INTERACTIVE CBIR USING RBF-BASED RELEVANCE FEEDBACK FOR WT/VQ CODED IMAGES

Paisarn Muneesawang[†], *and Ling Guan*[‡]

[†]School of Elec. and Info. Engineering, The University of Sydney, NSW 2006, Australia.
[‡]Dept. of Elec. and Comp. Engineering, Ryerson Polytechnic University, Toronto, Canada

ABSTRACT

Powerful interfaces provide great potential for retrieval systems to adapt to dynamic user needs and allow a more accurate modeling of image similarity from the users' point of view. In this paper, we propose a novel method within the interactive framework. It allows the users to directly modify the system characteristics by specifying their desired image attributes in the form of training samples. More specifically, we have adopted a radial basis function (RBF) method for implementing an adaptive metric which progressively models the notion of image similarity through continual feedback from the users. The proposed approach has been integrated into an image retrieval system using images compressed by wavelet transform and vector quantization coders. Comparisons with some of the recent systems using the standard texture database indicate that the proposed method provides the more favorable retrieval result.

1. INTRODUCTION

The explosion of digital media in diverse applications has created the need for new approaches that help users effectively access and manipulate large quantities of heterogeneous data. Content-based image retrieval (CBIR) has received a great deal of attention in the literature [1]-[5]. Early research in CBIR has been conducted with fully automated systems [1], in which 'index features' are adopted for characterizing image contents including color, texture or shape information. However, these index features have introduced key questions regarding the image retrieval task. This is because the majority of the semantics in an image or user request is lost when we replace its content by a set of features. As a result, the matching of an image to the query features is approximate (or vague). These difficulties have attracted renewed interest in image retrieval. In a number of recent papers [2][3][4], an alternative approach is based on human-computer interaction; the user interactively manages the retrieval system via an interface to extract information needs which are not realized by a one-pass retrieval procedure.

The human-computer interface is understood less well than other aspects of image retrieval although this situation is rapidly changing. This is partly because humans are more complex than computer systems and their motivations and behaviors are more difficult to measure and characterize. Recent studies have been conducted to simulate human perception of visual contents based on a similarity function [4] and by incorporating limited adaptivity in the form of a relevance feedback scheme [2][3], where system characteristics are modified according to the user's preference. However, the limited number of adjustable parameters and the restriction of the distance measure to a quadratic form may not be adequate for modeling perceptual difference as seen from the user's view point.

In this paper a non-linear technique is proposed to address some of the above problems. We adopt a specialized radial-basis function (RBF) method [8] to implement interactive mechanisms for learning the user's notion of *image similarity*. The proposed system allows the user to directly modify retrieval characteristics by specifying desired image attributes in the form of training examples to determine the centers and widths of the different RBFs. In other words, instead of relying on any pre-conceived notion of similarity through the enforcement of a fixed metric for comparison, the concept is adaptively re-defined according to different users' preferences and different types of images. Compared with the previous quadratic measure and the limited adaptivity allowed by its weighted form, the current approach offers an expanded parameter set, in the form of the RBF centers and widths, which allows a more accurate modeling of the notion of similarity from the user's view point.

The proposed approach has been integrated into an image retrieval system using images compressed by wavelet transform (WT) and vector quantization (VQ) coders [6]. Here, image indexing and retrieval are directly performed on the compressed data. This would be advantageous in terms of computational efficiency. In addition, we see that the proposed method has the potential to support the recent development of the compression techniques based on the WT/VQ coders.

2. CODING AND FEATURE EXTRACTION

A compressed-domain approach is highly desirable since it can be applied directly to the compressed streams of images without full decoding [5]. In this paper, we apply the proposed interactive approach to a compressed domain image retrieval. Specifically, the image matching is directly performed on the compressed data after vector quantization of wavelet decomposed images.

A compressed domain feature extraction strategy referred to as a multiresolution-histogram indexing (MHI) [5] is adopted to describe image features in our work. An important characteristic of MHI is that it preserves multiresolution representation of an image. This provides more accu-

racy for image matching. The coding and feature extraction process start with a multiscale pyramid decomposition of an input image. This results in multiresolution sub-images in which horizontal and vertical orientation are considered preferential. The resulting wavelet coefficients in each sub-image are then vector quantized by a multiresolution codebook that contains subcodebooks for each resolution level and preferential direction.

