Relevance Feedback Reinforced with Semantics Accumulation

Sangwook Oh[1], Min Gyo Chung[2], and Sanghoon Sull[1]*

[1] Dept. of Electronics and Computer Engineering, Korea University, Seoul, Korea
{osu,sull}@mpeg.korea.ac.kr
[2] Dept. of Computer Science, Seoul Women's University, Seoul, Korea
mchung@swu.ac.kr

Abstract. Relevance feedback (RF) is a mechanism introduced earlier to exploit a user's perceptual feedback in image retrieval. It refines a query by using the relevance information from the user to improve subsequent retrieval. However, the user's feedback information is generally lost after a search session terminates. In this paper, we propose an enhanced version of RF, which is designed to accumulate human perceptual responses over time through relevance feedback and to dynamically combine the accumulated high-level relevance information with low-level features to further improve the retrieval effectiveness. Experimental results are presented to demonstrate the potential of the proposed method.

1 Introduction

An image retrieval system solely based on low-level image features is limited in its applicability due to some reasons: for example, it is very hard to represent high-level human perceptions precisely by using low-level visual features, and those low-level features are also highly sensitive to a small change in image shape, size, orientation and color. Many active research efforts thus have been made to overcome such limitations. Among them, relevance feedback (RF) is one notable approach to integrate high-level human concepts and low-level features into image retrieval [1,2,3]. In RF approach, a user is able to interactively specify the amount of relevance between a query and resulting images, and such relevance information is used to refine the query continuously to the user's satisfaction.

Though RF is an intriguing concept for interactive image retrieval, it has one serious drawback, which is that RF ignores valuable feedback information generated from user interactions during search sessions. However, we discover that the feedback information thrown away in this way contains the important information that captures the semantics of images, thus can be more intuitive and informative description of visual content than low-level image features. Motivated by this discovery, we propose a novel RF mechanism strengthened with a capability to store and reuse the relevance feedback information effectively.

* Corresponding author

P. Enser et al. (Eds.): CIVR 2004, LNCS 3115, pp. 448–454, 2004.

Specifically, the proposed method constructs a semantic space for a large collection of images by accumulating human perceptual responses over time through relevance feedback, and dynamically combines the accumulated high-level relevance information with low-level features to further improve the retrieval effectiveness. Experimental results show that the retrieval performance of the proposed method is greatly enhanced compared with traditional RF methods and gets better and better as time passes.

The rest of this paper is organized as follows. In Sec. 2, we describe the details of the proposed method: construction of a semantic vector space for an image database, and dynamic integration of semantic information and low-level image features into RF framework. Experimental results are presented in Sec. 3 to validate some good characteristics of the proposed method. Finally, concluding remarks are given in Sec. 4.

2 New Image Retrieval

2.1 Semantic Space

Relevance feedback responses, which are generated during search sessions but destroyed immediately in traditional RF mechanisms, are now accumulated to build a semantically meaningful high-level feature space, called *semantic space* hereafter. A semantic space for an image database is represented by an $n \times n$ matrix $M = (m_{ij})$, where n is the number of images in the image database and m_{ij} denotes a total of relevance scores accumulated over a certain period of time between a query image i and an image j on the image database. The more conceptually similar the two images i and j, the greater the value of m_{ij}.

Figure 1 shows a simple example of the semantic matrix M for a collection of 5 images. m_{ij} is initially zero for all images i and j, but will be filled soon with relevance values as search processes go on. For the given query image i, if the image j is marked by a user as *relevant* in a search session, m_{ij} gets updated by a particular relevance score. In Fig. 1, for example, if the image 4 is the query image and the image 5 is selected as a relevant image, then $m_{4,5}$ is changed from 5 to $5 + \alpha$, where α is a relevance score. Although there are many possible ways to determine relevance scores, we simply use the following scoring rule: if marked as relevant, then $\alpha = 1$; otherwise, then $\alpha = 0$.

