

REGION-BASED IMAGE RETRIEVAL USING RADIAL BASIS FUNCTION NETWORK

Kui Wu, Kim-Hui Yap and Lap-Pui Chau

School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore

ABSTRACT

This paper presents a new framework that integrates relevance feedback into region-based image retrieval (RBIR) systems based on radial basis function network (RBFN). A modified unsupervised subtractive clustering algorithm is proposed for RBFN center selection according to the characteristics of region-based image representation. A new kernel function of RBFN is introduced for image similarity comparison under region-based representation. The underlying network parameters (weight and width) are then optimized using a supervised gradient-descent training strategy. Experimental results using a database of 10,000 images demonstrate the effectiveness of the proposed hybrid learning approach.

1. INTRODUCTION

Content-based image retrieval (CBIR) is a process of retrieving a set of desired images from an image collection on the basis of visual contents such as color, texture, shape or spatial relationship that are present in the images. These low-level features, however, may not correspond to the users' dynamic and subjective interpretation of image contents under various circumstances. In view of this, relevance feedback, as an interactive mechanism, has been introduced to facilitate image retrieval [1]. Many relevance feedback algorithms have been adopted in CBIR systems and demonstrated considerable performance improvement.

While users generally look for visual objects in images, the global features extracted from images sometimes fail to match the users' object-level perceptions. Consequently, region-based image retrieval systems have been developed to represent images at the object level to achieve higher retrieval accuracy [2-3]. They utilize image segmentation to decompose an image into regions, and then perform similarity comparison based on the local features extracted from the constituent regions.

Relevance feedback is traditionally used in CBIR systems that use global features. There are still relatively few works that incorporate it in the RBIR systems. Nevertheless, some preliminary works have been done to

combine RBIR and relevance feedback together to improve the retrieval performance. Existing feedback methods in RBIR systems can be mainly categorized into query point movement (QPM) [3], feature and region re-weighting [4], support vector machines (SVM) [3], and multiple instance learning (MIL) [5].

In this paper, we propose a new progressive RBFN-based relevance feedback method to model and learn the users' perception of image similarity in RBIR systems. Subtractive clustering is employed to construct the network. An efficient gradient-descent-based learning strategy is utilized to estimate the network parameters. The trained RBFN is then used in the next sessions to retrieve the images.

2. REGION-BASED IMAGE RETRIEVAL

We use mean shift algorithm as our image segmentation method to partition images into homogenous regions [6]. After image segmentation, for each region in the image, the color and texture descriptors of the region are extracted and used to represent the regions. Color moments are used as the color feature while wavelet moments are used as the texture feature. Further, we use region weights to reflect the importance of each region. For initialization, a uniform weighting scheme is adopted, that is, every region in the image is assumed to be equally important. The summation of the region weights within an image is equal to one. Later, with users' feedbacks, a weight updating scheme is designed to adaptively tune the regions' weights to reflect their importance. The detailed algorithm will be described in later sub-sections.

We employ the Earth Mover's Distance (EMD) to compare two images with their varying constituent regions [7]. The EMD is based on the minimal cost that must be paid to transform one distribution into the other, which can be solved as a linear optimization problem.

3. RBFN HYBRID LEARNING ALGORITHM

3.1. Structure of Progressive RBFN

In interactive RBIR systems, global modeling of image similarity may not address the local information contained in the query and feedback images adequately. To exploit the

local characteristics of image relevance, it is more desirable to adopt a multi-cluster local modeling strategy. RBFN enjoys the following characteristics that make it a good choice in the context of RBIR: fast learning speed, simple network structure, multiple local modeling feature, and global generalization power.