The outcome of the coding process, referred to as coding labels, are used to constitute a feature vector by computation of the labels histograms. Each sub-image is characterized by one histogram. This technique makes use of the fact that the usage of codewords in the subcodebook reflects the content of input sub-image encoded. To minimize the dimension of the feature vector and addressing invariant issues of the illumination level, only five sub-images (two-level decomposition) containing the wavelet detail coefficients are concatenated to obtain the MHI features, $MHI = [H_1, ..., H_i, ...H_5]^T$ where H_i is the histogram obtained from sub-image i.

3. RADIAL BASIS FUNCTION METHOD

Our goal is to establish a non-linear model to simulate human perception for proximity evaluation between images. The non-linear model is an input-output mapping function, $F(X)$, that uses feature values of input image X to evaluate the degree of similarity (according to a given query) by a combination of activation functions associated as a non-linear transformation.

The estimation of the input-output mapping is performed on the basis of a method called *regularization* [7]. In this context, the idea of regularization is based on the *a priori* assumption about the form of the solution (i.e., the input-output mapping function $F(X)$). In its most common form, the mapping function is *smooth*, in the sense that similar inputs correspond to similar outputs. Particularly, the solution function that satisfies this regularization problem is given by the expansion of the radial basis function (RBF) [8].

We adopt the Gaussain shape RBF as a basic model to estimate the input-output mapping function $F(X)$. The one-dimensional Gaussian RBF is associated with each component of the feature vector as follows:

$$F(X) = \sum_{i=1}^{P} G_i(x_i) = \sum_{i=1}^{P} \exp(-\frac{(x_i - z_i)^2}{2\sigma_i^2}) \qquad (1)$$

where $\sigma_i, i = 1, \ldots, P$ are the tuning parameters in the form of RBF widths, $G(\cdot)$ is the Gaussian transformation of the distance between the feature values $\mathbf{x} = [x_1, \ldots, x_i, \ldots, x_P]^T$ and the RBF center $\mathbf{z} = [z_1, \ldots, z_i, \ldots, z_P]^T$.

The activity of the function $G_i(\cdot)$ is to perform Gaussian transformation of the distance $d_i \equiv |x_i - z_i|$, in which its magnitude can be used to describe the similarity between the input x_i and the center of the function: the highest similarity is attained when $G_i(\cdot)$ is equal to unity. Specifically, this function provides a controlled process designed to emphasize some features (relevant ones) and de-emphasize others (non-relevant ones) through proximity evaluation. The transformation process is controlled by an expanded set of tuning parameters $\sigma_i, i = 1, \ldots, P$, which reflects the relevance of individual image features. These parameters are estimated via interactive learning which is described in the next section.

The second expanding parameter offered by the function $G_i(\cdot)$ is the adjustable query $z_i, i = 1, \ldots, P$, in the form of RBF centers. This adjustable center tries to modify the query location in such a way that the new center can enhance the quality of a decision region in the P-dimensional feature space, with regard to distance calculation. The RBF function through the associated RBF centers and widths is designed to perform a system of locally tuned processing units to approximate the target non-linear function $F(X)$ for modeling perceptual similarity.

4. INTERACTIVE LEARNING

We propose a learning method to enable the non-linear function $F(X)$ to progressively model the notion of image similarity through continual relevance feedback from users. This is implemented as an interactive search procedure which uses the information provided by the user to update the parameters of the function.

4.1. Center selection

In an interactive cycle, the user examines the top ranked images and separates them into two classes: the relevant ones $R_i, i = 1, \ldots, N$ and the non-relevant ones $Y_i, i = 1, \ldots, M$. The feature values extracted from these images are then formed as the training set used to select the new feature values for the RBF centers. The idea is to collect information contained in known relevant images. This information is then used to describe a larger cluster of relevant images in the database. In this case, the description of the larger cluster of relevant images is built interactively with assistance from the user.