There are two possible ways to update (and maintain) a semantic matrix: asymmetric or symmetric. In the asymmetric update scheme, if a user establishes a relevance between a query image i and an image j, then only m_{ij} in the semantic matrix is updated to a new value, but m_{ji} remains same. On the other hand, the symmetric update scheme makes both m_{ij} and m_{ji} get updated at the same time. Although it requires further studies to investigate the properties and effectiveness of the two update schemes, the asymmetric update scheme has a tendency to return different retrieval results depending on which of the images i or j is chosen as the query image, but the symmetric update scheme tends to yield similar retrieval results irrespective of the choice of the query image. We prefer to use asymmetric update scheme because it is more general than

	Im 1	Im 2	Im 3	Im 4	Im 5
Im 1	0	1	2	1	1
Im 2	3	0	1	0	2
Im 3	3	1	0	2	4
Im 4	1	2	2	0	5
Im 5	0	1	2	3	0

Fig. 1. A simple example of a semantic matrix for a collection of 5 images.

the symmetric scheme and has a capability to differentiate images with subtle differences in human perception.

2.2 Integration of Semantic and Low-Level Features

This section will give a detailed description of how to combine high-level semantic features and low-level visual features within RF framework. Without loss of generality, we assume that an image is associated with two low-level features, color and texture, and one high-level feature, semantic feature presented in the previous section. We take color correlogram [4] as the color feature, and use Shim and Choi's method [5] to obtain the texture feature. Assume further the following symbols and definitions for convenience of explanation:

- $S_C(i,j)$, $S_T(i,j)$, and $S_S(i,j)$ are a similarity measure of color, texture and semantic features, respectively, between two images i and j.
- W_C, W_T, and W_S are a weight associated with color, texture and semantic features, respectively.

Similarity Computation. The overall similarity between a query image i and an arbitrary image j, $S(i,j)$, can then be calculated using the above definitions as follows:

$$S(i,j) = W_C \frac{S_C(i,j)}{\max_j S_C(i,j)} + W_T \frac{S_T(i,j)}{\max_j S_T(i,j)} + W_S \frac{S_S(i,j)}{\max_j S_S(i,j)},$$

where we use m_{ij} in the semantic matrix M for the value of $S_S(i,j)$. In the above equation, $\max_j S_C(i,j)$, $\max_j S_T(i,j)$, and $\max_j S_S(i,j)$ indicate a maximum similarity value for the corresponding image feature, and are used to normalize each similarity measure. In other words, the overall similarity $S(i,j)$ is represented as a linear combination of individual normalized similarity measures.

Weight Update. For the query image presented by a user, an RF based retrieval system executes search algorithms and returns its results. The user then

views the retrieved images and judges which images are relevant and which images are not. The relevance information obtained in this way is used to dynamically update the weights of the image features as well as the semantic matrix. The weights should be changed in proportion to the relative importance of the image features.

Let R be the set consisting of k most similar images according to the overall similarity value $S(i,j)$, where k is the number of images the user wants to retrieve. Similarly, we define three more sets R_C, R_T, and R_S. That is, R_f is the set of k most similar images according to the similarity measure $S_f(i,j)$, where f can be any of three image features, C, T, and S. Then, the new weights for each image feature, f, are calculated using the following procedure, which is similar to the one in [1]:

1. $W_{sum} = 0$.
2. For each f in $[C, T, S]$, execute three steps below.
   a) Initialize $W_f = 0$.
   b) For each image p in R_f, $W_f = W_f + \alpha$ if p is in R.
   c) $W_{sum} = W_{sum} + W_f$.
3. The weights obtained in Step 2 are now normalized by the total weight W_{sum} as follows: $W_f = \frac{W_f}{W_{sum}}$.

The above procedure implies that the more overlap between R_f and R, the larger the weight of W_f. The weights updated in the current retrieval iteration are subsequently used to return more perceptually relevant images in the next iteration.

