Utilizing Inter- and Intra-Query Relevance Feedback for Content-Based Image Retrieval

Chi-Hang Chan, Ka-Cheung Sia, and Irwin King
Department of Computer Science and Engineering
The Chinese University of Hong Kong
Shatin, N.T., Hong Kong SAR, China
{chchan, kcsia, king}@cse.cuhk.edu.hk

Abstract— **The relevance feedback approach is a powerful technique in content-based image retrieval (CBIR) tasks. Many parametric and non-parametric estimation approaches to refine user's query that are based on low-level image features have been proposed. Most of them often fail when the underlying distribution of a user's target is not clustered in the low-level feature vector space. To address this, researchers have proposed the use of long-term relevance information to assist in discovering the semantics of images in retrieval process. However, these methods do not work properly when the number of query results is small and most of them are not concerned about the subjectivity of different users. In this paper, we propose a SOM-based technique to construct a vector space which represents the similarities of images under human perception from long-term relevance information. The target distribution of user's query is then estimated on the newly constructed vector space. Experiments indicate that retrieval performance is increased when a parameter estimation approach is used on the newly formed vector spaced after learning the inter-query relevance feedback information.**

I. Introduction

With the rapid growth in the volume of digit images, searching and browsing in a large collection of unannotated images is gaining importance. CBIR systems use low-level feature extraction methods, such as color, texture and shape, to index and search for similar images from a database. Previous systems are based on the one-shot approach, in which images with the smallest distance to the query in the feature vector space are retrieved. However, this approach is difficult to capture user's need in the retrieval process.

The goal of relevance feedback is to learn user's preference from their interaction, and it is a powerful technique to improve the retrieval result in CBIR. Under this framework, a set of images is presented to the user according to the query. The user marks those images as either relevant or non-relevant and then feeds back this into the system. Based on this feedback, the system estimates user's preference through this learning process.

Most of the current relevance feedback systems are based on the intra-query approach. In this approach, the system refine the query by using feedback information that the user provided. This learning process starts from ground up for each query. For example, *PicHunter* [1] presented a Bayesian framework to direct a search with relevance feedback information. Each image in the database is associated with a probability that it is the target image of user's query. The Bayes's rule is used to update the probability according to user's feedback information in every iteration. Rui et al. [8] proposed a weight updating approach to capture user's preference on different feature representations of images. The weight of each feature, its representation and each dimension is updated by their discriminant power between the set of relevant and non-relevant images in the current query. The similarity measure of images is based on the feature representations and their weights. The intra-query approach uses the feedback information to estimate user's target distribution, but it often fails when the underlying distribution is not clustered in the low-level feature space. This is due to the feedback information given by a single query is limited and unable to provide enough statistical information about the distribution [10].

Recent research propose the use of inter-query information to further improve retrieval result [5] [4]. The relevance feedback besides providing information about user's target distribution, but also a similarity measure of images under human perception. In this approach, feedback information from individual users are accumulated to train the system to determine what images are of the same semantic meaning. Heisterkamp [5] and He et al. [4] applied inter-query information and latent semantic indexing (LSI) [2] to CBIR. LSI is a classical document retrieval algorithm. It analyzes the correlation of documents and terms in the database. In this approach, previous feedback information are stored in the system to build the latent semantic index. Each query result is considered as a document and each image in the database is consider as a term, thus the correlation of image and different semantic meaning is revealed. However, these two systems can only retrieve images which have been marked as relevant in the results of query. Thus, the number of query results required is large in order to make sure all images can be retrieved. Moreover, various inter-query approachs do not concern about the subjectivity of different users, and the same initial query point always provide the same query result.

In this paper, we propose to use both inter-query and intra-query information for modifying the feature vector space and estimating user's target distribution. Self-Organizing Map (SOM) [7] is used to cluster and index the images in the database. We propose to use the inter-query information to modify the feature vector space, in which the SOM of images is stored. This allows for transforming the images distributions and improving the similarity measure. Moreover, We present an Expectation Maximization (EM) [3] [9] approach to estimate user's target distribution on the modified feature vector space.

II. The Basic Framework

A. Image Object Model

We perform a low-level feature extraction on the set of images $\{I_i|1 \leq i \leq n\}$ in the database, and each image is then represented by a feature vector in a high dimensional vector space. We construct and train a SOM M with feature vectors extracted from the images. After the training, the model vectors stored in the neurons of M are arranged to match the distribution of the feature space. The model vectors stored in the neurons are used to partition the feature space; as a result, images are grouped into different neurons according to a minimum distance classifier.