To attempt to obtain a description for a larger cluster of relevant images, the feature values of the retrieved relevant images $R_i, i = 1, \ldots, N$ are formed as an $N \times P$ feature metric $\mathbf{R}$, and the statistical measuring of feature values in this metric is conducted, where $\mathbf{R} = [x'_{ni}], n = 1, \ldots, N, \quad i = 1, \ldots, P$, and x'_{ni} is the i-th component of the feature vector $\mathbf{x}'_n = [x'_{n1}, \ldots, x'_{ni}, \ldots, x'_{nP}]^T$ corresponding to one of the images marked as relevant. The entries in the i-th column of this metric indicate the possible values that the i-th feature component will take on for a sample set of relevant images. Hence, a suitable statistical measure of values in this sequence should provide a good representation of the i-th feature component. In particular, the mean value of this sequence $\overline{x}'_i = (1/N) \sum_{n=1}^{N} x'_{ni}$ is a good statistical measure since this is the value which minimizes the average distance $(1/N) \sum_{m=1}^{N} (x'_{Ni} - \overline{x}'_i)^2$. As a result, a suitable candidate for the new RBF center is the mean of the set of row vectors in $\mathbf{R}$.

As well as the relevant images, the non-relevant images $Y_i, i = 1, \ldots, M$ imply that their associated vectors have a cluster close to a given query $\mathbf{z}$ according to distance calculation in the previous search operation. In this case, the new RBF center should move away from this cluster in order to avoid the presentation of these non-relevant images in

the next search operation. As a result, the information extracted from the non-relevant images is also included in the process of obtaining the RBF center. The complete formula for updating the RBF center is based on the local clusters of relevant images and nonrelevant images as follows:

$$\mathbf{z}(t+1) = \frac{1}{N}\sum_{n=1}^{N}\mathbf{x}'_n - \alpha_N\left(\frac{1}{M}\sum_{m=1}^{M}\mathbf{x}''_m - \mathbf{z}(t)\right) \quad (2)$$

where $\mathbf{z}(t)$ is the RBF center at the previous iteration, α_N is a suitable positive constant, $\mathbf{x}'_n$ and $\mathbf{x}''_m$ are feature vectors corresponding to the relevant images $R_n, n \in \{1,\ldots,N\}$ and the non-relevant images $Y_m, m \in \{1,\ldots,M\}$, respectively.

The second term of the right-hand side of Eq. (2) tries to shift the new RBF center relatively far from the center of non-relevant samples using the reference point at $\mathbf{z}(t)$ position. This formula allows a more definite movement toward the set of relevant images while permitting slight movement away from the non-relevant regions.

Comparison of the query modification in Eq. (2) to the query reformulation strategy in Eq. (5) shows the following: In the former, the query is modified in the original feature space, while in the latter it is modified in the $term-weighting\ vector$ space. This implies that the problem is tackled in different manners in the two methods.

4.2. Selection of RBF width

It is important that the user's judgment of image similarity can be captured by a small number of pictorial features so that unequal weights exist to dispose of the contribution different features make toward the evaluation of image similarity. That is, given a semantic context, some pictorial features exhibit greater importance or 'relevance' than others in the proximity evaluation. This is the same assumption which underlies image matching algorithms as in Peng et al. [4]. In our case, the estimated tuning parameters can reflect the relevance of individual features. If a feature is highly relevant, the value of σ_i should be small to allow greater sensitivity to any change of the distance $d_i \equiv |x_i - z_i|$. In contrast, a high value of σ_i is assigned to the non-relevant features so that the magnitude of $G_i(\cdot)$ is approximately equal to unity regardless of the distance d_i.

It was proposed in [4] that given a particular numerical value z_i for a component of the query vector, the length of the interval which completely encloses z_i and a predetermined number L of the set of values x'_{ni} in the relevant set is a good indication of the relevancy of the feature. In other words, the relevancy of the i-th feature is related to the density of x'_{ni} around z_i, which is inversely proportional to the length of the interval. A large density usually indicates high relevancy for a particular feature, while a low density implies that the corresponding feature is not critical to the similarity characterization. Setting $L = N$, the set of tuning parameters is thus estimated as follows

$$\sigma_i = \eta \max_n |x'_{ni} - z_i| \quad (3)$$

The factor η guarantees a reasonably large output G_i for each RBF unit, which indicates the degree of similarity, e.g., $\eta = 3$.