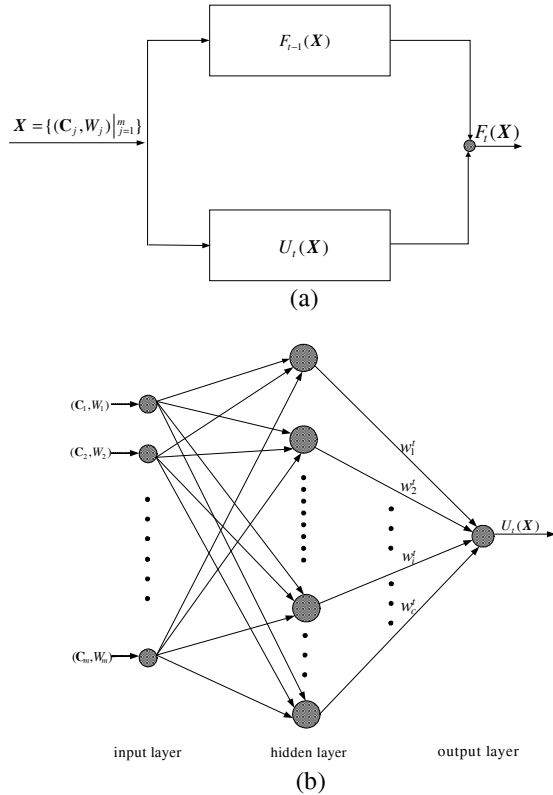


Fig.1. Schematic diagram of RBFN. (a) Recursive structure of RBFN, (b) Update network, $U_t(X)$

The schematic diagram of the network is given in Fig.1. It has a recursive structure (Fig.1(a)) because the RBFN during the previous iterations has been trained to learn the user information need, therefore it is unnecessary to re-train the whole network for each feedback iteration. The input data to the network is an image $X = \{(C_j, W_j) |_{j=1}^m\}$ that consists of m regions $\{R_j\}_{j=1}^m$ with C_j and W_j representing the feature vector and weight of the region R_j . A small update network $U_t(X)$ (Fig.1(b)) is developed to track the current user information need. It has an input layer, a hidden layer and an output layer. The input layer receives images of region-based representation. It is connected to the hidden layer which is constructed from the feedback images. The output layer has a single unit whose output value is the weighted combination of all the responses from each RBF unit. $\{w_i^t\}_{i=1}^c$ is the set of output connection weights. The t -

th iteration RBFN $F_t(X)$ is built from the $(t-1)$ -th iteration RBFN $F_{t-1}(X)$ progressively, and combined with the t -th iteration update network $U_t(X)$. $F_t(X)$ can be expressed as:

$$F_t(X) = \begin{cases} F_{t-1}(X) + U_t(X) & t = 1, 2, \dots, T \\ 0 & t = 0 \end{cases} \quad (1)$$

where T is the total number of feedback iterations. The method is computationally efficient since only the parameters of the update network need to be estimated. It is also algorithmically effective as all feedback samples accumulated over the previous iterations are used to train the update network.

3.2. Subtractive Clustering for RBF Center Selection

An important issue in the development of RBFN is the determination of RBF centers. Taking into account the fast-response requirement and the local multi-cluster nature of the samples, we adopt a clustering process to determine the RBF centers based on their types, relevant and irrelevant. In our retrieval system, subtractive clustering is used as it is fast, efficient and does not require the number of clusters to be specified *a priori* [8]. Subtractive clustering assumes each sample as a potential cluster center. It computes a potential field which determines the likelihood of a sample being a cluster center. Considering the variable-length, multi-region representations, and many-to-many relationship of region mapping between any two images, we propose a modified subtractive clustering algorithm in the context of the EMD as follows.

Let $\{X_i\}_{i=1}^n$ be the set of n relevant or irrelevant samples, where $X_i = \{(C_{ij}, W_{ij}) |_{j=1}^m\}$ contains m regions $\{R_{ij}\}_{j=1}^m$ with C_{ij} and W_{ij} representing the feature vector and weight of the region R_{ij} . The subtractive clustering algorithm can be summarized as:

(i) Compute the initial potential function $P_{k=1}(i)$ of the i -th sample X_i for $i = 1, \dots, n$, and select the sample with the highest potential as the first cluster center. $P_{k=1}(i)$ is expressed in terms of the EMD to the other samples X_j , and defined as:

$$P_{k=1}(i) = \sum_{j=1}^n \exp\left(-\frac{\text{EMD}^2(X_i, X_j)}{r_a^2}\right), \quad i = 1, \dots, n \quad (2)$$

where r_a is a positive coefficient defining the range of the field.

