Improving Retrieval Performance by Long-term Relevance Information

Peng-Yeng Yin[1], Bir Bhanu[2], Kuang-Cheng Chang[1] and Anlei Dong[2]
[1]Department of Computer Science, Ming Chuan University, Taoyuan 333, Taiwan
[2]Center for Research in Intelligent Systems, University of California, Riverside, California 92521, USA
{pyyin, kcchang}@mcu.edu.tw, {bhanu, adong}@vislab.ee.ucr.edu

Abstract

Relevance feedback (RF) is an iterative process which improves the retrieval performance by utilizing the user's feedback on retrieved results. Traditional RF techniques use solely the short-term experience and are short of knowledge of cross-session agreement. In this paper, we propose a novel RF framework which facilitates the combination of short-term and long-term experiences by integrating the traditional methods and a new technique called the virtual feature. The feedback history of all the users is digested by the system and is represented as a virtual feature of the images. As such, the dissimilarity measure can be adapted dynamically depending on the estimate of the relevance probability derived from the virtual features. The results manifest that the proposed framework outperforms the one that adopts a single traditional RF technique.

1. Introduction

Since the users in general do not know the make-up of the image database and the techniques used for indexing, the query formulation process should be treated as a series of tentative trials until the target images are found. Relevance feedback (RF) is an automatic process which fulfills the requirements of the query formulation.

Let a user initialize a query session by submitting an image represented by $Q=(q_1,q_2,...,q_t)$ where t is the number of selected features and q_i is the calculated value of the ith feature. The retrieval system compares the query image with each of the database images, say $D=(d_1,d_2,...,d_t)$, by deriving the dissimilarity measure $Dist(Q,D)$. The top k database images that have the smallest dissimilarity score are then returned to the user. If the user is not satisfied with the retrieved result, he or she can activate an RF process by identifying which retrieved images are relevant to the query and which are not. The system will adapt its internal parameters to involve as many desirable images as possible in the next retrieved result. The process is repeated until the user is satisfied with the retrievals. The general system flow chart of the RF process is depicted in Fig. 1. In the following, the main RF techniques for image retrieval are presented.

The query vector modification (QVM) approach [1] repeatedly reformulates the query vector as the mean difference vector between relevant images and nonrelevant ones, in the attempt to redirect the query vector toward a more desired area. The feature relevance estimation (FRE) approach [2] assumes, for a given query, some specific features may be more important than others when computing the similarities between images and the query. The most natural way of estimating the individual feature relevance is to verify the retrieval ability using each feature alone. Finally the feature relevance is used as a weight incorporated into the dissimilarity metric. The Bayesian inference-based (BI) approach [3] estimates the posterior probability that a database image is relevant to the query given the prior feedback history. The probability distribution over all database images is updated after each feedback iteration, the system is therefore able to improve the future retrieval performance.

These methods suffer their respective shortcomings. First, the QVM put equal emphasis on every relevant image by averaging their feature vectors, however, not every relevant image has the same magnitude of relevance. Second, the success of both QVM and FRE is based on the assumption that the distributions of the feature vectors of the relevant images form an intrinsic cluster. Whereas, no matter how sophisticated features are selected, they are insufficient to fully represent the image semantics, and the relevant images will usually do not form a single cluster. Third, without storing the

1051-4651/02 $17.00 © 2002 IEEE

relevance information directly, some information is lost such as the relevance significance of each individual image. The BI approach is theoretically the most flexible one since it does not rely on the nearest neighbor criterion. However, the BI approach needs more feedback iterations to accurately approximate the posterior probability distribution. So it is less efficient than the other RF techniques.

Moreover, all the three kinds of RF approaches improve the retrieval performance based on the feedback history within one query session. Hence, the previous approaches maintain a form of short-term memory that captures the user's intention for only this specific query. There is no consideration being taken for the cross-session feedback history, which is a form of long-term memory that captures the common agreement among various query sessions. The long-term memory is useful in leading the feedback process to converge at an earlier iteration.

2. The Proposed Approach

To digest the relevance information accumulated from within- or cross-session query experiences, we add a virtual feature (VF) to the feature vector of each of the database images. The VF is determined by the set of relevant images and is used to assist the original pictorial features to evaluate the similarity degree between images in accordance with the human subject. The details of the proposed approach are presented as follows.