B. Inter-Query Feedback

In various relevance feedback systems, only the intra-query feedback information is used to estimate the user's target distribution. However, a small training data set is difficult to provide enough statistical information for estimating the underlying distribution and providing good retrieval result. If the form of underlying density is known, the parameters of the density can be estimated classically by maximum likelihood. However, the underlying distribution is often not clustered and difficult to assume it to follow any particular form of density.

In order to address the above difficulties, we use the inter-query information to modify the feature vector space and organize the neurons into Gaussian-like distributions. Thus, an prior form of density can be assumed for user's target distribution in the modified feature vector space. The key idea in inter-query feedback is that a user's feedback is not only providing information for optimizing his own query, but also similarity measure between images in the database in the sense of human perception. In the proposed approach, we update the similarity measure between images dynamically according to the feedback information given by each query. It is achieved by further training the neurons on the SOM. Neurons represent relevant images are moved closer to the estimated user target and those represent non-relevant images are moved away from the estimated user target.

Figure 1 shows a two-dimensional feature vector space of a collection of images with 4 different classes. A SOM is trained based on the underlying distribution. In analyzing the image data, images from the same class often form clusters which are sparse and irregular in shape. This makes the retrieval process more difficult to find target images. With the help of inter-query feedback information described above, we organize the feature vector space in a fashion that ease the retrieval process.

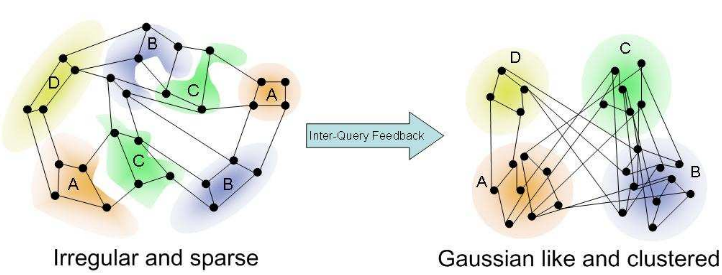

Fig. 1. Inter-query feedback illustration

C. Intra-Query Feedback

Most of the current inter-query relevance feedback systems use feedback information to refine or modify the similarity measure between images only. However, the target distribution may be different for different users even if their query points are the same. In order to address this, we use a intra-query feedback approach to estimate the user's target distribution, but the process is performed on the SOM in the modified vector space instead of the original feature vector space.

The intra-query relevance feedback process is divided into two phases, one is learning and the other one is display selection. The estimation is based on EM algorithm, the computer try to figure out a distribution representing user's target based on his feedback given. The display selection part makes use of most informative display. As the computer already obtained a model describing the distribution of user's target, it selects the data located along the boundary of this model for display in order to get maximum amount of information from user's feedback. The derivation of estimation and display selection is described in [9], we present a modified version here to deal with the SOM trained image data.

III. Algorithm

A. Initialization

Given the feature vectors of all the images in the database, we train the SOM with these feature vectors and the trained SOM represents the underlying distribution of these feature vectors. The model vectors $\vec{m}_i \in M$ of neurons in the SOM are used to partition the feature vector space based on the minimum distance classifier, each image I_i is classified into different groups represented by m_k. By doing so, we reduce the size of data from $|I|$ to $|M|$, where $|I|$ and $|M|$ are the number of images and neurons in the SOM respectively. The similarity measure between images is then defined as the Euclidean distance between the model vectors which represent them.

B. SOM Duplication

The relationship between the neurons and the images in the database depends on the coordinates of the model vectors, any changes on the model vectors of neurons may alter this relationship. Our proposed approach is to modify the model vectors in the SOM to update the similarity measure. Thus, we duplicate another SOM from the original one. The new SOM contain a set of neurons with model vectors $\vec{m}'_i \in M'$ and has a one-to-one mapping, $f : M \to M'$, between the set M and M'. To obtain the set of images represented by model vector $\vec{m}'_i$, we can get the original model vector $\vec{m}_i$ by f^{-1}, and then by minimum distance classifier in M. Initially, the layout of the two SOMs are the same, that is

$$\begin{aligned} &\forall \vec{m}' \in M', \forall \vec{m} \in M \\ &\qquad \vec{m}' = f(\vec{m}) \quad \Rightarrow \quad \vec{m}' = \vec{m}. \end{aligned} \tag{1}$$

We update the similarity measure by modify the model vectors in M' instead of M, so that the relationship between images in the database and the model vectors in M can be preserved during the whole learning process.