(ii) For $k = 2, \dots, K$, update the potential of each sample according to:

$$P_k(i) = P_{k-1}(i) - P_{k-1}^* \exp\left(-\frac{\text{EMD}^2(\mathbf{X}_i, \mathbf{X}_{k-1}^*)}{r_b^2}\right), \quad i = 1, \dots, n \quad (3)$$

where \mathbf{X}_{k-1}^* and P_{k-1}^* are the $(k-1)$ -th cluster center and its potential value, r_b is a positive coefficient defining the neighborhood radius for potential reduction, and K is the maximum number of potential clusters.

(iii) Repeat step (ii) until the termination criterion is satisfied.

3.3. Probabilistic Region Weight Learning

After subtractive clustering, we obtain a set of c cluster centers at the t -th feedback iteration $\{\mathbf{V}_i^t\}_{i=1}^c$, where

$\mathbf{V}_i^t = \{(\mathbf{C}_{ij}, W_{ij})\}_{j=1}^m$ has a region-based representation.

These estimated centers can belong to either the relevant class ω_r or irrelevant class ω_r . Since different regions in an image have unequal importance for computing image similarity, we propose a probabilistic region weight learning method to capture the relevance of the constituent regions in each cluster center. Inspired by the idea of *a posteriori* estimation, we formulate the problem into the estimation of the conditional probability $P(\omega | R_{ij})$, namely, given a region R_{ij} of a cluster center \mathbf{V}_i^t from class $\omega \in \{\omega_r, \bar{\omega}_r\}$, what is the probability that it belongs to ω . Intuitively, the larger this probability, the more likely and relevant the region is in reflecting its class information. The feedback region importance/weight can be expressed as:

$$W_{ij} = P(\omega | R_{ij}) \quad (4)$$

According to the Bayesian theorem, we can rewrite (4) as:

$$W_{ij} = \frac{p(R_{ij} | \omega)P(\omega)}{p(R_{ij} | \omega)P(\omega) + p(R_{ij} | \bar{\omega})P(\bar{\omega})} \quad (5)$$

where $\bar{\omega}$ denotes the complement of ω . $P(\omega)$ and $P(\bar{\omega})$ are the prior probabilities of class ω and $\bar{\omega}$, respectively. They can be estimated from the feedback samples using:

$$P(\omega) = \frac{N_\omega}{N_\omega + N_{\bar{\omega}}} \quad (6)$$

$$P(\bar{\omega}) = 1 - P(\omega) \quad (7)$$

where N_ω and $N_{\bar{\omega}}$ are the total number of feedback samples accumulated over previous feedback iterations from class ω and $\bar{\omega}$, respectively. $p(R_{ij} | \omega)$ and $p(R_{ij} | \bar{\omega})$ are the class conditional probability density functions (pdf) of region R_{ij} for class ω and $\bar{\omega}$, respectively. They can be estimated using the following approximation principle:

$$p(R_{ij} | \omega) = \frac{N_{\omega, R_{ij}}}{N_\omega} \quad (8)$$

$$p(R_{ij} | \bar{\omega}) = \frac{N_{\bar{\omega}, R_{ij}}}{N_{\bar{\omega}}} \quad (9)$$

where $N_{\omega, R_{ij}}$ and $N_{\bar{\omega}, R_{ij}}$ represent the number of sample images in classes ω and $\bar{\omega}$ that are similar to region R_{ij} , respectively. Region R_{ij} is deemed to be similar to a sample image if there is at least one region of the sample image that is similar to R_{ij} . Further, two regions are considered to be similar if the Euclidean distance between them is below a predefined threshold.