2.1 Virtual Feature Computation

Given a query $Q=(q_1,q_2,\ldots,q_t)$, the retrieval system firstly searches the top k nearest images using the dissimilarity metric of the adopted short-term RF technique. If the user is not satisfied with the result, he would activate an RF process by identifying relevant and nonrelevant images. We denote by R the set of identified relevant images. Initially, the VF of each database image is empty. Each relevant image $D=(d_1,d_2,\ldots,d_t)$ in R will derive its VF by requesting a number from a system counter. The system counter starts counting from 1 and is increased by 1 after every time it is requested. Therefore, all the images in R will be assigned the same value as their VFs to mark that these images deliver the same concept possessed by the query.

As the feedback process repeated, one case may arise that some of the images in R have been already assigned the VFs. The relevant images that have not determined their VFs yet will be firstly given a number from the system counter and this number is then concatenated by the VFs of the other relevant images and converted into a canonical form. We define the canonical form $\Im$ with the set of positive integers Z^+ and a concatenation operator $\otimes$ as follows. The concatenation operator $\otimes$ is defined by

$$c_i^{e_i} \otimes c_j^{e_j} = \begin{cases} c_i^{e_i} \otimes c_j^{e_j} & \text{if } c_i < c_j, \\ c_j^{e_j} \otimes c_i^{e_i} & \text{if } c_i > c_j, \\ c_i^{e_i+e_j} & \text{if } c_i = c_j, \end{cases}$$

where c_i, c_j, e_i and e_j are in Z^+. An expression f is in canonical form $\Im$ if $f = c_1^{e_1} \otimes c_2^{e_2} \otimes \cdots \otimes c_m^{e_m}$, where $c_i < c_j$ if $i < j$. Apparently, canonical form $\Im$ holds a closure property. Formally, $f_1, f_2 \in \Im \rightarrow f_1 \otimes f_2 \in \Im$. The closure property guarantees that the VF yielded by concatenating several VFs is still in the canonical form $\Im$.

In this way, each value in the VFs represents a relevance concept impressed by a certain user, and the system can digest multiple concepts of image relevance in the VFs. To estimate the relevance between the query and database images, the VF of the query is computed as the concatenation of the VFs of all images in R which are specified in the previous feedback iteration, i.e., $VF(Q)=VF(D_1)\otimes VF(D_2)\otimes\cdots\otimes VF(D_{|R|})$, $D_i \in R$, where $VF(\cdot)$ denotes the VF of the corresponding image. The VFs of the query and the database images are used to define the dynamic dissimilarity measure which will be discussed in the next subsection.

2.2 Probabilistic dissimilarity measure

Let the VF of an image D be $c_1^{e_1} \otimes c_2^{e_2} \otimes \cdots \otimes c_m^{e_m}$, we firstly define the concept set of image D as $C(D)=\{c_1,c_2,\ldots,c_m\}$, each concept c_i is associated with a support value e_i. The larger the cardinality of the concept set, the more general the overall concept delivered by the image. Also, the larger the support value, the more important to the image the corresponding concept. We define the probability that D is semantically recognized as concept c_i, or the confidence that D is delivering concept c_i, given $VF(D)$ as

$$p(D \quad c_i | VF(D)) = \frac{e_i}{\sum_{j=1}^{m} e_j}.$$

Assume the two events that the concepts delivered by the query and by the database image are independent given their VFs. The probability, denoted by $p_{Q \equiv D|VFs}$, that the query Q and the database image D are delivering the same concept given their VFs is calculated by

$$p_{Q\equiv D|VFs} = \underset{c_i \in C(Q) \cap C(D)}{} p(Q \quad c_i \text{ and } D \quad c_i | VF(Q), VF(D))$$

$$= \underset{c_i \in C(Q) \cap C(D)}{} p(Q \quad c_i | VF(Q))\, p(D \quad c_i | VF(D)) \cdot$$