C. Similarity Measure Updating

In order to update the similarity measure based on the inter-query feedback information, we modify the model vectors $\vec{m}'_i$ in the new SOM, such that neurons contain similar images as indicated in the feedback are moved closer to each others. In inter-query feedback learning, we consider each query by an user as an iteration in a learning process. Assume in the k^{th} iteration, the user marked a set of relevant images I_R and a set of non-relevant images I_N during the whole retrieval process, M'_R and M'_N are the corresponding sets of model vectors respectively. Let $c(k)$ be the vector has the largest probability to be the user's target in the vector space of M', and it is defined by

$$\vec{c} = \arg\max_{\vec{v}} P(\vec{v}|\Theta), \tag{2}$$

where $P(\vec{v}|\Theta)$ is the probability function of $\vec{v}$ to be the user's target and it is described in section III-D. We modify the model vectors with the following equations,

$$\begin{aligned} &\forall \vec{m}'_i \in M'_R \\ &\vec{m}'_i(k+1) = \vec{m}'_i(k) + \alpha_R(k)[\vec{c}(k) - \vec{m}'_i(k)], \quad (3) \\ &\forall \vec{m}'_i \in M'_N \\ &\vec{m}'_i(k+1) = \vec{m}'_i(k) + \alpha_N(k)[\vec{m}'_i(k) - \vec{c}(k)], \quad (4) \end{aligned}$$

where $\alpha_R(k)$ and $\alpha_N(k)$ are the learning rates and they are monotonic decreasing functions. Thus, neurons represent relevant images are moved closer to the estimated user's target and those represent non-relevant images are moved away from the estimated user's target. For a long run, the vector space will be modified, in which neurons represent different image classes are organized as separated Gaussian-like clusters.

In a SOM, the nearby neurons in the topology are representing similar units, so that the learning process can be improved by moving also the neurons that near to the neurons in the sets M_R and M_N. Thus, the equations for modifying the model vectors are defined by,

$$\begin{aligned} &\forall \vec{m}'_i \in N(M'_R) \\ &\vec{m}'_i(k+1) = \vec{m}'_i(k) + h_{Ri}(k)[\vec{c}(k) - \vec{m}'_i(k)], \quad (5) \\ &\forall \vec{m}'_i \in N(M'_N) \\ &\vec{m}'_i(k+1) = \vec{m}'_i(k) + h_{Ni}(k)[\vec{m}'_i(k) - \vec{c}(k)], \quad (6) \end{aligned}$$

where $N(M)$ is the set of nearby neurons for M, $h_{Rci}(k)$ and $h_{Nci}(k)$ are the neighborhood functions. The neighborhood functions are defined by

$$h_{Ri}(k) = \alpha_R(k) \cdot \exp\left(-\frac{\mathrm{dis}(\vec{m}'_i(k), M'_R)}{2\sigma_R^2(k)}\right), \text{ and} \tag{7}$$

$$h_{Ni}(k) = \alpha_N(k) \cdot \exp\left(-\frac{\mathrm{dis}(\vec{m}'_i(k), M'_N)}{2\sigma_N^2(k)}\right), \tag{8}$$

where $\sigma_R(k)$ and $\sigma_N(k)$ are some monotonic decreasing functions, and dis($\vec{m}'_i(k)$,M) denote the distance between the model vector $\vec{m}'_i(k)$ and the set M in the SOM topology.

D. User's Target Estimation

In the intra-query learning process, the system presents a set of images I_t to the user in each iteration, and the user marks them as either relevant or non-relevant. The system uses the set of relevant images I_{Rt} and the set of non-relevant images I_{Nt} to refine the query. Since the distribution of similar neurons in the vector space of M' is more or less follow Gaussian distribution, we perform the user's target estimation on it instead of the feature vector space. We define M'_R as the set of relevant model vectors in M', and we use the EM-based approach proposed in [9] to estimate user's target distribution. Since the most ambiguous images, that will be discussed in III-E, are chosen to present to the user. The relevant model vectors should be the boundary cases in the target distribution. Thus, we maximize the probability that model vectors are located in the boundary region, we illustrate this idea in Fig. 2. This is achieved by assuming no correlation between dimensions and maximizing the following equation,