We also consider the sample variance in the relevant set for estimating the tuning parameters as:

$$\sigma_i = \exp(\beta \cdot Di) \quad (4)$$

where D_i is the standard deviation of the samples $x'_{ni}, n = 1,\ldots,N$, which is inversely proportional to their density (Gaussian distribution). The parameter β can be chosen to maximize (minimize) the influence of D_i on σ_i. For example, when β is large a change in D_i will be exponentially reflected in σ_i. The exponential relationship is more sensitive to the changes in relevancy and gives rise to better performance improvement, as we shall see in the experiment.

As a result, if the i-th feature is highly relevant (i.e., the sample variance in the relevant set $\{x'_{ni}\}_{n=1}^{N}$, is small), Eqs. (3)-(4) provide a small value of σ_i to allow higher sensitivity to any change in distance. In contrast, a high value of σ_i is assigned to the non-relevant features so that the corresponding vector component can be disregarded when determining the similarity.

5. EXPERIMENTAL RESULTS

In the experiments, the proposed RBF method is compared with some of recently proposed interactive methods using texture database and the MHI feature representation.

For the RBF method, the Euclidean distance based nearest object search is applied for the first search iteration. In the subsequent search operation, the RBF method based on Eqs. (1)-(3) is applied for discriminating image similarity. This method is denoted as RBF1. To study how the performance of the RBF changes when using different criterion for obtaining RBF widths, Eq. (4) is applied for tuning RBF parameters. This method is denoted as RBF2.

Comparisons are made with the conventional relevance feedback approaches used in the MARS [2] and PicToSeek [3] systems. In MARS system [2], a tf $\times$idf factor is applied for conversion of image features to weighted vectors, before using relevance feedback to revise the weights of the original query. The new query weights are updated according to the following formula:

$$\mathbf{z}^{(new)} = \alpha\mathbf{z}^{(ori)} + \frac{\gamma}{M}\sum_{n\in D_R}\mathbf{x}_n - \frac{\epsilon}{Q}\sum_{n\in D_{IR}}\mathbf{x}_n \quad (5)$$

where $\mathbf{z}^{(ori)}$ represents the original query, D_R and D_{IR} represent the set of relevant and non-relevant images respectively. The similarity between weighted vectors is computed using the cosine measure. The main idea of query reformulation strategy (Eq. (5)) consists of selecting important $terms$ (i.e., important features) of the relevant images, and enhancing the importance of these terms in a new query formulation. In other words, all $terms$ appearing in the relevant images are considered as important and added to the original query, whereas the ones appearing in the non-relevant images are considered as insignificant and deleted from the original query.

In PicToSeek [3], weighting of image features is based on another tf $\times$ idf factor (which is different from the one used by MARS). Histogram intersection is employed as a similarity function in the first search operation. During the

interactive searching, the query's weights are modified by the query reformulation strategy in Eq. (5).

The image database was obtained from MIT Media Laboratories as in [2] and [4]. It consisted of 624 texture images that were classified into 39 different classes, where each class contained 16 similar images. The MHI feature representation was applied to the texture images in the database, where it was performed with simultaneous WT/VQ coding. Each image was decomposed using the 15-tab biorthogonal filters, with two-level decomposition. The transform coefficients were vector quantized, using the multiresolution codebook at the total bit rate of 1 bpp [6]. During the VQ of the coefficients, the MHI was obtained by recording the usage of codewords. Each image was represented by a MHI vector and a set of labels.

A total of 39 images, one from each class, were selected as the query images. For each query, the top 16 images were retrieved to provide necessary relevance feedback. The performance was measured in terms of average retrieval rate [1] which was defined as the average percentage number of images belonging to the same class as the query in the top 16 matches. Note that, in the simulations, we used a known "ground truth" to introduce the user feedback.

The average retrieval rate of the 39 query images is summarized in Table 1. In the table, t denotes the number of iterations. The following main trends can be observed from the results. Firstly, for all methods, the performance with interactive learning after 3 iterations (t=3) was substantially better than the non-interactive case (t=0), i.e., more than 45% improvement. In particular, the greatest improvement is achieved in the first iteration. These results also imply that with or without learning, the MHI feature provides a very good representation in retrieving all other 15 images from the same class.