Based on the probability estimate, we define a dynamic dissimilarity measure as $Dist_{VF}(Q,D) = p_{Q\equiv D|VFs}(Dist(Q,D) - \Delta) + (1 - p_{Q\equiv D|VFs})(Dist(Q,D) + \Delta)$, if both $VF(Q)$ and $VF(D)$ are known, and $Dist_{VF}(Q,D) = Dist(Q,D)$ otherwise, where Δ is the quantity of the maximal distance adjustment, and $Dist(Q,D)$ is the distance metric defined by the short-term RF technique incorporated into our approach. The first equality can be rewritten as $Dist_{VF}(Q,D) = (1 - 2p_{Q\equiv D|VFs})\Delta + Dist(Q,D)$. It is observed that $Dist_{VF}(Q,D) < Dist(Q,D)$ if $p_{Q\equiv D|VFs} > 0.5$, $Dist_{VF}(Q,D) > Dist(Q,D)$ if $p_{Q\equiv D|VFs} < 0.5$, and $Dist_{VF}(Q,D) = Dist(Q,D)$ if $p_{Q\equiv D|VFs} = 0.5$. Therefore, the proposed method dynamically adjusts the distance between the query and the database images based on the estimate of $p_{Q\equiv D|VFs}$ which is derived from the long-term feedback history.

Compared with the existing RF techniques, the proposed method has the following features.

- We assume neither the shape of the nearest neighborhood of the query nor the presence of one cluster containing all relevant images.
- The relevance information of the original users' intention is stored directly in the VFs. This mechanism enables us to define a flexible dissimilarity measure.
- The proposed method combines the short-term and long-term RF techniques to establish an effective retrieval system.

3. Experimental Results

We have implemented the QVM approach [1] and the proposed VF technique. The UCR database is chosen for the experiments. The database is obtained from the UCR Visualization and Intelligent Systems Lab (VISLab) [4]. There are 10038 images covering a variety of outdoor scenes such as castles, cars, humans, animals, etc. Some sample images are shown in Fig. 2. Since the number of images in the database is tremendous, it is laborious to classify these images manually. As such, we employ the c-means clustering algorithm [5] to automatically classify these images into 70 classes for performance evaluation purpose. Each image is represented by a 16-dimensional feature vector using the Gabor filters [6].

Also, in all the experiments, the performance is measured using the precision rate defined as

$$\text{Precision Rate} = \frac{\text{Relevance Retrievals}}{\text{Total Retrievals}} \times 100\% \cdot$$

To simulate the practical situation of online users, the sequence of query images is generated randomly until each database image has been chosen at least once. Each query session is allowed to refine its retrievals by executing the RF process for two iterations. The average precision rates obtained at three different stages, namely the one without any relevance feedback (PR0), the one after the first feedback iteration (PR1), and the one after the second feedback iteration (PR2), are computed, respectively.

To understand the influence on the growing of precision rates by using the proposed VF technique, the accumulated precision rates that are averaged over the number of processed queries are plotted in Fig. 3. There is a fluctuating period in the beginning of the plotted curves depending on which images are firstly selected as query images. After this period, the accumulated precision rates climb up rapidly due to the contribution of the use of the active nearest neighborhood learned by the VFs. Looking at the curve of PR0, it reveals that the precision rate obtained even before performing the feedback iterations can be as high as 95% because the relevance information of the previously processed queries provides a valuable clue. Also, the improving ratio from PR0 to PR1 is higher than that from PR1 to PR2. This is a desired property since the users can not stand too many feedback iterations and they expect a greatly improved result after the first feedback. On the other hand, if we use solely the QVM method, there is no gain on the retrieval precision along the number of processed queries. As a result, the accumulated precision rates hold themselves to a relatively fixed value as shown in Fig. 4.

Next, we analyze the scalability of the proposed approach on the storage requirement of the VFs. First, we construct nine subdatabases from the UCR database. Each subdatabase consists of the images that are a certain amount of percentages of the original database volume (from 10% to 90%) and includes at least one image from every labeled class. Fig. 5 shows the storage requirement for the average length of the VFs. It is observed that the memory needs of the VFs grow less than three times when the test subdatabase size varies from 10% to 100% of the original database volume. Thus the proposed method is scalable against the variations of database size.

4. Conclusions

In this paper, we have presented a new RF approach for content-based image retrieval. The traditional RF techniques use only within-session query experience to improve the retrieval precision. We devise a new technique called the virtual feature which digests the cross-session query experiences to give the retrieval results that are more satisfactory. Experimental results show that the proposed retrieval system which uses a combination of short-term and long-term relevance information performs better than that adopting the short-term RF technique only.