3.4. Progressive RBFN Gradient-Descent Learning

In this study, relevance feedback is implemented as a hybrid learning procedure through unsupervised selection of RBF centers and supervised error correction learning of the underlying parameters (weight and width) of the network. The error function at the t -th feedback iteration is defined as:

$$E_t = \frac{1}{2} \sum_{j=1}^{N_T} e_{jt}^2 = \frac{1}{2} \sum_{j=1}^{N_T} (Y_t(\mathbf{X}_j) - F_t(\mathbf{X}_j))^2 \quad (10)$$

where N_T is the total number of training samples, e_{jt} is the error signal for the j -th training sample \mathbf{X}_j at t -th iteration. $F_t(\mathbf{X}_j)$ and $Y_t(\mathbf{X}_j)$ represent the actual and desired network output for \mathbf{X}_j , respectively. We set $Y_t(\mathbf{X}_j)$ to 1 and 0 for \mathbf{X}_j associated with the relevant and irrelevant feedback, respectively. The RBFN output $F_t(\mathbf{X}_j)$ for an input image \mathbf{X}_j is given as:

$$F_t(\mathbf{X}_j) = \begin{cases} F_{t-1}(\mathbf{X}_j) + \sum_{i=1}^c w_i^t f(\mathbf{X}_j, \mathbf{V}_i^t, \sigma_i^t) & t = 1, 2, \dots, T \\ 0 & t = 0 \end{cases} \quad (11)$$

where $F_{t-1}(\mathbf{X}_j)$ is the $(t-1)$ -th iteration RBFN output,

$U_t(\mathbf{X}_j) = \sum_{i=1}^c w_i^t f(\mathbf{X}_j, \mathbf{V}_i^t, \sigma_i^t)$ is the update network output,

w_i^t is the connection weight of the output layer, \mathbf{V}_i^t , σ_i^t are the center and corresponding width of the i th RBF unit, and c is the number of RBF centers for the update network. $f(\mathbf{X}_j, \mathbf{V}_i^t, \sigma_i^t)$ is the new kernel function of the RBF units, which is proposed for region-based representations with the Euclidean distance replaced by the EMD and given by:

$$f(\mathbf{X}_j, \mathbf{V}_i^t, \sigma_i^t) = \exp\left(-\frac{\text{EMD}^2(\mathbf{X}_j, \mathbf{V}_i^t)}{2(\sigma_i^t)^2}\right) \quad (12)$$

The training is achieved using gradient-descent algorithm by minimizing the cost function E_t to update the network parameters $\boldsymbol{\theta} = \{w_i^t, \sigma_i^t | i = 1, 2, \dots, c\}$:

$$\boldsymbol{\theta} = \arg \min_{\boldsymbol{\theta} \in \Theta} (E_t) = \arg \min_{\boldsymbol{\theta} \in \Theta} \left(\frac{1}{2} \sum_{j=1}^{N_T} (Y_t(\mathbf{X}_j) - F_t(\mathbf{X}_j))^2 \right) \quad (13)$$

where Θ is the solution space of the parametric vector θ . The update equations for w_i^t and σ_i^t are summarized as follows:

(i) Weight estimation at the k -th learning iteration:

$$\frac{\partial E_t(k)}{\partial w_i^t(k)} = -\sum_{j=1}^{N_r} e_{jt}(k) f(\mathbf{X}_j, \mathbf{V}_i^t(k), \sigma_i^t(k)) \quad (14)$$

$$w_i^t(k+1) = w_i^t(k) - \eta_1 \frac{\partial E_t(k)}{\partial w_i^t(k)}, \quad i = 1, 2, \dots, c \quad (15)$$

(ii) Width estimation at the k -th learning iteration:

$$\frac{\partial E_t(k)}{\partial \sigma_i^t(k)} = -w_i^t(k) \sum_{j=1}^{N_r} e_{jt}(k) f(\mathbf{X}_j, \mathbf{V}_i^t(k), \sigma_i^t(k)) \frac{\text{EMD}^2(\mathbf{X}_j, \mathbf{V}_i^t(k))}{(\sigma_i^t(k))^3} \quad (16)$$

$$\sigma_i^t(k+1) = \sigma_i^t(k) - \eta_2 \frac{\partial E_t(k)}{\partial \sigma_i^t(k)}, \quad i = 1, 2, \dots, c \quad (17)$$

(iii) Repeat steps (i)-(ii) until convergence or a maximum number of iterations is reached

The term $e_{jt}(k)$ is the error signal of the j -th training sample \mathbf{X}_j at the k -th learning iteration, η_1 and η_2 are learning parameters for w_i^t and σ_i^t , respectively.