Secondly, comparing the results after learning, we observed that the proposed method RBF2 gave the best performance: on average 92.6% of the correct images are in the top 16 retrieved images (i.e., more than 14 correct images were present out of 16). This is closely followed by RBF1 at 90.55%. The MARS system performed slightly better (85.4%) than the PicToSeek system (84.8%). This result implies that RBF2 consistently displays superior performance over the other retrieval systems discussed.

NOTE: The procedural parameters input to all methods reported in the above results were determined empirically. The parameters that gave the best retrieval results are listed at the bottom of Table 1. The parameters $(\alpha, \gamma, \epsilon)$ in Eq. (5) input to MARS and PicToSeek were determined according to the standard formula studied by Richio [9]. The constant α is fixed to 1 and the constants γ, ϵ are varied to obtain the best retrieval results.

6. CONCLUSION

An interactive approach for content-based image retrieval is proposed and its application to WT/VQ coded images is demonstrated. More specifically, we propose the adoption of a non-linear RBF for characterizing the behavior of human users in an interactive retrieval session where relevance feedback is applied. Image matching is directly performed on the compressed domain using MHI feature representation. As demonstrated by the experimental results, the combination of MHI features and interactive learning significantly enhance the retrieval performance. A comparison with some well known conventional relevance feedback methods indicates that the proposed approach is robust and favorable over preceding methods.

Method	t=0	t=1	t=2	t=3
RBF1	63.46	82.85	88.62	90.55
RBF2	63.46	83.65	89.74	92.63
MARS	63.94	80.77	84.46	85.42
PicToSeek	62.18	77.08	83.97	84.87

Table 1: Average retrieval rate (%) for the 39 queries images in the MIT database; comparison of the performance with and without learning similarity using MHI feature representation. NOTE: the procedural parameters are; RBF1: $\alpha_N = 0.7$, RBF2: $\alpha_N = 0.7$, $\beta = 0.8$, MARS: $(\alpha, \gamma, \varepsilon) = (1, 8, 3)$, PicToSeek: $(\alpha, \gamma, \varepsilon) = (1, 8, 3)$

7. REFERENCES

[1] B. S. Manjunath and W. Y. Ma, "Texture features for browsing and retrieval of image data," *IEEE Trans. of Pattern Analysis And Machine Intelligence*, August, vol. 18, no. 8, pp. 837-842, 1996.

[2] Y.Rui, T.S. Huang, and S.Mehrotra, "Content-based image retrieval with relevance feedback in MARS," *Proc. IEEE Int. Conf. on Image Processing*, pp. 815-818, 1997.

[3] T.Gevers and A.W.M.Smeulders, "PicToSeek: combining color and shape invariant features for image retrieval," *IEEE Trans. on Image Proc.*, vol.9, no.1, pp.102-119, 2000.

[4] J. Peng, B. Bhanu, and S. Qing, "Probabilistic feature relevance learning for content-based image retrieval," *Computer Vision and Image Understanding*, vol.75, nos.1/2, pp. 150-164, 1999.

[5] P. Muneesawang and L.Guan, "Multiresolution-histogram indexing for wavelet-compressed images and relevant feedback learning for image retrieval," *Proc. IEEE Int. Conf. on Image Processing*, 2000.

[6] N. B. Karayiannis, P.-I. Pai, and N.Zervos, "Image compression based on fuzzy algorithms for learning vector quantization and wavelet image decomposition," *IEEE Trans. on Image proc.*, vol.7, no.8, pp.1223-1230, 1998.

[7] A. N. Tikhonov, "On solving incorrectly posed problems and method of regularization," In S. Haykin editor, *Neural Networks*, Prentice Hall, 1999.

[8] T. Sigitani, Y. Liguni, H. Maeda, "Image interpolation for progressive transmission by using radial basis function networks," *IEEE Trans. on Neural Networks*, vol.10, no. 2, pp. 381-390, 1999.

[9] J.J. Richio, "Relevance feedback in information retrieval," In G. Salton, editor, *The SMART Retrieval System – Experiments in Automatic Document Processing*, Prentice Hall Inc., Englewood Cliffs, NJ, 1971.