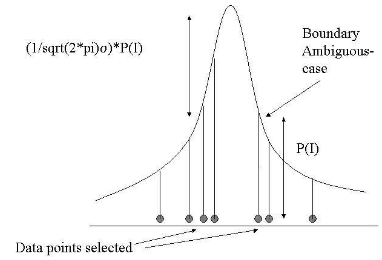

Fig. 2. Fitting μ and σ for data points selected to display

$$L(\mu_j, \delta_j) = \sum_{\vec{m}'_i \in M'_R} \sum_{j=1}^{d} P(m'_{ij}|\mu_j, \delta_j) \times \left(\frac{1}{\sqrt{2\pi}\delta_j} - P(m'_{ij}|\mu_j, \delta_j)\right), \tag{9}$$

$$P(m'_{ij}|\mu_j, \delta_j) = \frac{1}{\sqrt{2\pi}\delta_j} \exp^{-\frac{(m_{ij}-\mu_j)^2}{2\delta_j^2}}, \tag{10}$$

where j is the subscript for dimension and $P(.)$ is a Gaussian density function for a model vector to be the user's target in a particular dimension. By differentiate the function $L(\mu_j, \delta_j)$, we update the parameters by the following equations,

$$\mu_j = \frac{\sum_{\vec{m}'_i \in M'_R} m'_{ij}}{|M'_R|}, \tag{11}$$

$$\delta_j^2 = \frac{\sum_{\vec{m}'_i \in M'_R} ((m'_{ij} - \mu_j)^2 2^{\frac{1}{4}} \pi^{-\frac{3}{4}} - \delta_{old_j}^{\frac{1}{2}} P_{old}^{\frac{1}{2}} (m'_{ij} - \mu_j)^2 2^{\frac{1}{2}} \pi^{-\frac{1}{2}})}{\sum_{\vec{m}'_i \in M'_R} (2^{\frac{1}{4}} \pi^{-\frac{3}{4}} - \delta_{old_j}^{\frac{1}{2}} P_{old}^{\frac{1}{2}} 2^{\frac{1}{2}} \pi^{-1})}. \tag{12}$$

E. Images Display Selection

In order to maximize the user's distinguishing power, we choose the most ambiguous images to present to the user, and acquire the feedback information from the user. This images selection strategy is analogous to the maximum entropy display [6]. Specifically, we choose the model vectors that are $\pm k\delta$ away from the μ such that $P(\mu \pm k\delta) = \frac{1}{\sqrt{2\pi}\delta} - P(\mu \pm k\delta)$, and we use the function f^{-1} to obtain the corresponding images. Thus, images which are best for identifying user's target distribution are selected.

IV. Experiment

We perform the experiment on the Corel image collection, which contains 40,000 images in different categories. Among the 400 categories, we selected 40 categories, each contains 100 images and ranges from the category of **buildings**, **protait**, **outdoor scenary**, etc. We use the default groupings of images as the ground truth and human knowledge to run automated tests. We use *Color Moment* and *Cooccurence Matrix* as the image features. *Color Moment* computes the mean, variance and skewness values of each color channel in the image. *Cooccurence Matrix* describes the texture of an image by measuring the occurrence of its graylevel configurations.

The feature vectors of images are first extracted and normalized, then it is used to train a SOM structure of dimension 18×18. Queries are generated and evenly distributed among the 40 classes, while the relevance feedbacks are generated based on the ground truth. All displayed images are marked relevant if they are in the same class as the query, and the others are marked non-relevant. The experiment is divided into 3 stages. In the first stage, 80 queries are generated to find out the average recall and precision when using the intra-weight updating version of Rui's approach and the intra-query approach in this paper. In the second stage, a number of queries are generated and our SOM-based approach is used to retrieve images and train the system. In the final stage, 80 queries are generated again to find out the average recall and precision of our intra-query approach after inter-query feedback is applied. Table I shows the parameters used in this experiment. We then compare the result of Rui's intra-weight approach and the intra-query approach before and after our SOM-based inter-query feedback training.

