features are employed as the input of a radial basis function network for relevance feedback. This approach reduces the semantic gap and feature dimensionality considerably. In low-level into high-level feature transformation, we employ one thousand images of Corel database in training phase and, ten thousands images from the same database to test the relevance feedback. The experimental result shows the improvement of precision rate due to fast convergence after third iteration.**

***Index Terms*— Relevance feedback, Multilayer neural network, content-based image retrieval, radial basis function network**

I. INTRODUCTION

Content based image retrieval (CBIR) system employs low-level features such as colour, shape and texture as a measure to find similar images [1]. These low-level features are insufficient to describe image contents similar to human visual perception (HVP). The most important reason is the semantic gap between low level features and VHP [2].

The lost ring in the CBIR systems is the user interaction, in which the user gets involved in the retrieval process by giving feedback about the images displayed by the system and refining the search until satisfied, and transferring his divine visual perception to the CBIR systems [2].

Another factor to be taken into account is computational time, because working with high dimensional feature vectors may become too time-consuming, where the environment of CBIR systems is real-time. Therefore, transforming low-level attributes into relevant vectors is an important step of the CBIR lifecycle. In other words, a feature vector with low dimension should be able to represent the semantic information familiar to HVP [3].

Feature transformation is employed to enrich image representation by high-level vectors from low-level features (colour, shape and texture) and reduce the dimension of feature vector [4]. Relevance feedback tends to minimize the semantic gap by enriching the retrieval process based on HVP.

This work proposes to join these two approaches. The power of multilayer neural network which will be used beyond its adaptation and learning ability in connection with the power of fuzzy radial basis function network (FRBFN) to enhance the RF in the retrieval cycle. This technique reduces the data dimensionality and semantic gap in parallel. The use of FRBFN is beneficial for two reasons. First of all, it incorporates fuzzy nature of human decision into the system. Moreover the convergence time is low due to its fast learning algorithm which lacks back-propagation

.

II. RELATED WORK

The problem of retrieving and recognizing patterns in images has been investigated for several decades by the image processing and computer vision research communities. Learning approaches such as multilayer neural networks, RBF networks and Kohonen feature map can be employed to obtain satisfactory results for very specific applications. Image retrieval is a well-known field of research in which a large number of methods have been designed but still no satisfying general solution exists.

Multilayer neural networks have been used in many image retrieval and characterization applications mainly to generate feature vectors also called "high-level semantic vectors" through image classification by machine learning [5,6].In [7], a retrieval system via neural network is proposed where each image is characterized by low-level descriptors and associated to one or more concepts preset in the image. The result shows that the neural network is able to memorize these categories accordingly. However in images which contain several semantic concepts, the result deteriorates. In contrast, in [8] the neural network is used in its final stage (after training) to characterize images in generic form based on the intrinsic semantic categories of the query image instead of predefined categories. These semantic aspects are acquired during training and used in the low-level into high-level feature transformation.

From the relevance feedback point of view, many algorithms such as query refinement [9, 10], reweighting

Manuscript received October 2010 and revised January 2011-01-08. This work is part of MSc thesis in Payame Noor University- Tehran, I. R. Iran.

Shahrooz Nematipour is a postgraduate student in the field of computer engineering, in Department of Information technology at Payame Noor University, in Iran. He is interested in image retrieval and its application on search engines. (e-mail : shahrooz26382@yahoo.com).

Dr. Jamshid Shanbehzadeh is an Associate Professor in the Department of Computer Engineering at Tarbiat Moallem University in Tehran, I. R. Iran. He has published over eighty international journal and conference papers in the areas of image processing, computer vision, emotion detection and affective tutoring systems. He is an expert in e-learning and artificial intelligence in education. (Corresponding author.Phone: +98 9121484177; e-mail: jamshid@tmu.ac.ir)

Dr. Reza Askari Moghadam is an assistant professor in Department of Information technology at Payame Noor University, in Iran. He has published many journal and conference papers in the area of neural networks.(Corresponding author.Phone:+989123060417;e-mail: askari@pnu.ac.ir)

ISBN: 978-988-18210-3-4
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

[11], Bayesian learning [12] and kernel-based learning [13] have been adopted in CBIR systems, neither of which have been efficient enough to boost the retrieval performance due to the lack of attention to the fuzzy logic embedded in the human decision.