5. References

[1] G. Salton and C. Buckley, Improving retrieval performance by relevance feedback, Journal of Am. Soc. Information Sci. 41 (1990) 288-297.

[2] Y. Rui, T. S. Huang, M. Ortega, and S. Mehrotra, Relevance feedback: a power tool for interactive content-based image retrieval, IEEE Trans. Circuits and Systems for Video Technology 8 (1998) 644-655.

[3] I. Cox, M. Miller, T. Minka, T. Papathomas and P. Yianilos, The Bayesian image retrieval system, PicHunter: theory, implementation, and psychophysical experiments, IEEE Trans. Image Processing 9 (2000) 20-37.

[4] Vision and Intelligent System Lab (VISLab), University of California, Riverside. Http://www.vislab.ucr.edu.

[5] L. Bobrowski and J. C. Bezdek, c-Means clustering with the l_1 and l_∞ norms, IEEE Trans. Syst. Man Cybernet. 21 (1991) 545-554.

[6] J. Peng, B. Bhanu, and S. Qing, Probabilistic feature relevance learning for content-based image retrieval, Computer Vision and Image Understanding 75 (1999) 150-164.

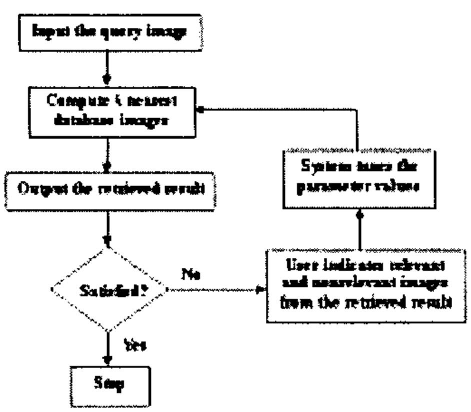

Fig. 1 System flow chart of the RF process.

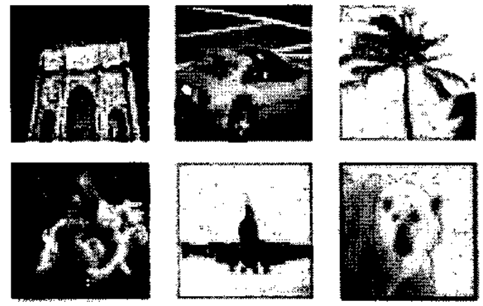

Fig. 2 Sample images from UCR database.

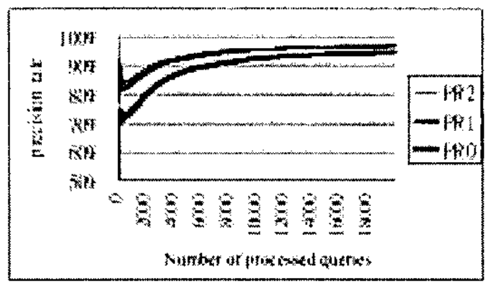

Fig. 3 Performance of the proposed method.

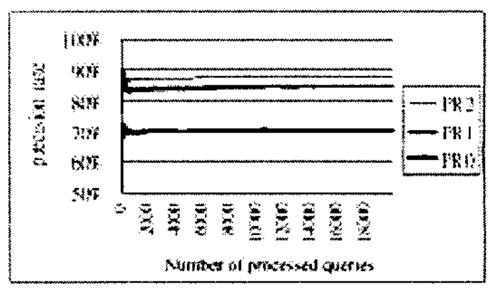

Fig. 4 Performance of the QVM approach.

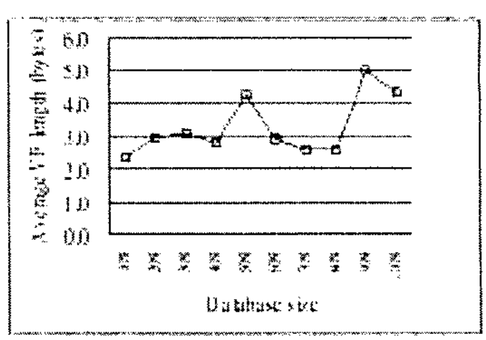

Fig. 5 The average VF length v.s. database size.