5. EXPERIMENTAL RESULTS

The performance of our framework is evaluated on an image database containing 10,000 color images of 100 different categories obtained from the Corel Gallery product. A total of 100 query images with one from each category are used for evaluation. For each query, the top 25 retrieved images are displayed for feedback. The retrieved images are considered relevant if they belong to the same category as the query image. We compare the proposed RBFN method for RBIR (RRBFN) with the approach using conventional RBFN for global image-based CBIR (GRBFN). Further, the proposed scheme is also compared with another feedback approach called QPM for RBIR [3] (RQPM).

We adopt the following performance measure, called retrieval accuracy:

$$\text{Retrieval accuracy} = \frac{\text{relevant images retrieved in top } T \text{ returns}}{T} \quad (18)$$

where T is the number of retrieved images. The retrieval accuracy comparison is given in Fig.2. It is observed that our method consistently achieves higher retrieval accuracy than the GRBFN and RQPM methods. Further it is observed that to achieve a specific retrieval accuracy, the RRBFN method requires less number of iterations when compared to the GRBFN and RQPM methods. The superiority of our method over the RQPM method mainly lies in the local modeling of image similarity and effective learning to estimate the underlying network parameters. With integration of region-based representation and

effective interactive mechanism, our method can outperform the GRBFN method that employs conventional whole image-based retrieval scheme.

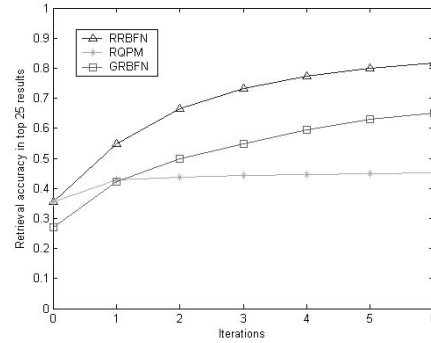


Fig.2. Comparison of retrieval accuracy

6. CONCLUSION

This paper presents an efficient and effective RBFN-based relevance feedback method in interactive RBIR systems. During the feedback iterations, an RBFN is constructed and progressively trained to achieve fast and better retrieval results. Experimental results confirm the effectiveness of our proposed approach.

7. REFERENCES

- [1] Y. Rui, T.S. Huang, M. Ortega, and S. Mehrotra, "Relevance feedback: A power tool for interactive content-based image retrieval," *IEEE Trans. on Circuits and Video Technology*, vol. 8, no. 5, pp. 644-655, 1998.
- [2] J. Z. Wang, J. Li, and G. Weiderhold, "SIMPLcity: semantic-sensitive integrated matching for picture libraries," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 23, no.9, pp.947-963, 2001.
- [3] F. Jing, M.J. Li, H.J. Zhang, and B. Zhang, "Relevance feedback in region-based image retrieval," *IEEE Transactions on Circuits and Systems for Video Technology*, vol.14, no. 5, pp. 672-681, 2004.
- [4] Z. Stejic, Y. Takama, and K. Hirota, "Relevance feedback-based image retrieval interface incorporating region and feature saliency patterns as visualizable image similarity criteria," *IEEE Transactions on Industrial Electronics*, vol. 50, no. 5, pp. 839-852, 2003.
- [5] C. Yang and T. Lozano-Perez, "Image database retrieval with multiple-instance learning techniques," *Proc. 16th Int. Conf. Data Engineering*, San Diego, CA, 2000.
- [6] D. Comaniciu and P. Meer, "Mean shift : A robust approach toward feature space analysis," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no.5, pp.603-619, 2002.
- [7] Y. Rubner, C. Tomasi, and L. Guibas, "The Earth Mover's Distance as a metric for image retrieval," *International Journal of Computer Vision*, Vol. 40, pp.99-123, 2000.
- [8] S. Chiu, "Fuzzy model identification based on cluster estimation," *Journal of Intelligent & Fuzzy Systems*, vol.2, no. 3, 1994.