Traditionally, users are limited to binary classification as to label an image "fully relevant" or "totally irrelevant" [9,12].Therefore it is unable to show the nature of user interpretation which is completely vague due to the fact that information needs are always changing and different among different people. Hence neither multi-level labeling nor binary labeling can help in this regard while users have the tendency toward using linguistic expressions [14].

In [15], the aforementioned approach has been implemented via a FRBFN. The FRBFN is constructed dynamically based on all the accumulated training samples over previous feedback sessions. Actually it is an online learning process in which the users interact with the system in real time until pertinent result is achieved.

Many content based applications are self-insufficient, meaning that it is impossible to utilize them in all areas. In addition the semantic gap between human & computers differ for varying applications. There is a long road to reach an ideal method which is efficient in all cases, because there is not enough knowledge about visualization process taking place in the brain. Based on this fact the CBIR is the junction point for many scientific fields like information retrieval, machine learning and human-computer interaction which should work together so that an efficient CBIR system is reached.

The proposed model in this paper makes use of two principles: 1-Feature transformation by using multi-layer neural network as a pre-processing phase to reduce the semantic gap.2-Relevance feedback integration by implementing FRBFN whose inputs, instead of low-level feature vectors, are now high-level vectors generated in the pre-processing phase by multi-layer neural network for maximizing semantic gap reduction. In other words, the research proposal in this work aims to tackle semantic gap from two sides, one from bottom , by low-level into high-level transformation and one from top, by using relevance feedback, which in this case, is implemented by FRBFN whose structure is perfect for classification and accuracy due to its fuzzy structure.

III. SUGGESTED MODEL

Our model consists of two main parts. The first part is knowledge representation and high-level semantics through examples to transform the low-level features into high-level vectors. The second part is relevance feedback through dynamically constructing FRBFN for learning different degrees of relevance embedded in the users' interpretation of visual contents. In the following subsections each of the aforementioned parts will be explained briefly.

IV. NEURAL NETWORK IMPLEMENTATION

This implementation of multilayer neural network consists of two phases. The first phase is training where a sample database, whose images are labeled, is chosen. After labeled, the images are passed through low-level descriptors such as colour, shape and texture for the formation of the network input. During the neural network training, each feature vector C_i is associated with a target vector A_i whose cardinality is equal to the number of groups in the training database. A fully connected feed-forward network is chosen for our implementation. The back-propagation learning rule is applied until the network convergence is reached. Convergence is determined in relation to the labels given by the images.

After the network is trained, the next phase will use this information in other images in another database to generate the high-level feature vectors. The network without the activation function will produce an output vector N_i for each low-level vector which is subsequently stored in database along with the image for use in retrieval process.

V. RETRIEVAL PROCESS

The complete model of the retrieval system is an ordinary CBIR system, where the user inserts the query image Q into the system, and then the low-level features, whose values are used as the input to the neural network in the previous phase, are extracted. The output vector will be compared with the high-level vectors stored in the database by means of a standard similarity measure like Euclidean distance. By means of K-nearest algorithm, top k images are displayed for feedback.

VI. FRBFN LEARNING

The architecture of the FRBFN is given in Figure 1.It is composed of an input layer, A Gaussian kernel layer and an output layer. The input data will be p-dimensional high-level feature vector. They are connected to the Gaussian kernel layer which is constructed dynamically based on relevant, irrelevant and fuzzy samples. The procedure is to cluster samples in each relevant, irrelevant and fuzzy categories.
Let V= $\{v_1, v_2, \ldots, v_k\}$ be a set of p-dimensional FRBF centers. The output for an input vector of an image X is:

$$F(x) = \sum_{i=1}^{k} W_i\, f(x, V_i, \delta_i) = \sum_{i=1}^{k} W_i\, e \times p\left(\frac{-(x-V_i)\Lambda(x-V_i)^T}{2\delta_i^2}\right) \quad (1)$$

W_i: Connection weight of the output layer.
V_i, δ_i: center & corresponding width of the ith FRBF unit

$$\sigma_i = \gamma . \min_j \|V_i - V_j\| \quad j = 1,2,\ldots,k \quad j \neq i \quad (2)$$

Where Υ is a factor which controls the overlapping of different FRBF centers.Λ=diag[$a_1,\ldots,a_p$] is a diagonal matrix that denotes the relative importance of different feature components determined by positive samples.