TABLE I
Parameters used in experiment

Parameter	Value
Number of images	4,000
Number of categories	40
Number of iterations used in inter-query feedback	300,500
Number of iterations used in intra-query feedback	9
Ratio of push and pull	20
Feature used	Color Moment (9-dimensions) Cooccurence Matrix (20-dimensions)

Figure 3 shows the Recall-Precision graph, averaged among 80 queries, of Rui's approach and our approach before and after the SOM-based training. We make two observations from the experiment. The first one is the intra-query approach performs better after the SOM-based inter-query feedback training. It shows that the retrieval precision can be improved by re-organizing the feature vector space with SOM. The second one is Rui's approach may perform better than our intra-query approach initially, but when more relevant images are retrieved, our performance on precision becomes better. It is because Rui's approach is better in estimating the most relevant image and our approach is better in estimating the category of relevant images. Moreover, the experiment indicates that the improvement is sensitive to the push-pull value, their ratio and the number of iterations used in inter-query feedback.

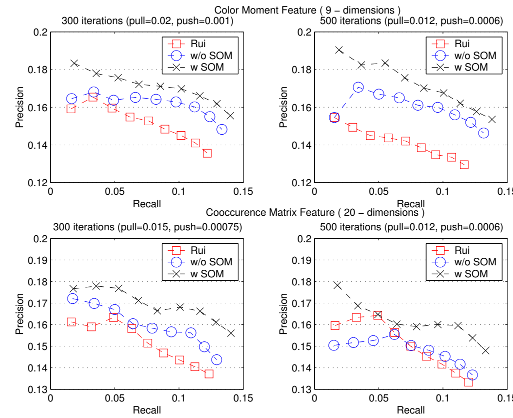

Fig. 3. Recall-Precision graph

V. Conclusion

The distribution of image classes in low-level feature space is not well clustered, in which user's target distribution is difficult to estimate. In this paper, we propose a SOM-based approach for re-organizing the feature space of images using inter-query feedback information. As a result, the distribution of similar images more or less follows a Gaussian-like distribution which is more efficient for estimation. We also demonstrate improvement in retrieval accuracy through experiments.

References

[1] Ingemar J. and Cox. The Bayesian Image Retrieval System, PicHunter, Theory, Implementation, and Psychophysical Experiments. *IEEE Transactions on Image Processing*, 9(20-37), January 2000.

[2] Scott C. Deerwester, Susan T. Dumais, Thomas K. Landauer, George W. Furnas, and Richard A. Harshman. Indexing by Latent Semantic Analysis. *Journal of the American Society of Information Science*, 41(6):391–407, 1990.

[3] A.P. Dempster, N.M. Laird, and Rubin D.B. Maximum Likelihood from Incomplete Data via the EM Algorithm. *J. R. Statist. Soc. B*, 39:185–197, 1977.

[4] Xiaofei He, Oliver King, Wei-Ying Ma, Mingjing Li, and Hong-Jiang Zhang. Learning a Semantic Space From User's Relevance Feedback for Image Retrieval. *IEEE Transactions on Circuits and Systems for Video Technology*, 13(1), January 2003.

[5] Douglas R. Heisterkamp. Building a Latent Semantic Index of an Image Database from Patterns of Relevance Feedback. In *Proceedings of the 16th International Conference on Pattern Recognition (ICPR 2002)*, pages 134–137, August 2002.

[6] Irwin King and Zhong Jin. Relevance Feedback Content-Based Image Retrieval Using Query Distribution Estimation Based on Maximum Entropy Principle. In Liming Zhang and Fanji Gu, editors, *Proceedings to the International Conference on Neural Information Processing (ICONIP2001)*, volume 2, pages 699–704, Shanghai, China, November 14-18 2001. Fudan University, Fudan University Press.

[7] Teuvo Kohonen. *Self-Organizing Maps*, volume 30 of *Springer Series in Information Sciences*. Springer, third edition, 2001.

[8] Yong Rui, Thomas S. Huang, Michael Ortega, and Sharad Mehrota. Relevance Feedback: A Power Tool for Interactive Content-Based Image Retrieval. *IEEE Transactions on Circuits and Video Technology*, 8(644-655), September 1998.

[9] K. C. Sia and Irwin King. Relevance Feedback Based on Parameter Estimation of Target Distribution. In *Proceedings of IEEE International Joint Conference on Neural Networks*, pages 1974–1979, May 2002.

[10] Y. Wu, Q. Tian, and T.S. Huang. Discriminant-EM Algorithm with Application to Image Retrieval. In *Proceedings to the IEEE Conference on Computer Vision and Pattern Recognition*, volume 1, pages 222–227, June 2000.