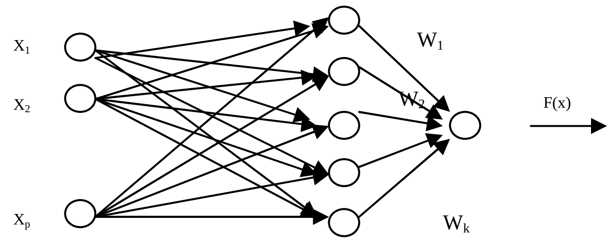

Fig. 1 Architecture of RBF network

ISBN: 978-988-18210-3-4
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

VII. FRBFN LEARNING ALGORITHM

The parameters of the network are adjusted in an online error correction procedure. The error function is defined as:

$$E = \frac{1}{2}\sum_{j=1}^{N} e^2 j = \frac{1}{2}\sum_{j=1}^{N}(Y_j - F(X_j))^2 \quad (3)$$

N: Number of total training samples

$F(x_j)$: actual network output for the jth training sample x_j

Y_j: desired network output for x_j,which is 1 for relevant 0 for irrelevant and the range [0,1] for fuzzy samples.
In particular we want to find the probability $p(w_r|x_j)$that a fuzzy sample belongs to the relevant class w_r.
The following formulas (4-8) calculate this probability.

$$p(w_r|x_j) = \frac{1}{M}\sum_{i=1}^{M} p(w_r|x_j i) \quad (4)$$

$$p(w_r|x_j i) = \frac{p(x_j i|w_r)p(w_r)}{p(x_j i|w_r)p(w_r)+p(x_j i|w_{ir})p(w_{ir})} \quad (5)$$

$$\Rightarrow p(x_j i|w_r) = \frac{1}{(2_\pi)^{di/2}\left|\sum_i^M\right|^{\frac{1}{2}}}$$

$$.e \times p\left[\frac{-1}{2}(x_{ji} - \mu_i^m)^T \sum_i^{m-1}(x_{ji} - \mu_i^m)\right] \quad (6)$$

$\mu_i^m = mean\ vector$

$\sum_i^m = covariance\ matrix\ for\ the\ ith\ feature\ vector\ of\ class\ W_m$

$$\mu_i^m = \frac{1}{N_m}\sum_{j=1}^{N_m} x_{ji}^m, \quad (7)$$

$$\sum_i^m = \frac{1}{(N_m - 1)}\sum_{j=1}^{N_m}(x_{ji}^m - \mu_i^m)(x_{ji}^m - \mu_i^m)^T \quad (8)$$

By minimizing the cost function using gradient-descent method, we update 3 parameters of the network:

$$w_i(t+1) = w_i(t) - \eta_1 \frac{\partial E(t)}{\partial w_i(t)} \quad i - 1,2,\ldots,k \quad (9)$$

$$\frac{\partial E(t)}{\partial w_i(t)} = -\sum_{j=1}^{N} e_j(t)\, t(x_j, v_i), \delta_i(t)) \quad (10)$$

$$v_i(t+1) = v_i(t) - n_2 \frac{\partial E(t)}{\partial v_i(t)} \quad (11)$$

$$\frac{\partial E(t)}{\partial v_i(t)} = -w_i(t)\sum_{j=1}^{n} e_j(t) f(x_j, v_i(t), \partial_i(t)) \frac{\wedge(x_j - v_i(t))}{\delta_i^2(t)} \quad (12)$$

$$\delta_i(t+1) = \delta_i(t) - n_3 \frac{\partial E(t)}{\partial \delta_i(t)} \quad (13)$$

$$\frac{\partial E(t)}{\partial \delta_i(t)} = -w_i(t)\sum_{j=1}^{n} e_j(t) f(x_j, v_i(t), \delta_i(t)) \frac{(x_j - v_i)^T \wedge (x_j - v_i)}{\delta_i^3(t)} \quad (14)$$

Figure 2 shows the general model of the retrieval process for the proposed system.

VIII. EXPERIMENTAL RESULTS

This section describes the test done based on the proposed model and optimization obtained by the model is depicted via a diagram. This diagram compares two models, one with normal and pure feature vectors and the proposed one with low-level into high-level transformation procedure included.

IX. LOW-LEVEL FEATURE EXTRACTION

The low-level features extracted in our model were:128 colour histogram bins,9 colour moments and 7 hue's moments totaling 144 neurons in the input layer along with 100 neurons in the hidden layer and 10 in the output layer in as much as there are 10 distinct semantic categories in our training database. The Corel 1000 database from Corel Corporation which contains 1000 photos from 10 varying categories was used for training the network. The test database was also from Corel gallery. It has about 10000 images. The low-level features extracted by our algorithm are not saved directly in the database. Instead, they serve the input for the previously trained network by Corel 1000 gallery in order to generate high-level feature vectors which are saved for each of the 10000 images in the database.

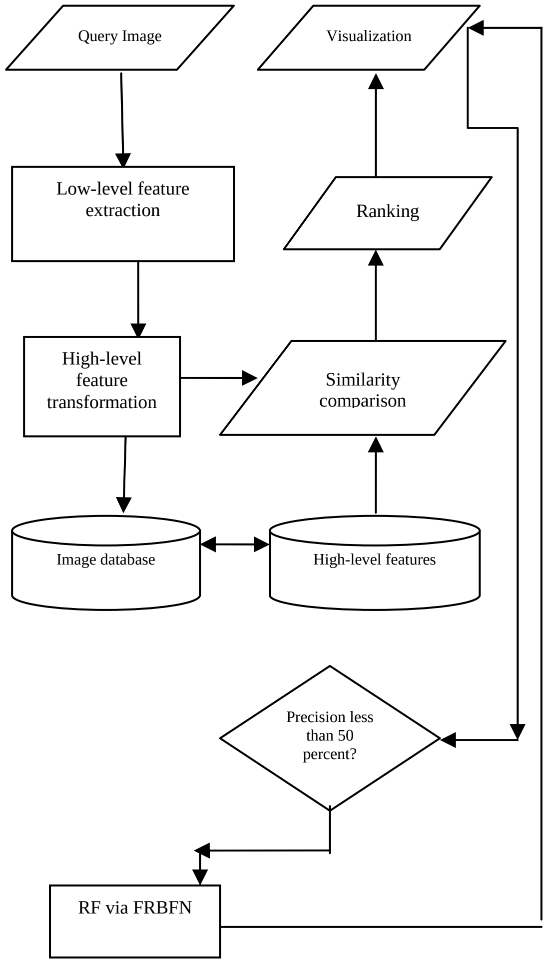

Fig. 2 The general depiction of the proposed model

When a query image is given to the system, the procedure which took place for the training database images, is done for the query image. Based on k-nearest neighborhood, the initial K results are returned and displayed to the user for feedback. In our test, top 20 images are displayed for feedback. Therefore 'k' here is 20.

We perform subjective test to evaluate the effectiveness of our proposed method. A total of 50 query images are used for evaluation. For each query the top 20 images are displayed for feedback. The measure used is defined here:
TRA= (relevant and fuzzy images in top T returns) /T
TRA parameter is compared for two methods: 1-simple FRBFN with low-level features as input and 2-optimized FRBFN with high-level features as input. Figure 3 depicts this comparison. The blue curve shows the results with simple RBF as relevance feedback and the red curve shows

ISBN: 978-988-18210-3-4
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

the experiments with the optimized RBF network as RF instead.

It can be observed that the second method requires less iteration to reach a specific retrieval precision in spite of the fact that the dimensionality of the feature vector has been declined tremendously.

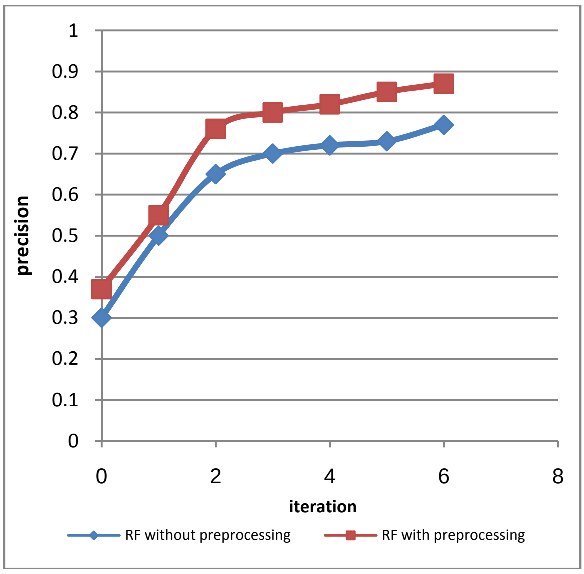

Fig. 3 Performance comparison of simple FRBFN and enhanced FRBFN

X. CONCLUSION

In this work a new model was presented which makes use of two contrasting ideas to fill the gap generated by low-level computation in CBIR systems. The first idea was to characterize images through neural network which involves the use of low-level features as support for the high-level vector generation represented by the neural network knowledge. The second idea was to integrate the user's fuzzy interpretation of image similarity into CBIR system by using FRBFN which tends to be more flexible than the ordinary relevance feedback algorithms. We have tested the technique on the Corel 10000 Gallery which is suitable for CBIR simulation. The test compared simple FRBFN with its enhanced version whose inputs were high-level vectors. The experimental result showed effectiveness in accordance to retrieval precision.

For further semantic gap reduction, it is recommended to work on relevance feedback logging and use statistical image retrieval as a big resource for forecasting the upcoming images.

REFERENCES

[1] E Dorileo, AG Frade, MAC Roselino, "Color image processing and content-based image retrieval techniques for the analysis of dermatological leisons"Engineering in Medicine and Biology Society,IEEE, 2008.

[2] G Carneiro, AB Chan, PJ Moreno, N Vasconcelos ,"Supervised learning of semantic classes for image annotation and retrieval ", - IEEE Transactions on Pattern Analysis and Machine, 2007

[3] H.Liu and H.Motoda.Feature transformation and subset selection.IEEE intelligent systems,13(2):26-28,1998

[4] Nixon, Mark S., Aguado, Alberto S., "Feature extraction and image processing",Second Edition.Academic Press,2008

[5] E Nowak, F Jurie, B Triggs ,"Sampling strategies for bag-of-features image classification", - Lecture Notes in Computer Science, - Springer , 2006

[6] Y Liu, D Zhang, G Lu, WY Ma ,"A survey of content-based image retrieval with high-level semantics", Pattern Recognition, 2007 , Elsevier

[7] X.Ma and D.Wang.Semantics modeling based image retrieval system using neural networks.In ICIP,pages 1165-1168,2005

[8] Celia A. Z., Barcelos, Eduardo F., Ribeiro ,Marcos A. Batista,"Image Characterization via Multilayer Neural Networks",2008 20th IEEE International Conference on Tools with Artificial Intelligence,2008

[9] DS Yang, M Li, L Liu ," New relevance feedback based on re-ranking mechanism in content-based image Retrieval" Gongcheng yu Yingyong(Computer Engineering and …, 2008 - csa.com

[10] Danzhou Liu, Kien A. Hua, Khanh Vu, and Ning Yu,"Fast Query Point Movement Techniques with Relevance Feedback for Content-based Image Retrieval",ACM,2006

[11] ZP Luo, XM Zhang ," Learning Based Relevance Feedback Algorithm in Content-Based Image Retrieval" Pattern Recognition, 2008. CCPR'08., 2008 - ieeexplore.ieee.org

[12] Zhang ,Ruofei *,Zhang ,Zhongfei (Mark) "BALAS: Empirical Bayesian learning in the relevance feedback for image retrieval", Image and Vision Computing,Elsevier,2005

[13] *Rongrong Ji, Hongxun Yao, Jicheng Wang, Pengfei Xu,1Xianming Liu,"* Clustering-Based Subspace Svm Ensemble For Relevance Feedback Learning", 2008 IEEE

[14] Miguel Arevalillo-Herra´ez a, Mario Zacare´ s b, Xaro Benavent c, Esther de Ves a,"A relevance feedback CBIR algorithm based on fuzzy sets",Signal Processing: Image Communication,Elsevier,2008

[15] Kim-Hui, Yap,Kui, Wu,"Fuzzy relevance feedback in content-based image retrieval systems using radial basis function network",IEEE,2005

ISBN: 978-988-18210-3-4
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)