As the retrieval process continues, the weights change dynamically according to the user's information need and intention. Since the user selects only perceptually relevant images to the query through relevance feedback, as the retrieval process continues, more and more images governed by the similarity measure $S_S(i,j)$ tend to appear on the retrieval result list. As a result, the weight for the color or texture feature gets smaller, but the weight for the semantic feature gets bigger, which means the semantic feature plays a critical role in finding relevant images quickly. Due to this favorable phenomenon, false positive rates are also dramatically reduced.

3 Experiments

3.1 Experimental Setup

To study the effectiveness of the proposed method, we have implemented our image retrieval system that works as illustrated in Fig. 2. The initial weights are all $\frac{1}{3}$ for color, texture and semantic features. When it comes to updating the system-wide semantic matrix, the asymmetric update scheme is chosen because it is more general than the symmetric scheme and can afford to differentiate images with subtle differences in human perception.

Our image database contains 2700 natural images, which implies the dimension of the semantic matrix $M = (m_{ij})$ is 2700×2700. According to their content,

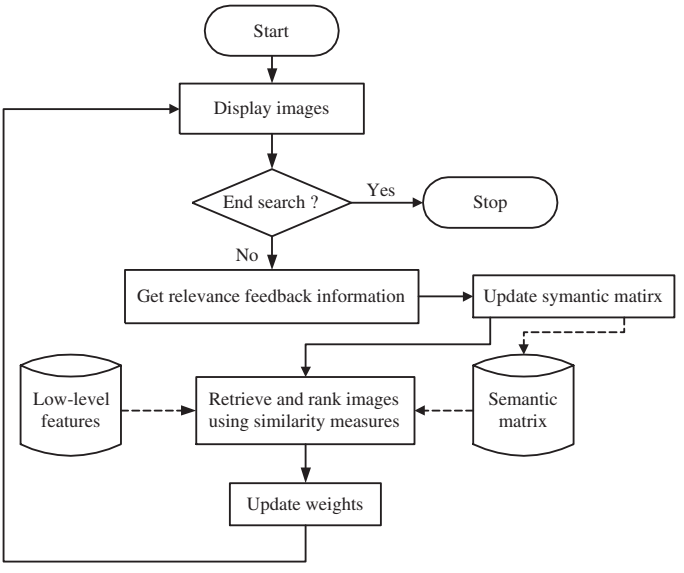

Fig. 2. Flowchart of the proposed retrieval system

the images in the image database are categorized into several groups: humans (celebrities, entertainers), animals, vehicles (cars, airplanes, motorcycles), stars, plants, natural scenes (sunrise, sunset, clouds, lightning), sports, cartoon characters, etc. A few users are asked to use our system to gather relevance feedback information into the semantic matrix. Some statistical figures to describe the initial semantic matrix are shown below:

- The total number of positive feedbacks (i.e., $\sum_{i,j} m_{ij}$) is 17823.
- The total number of row vectors in M that are not zero vector is 271. Therefore, the average positive feedbacks contained in one row vector are $\frac{17823}{271} = 66$.

As the size of an image database increases, the semantic matrix also requires a larger storage. It is, however, found that the semantic matrix is actually a kind of sparse matrix, which means that the large portion of the semantic matrix is filled with zero or empty. There are many well-known approaches to represent a sparse matrix in a memory-efficient manner. For instance, one way to relieve the memory requirement is to maintain only the non-zero elements for each row vector in the semantic matrix.

3.2 Experimental Results

Figure 3(a) compares the precision-recall graphs of the traditional RF method and the proposed RF method, where the nine values are obtained by varying the number of resultant images, such as 9, 18, 27, $\cdots$, 81. Here, the traditional

method employs only two low-level features (i.e., color and texture) while the proposed method uses the high-level semantic feature as well as the two low-level features. The above precision-recall graphs are obtained by averaging the search results of any 20 query images. Furthermore, the precision-recall graph of the proposed method is generated using the initial semantic matrix described in the previous section. In terms of search performance, the proposed method seems to be about 20% better than the traditional method.

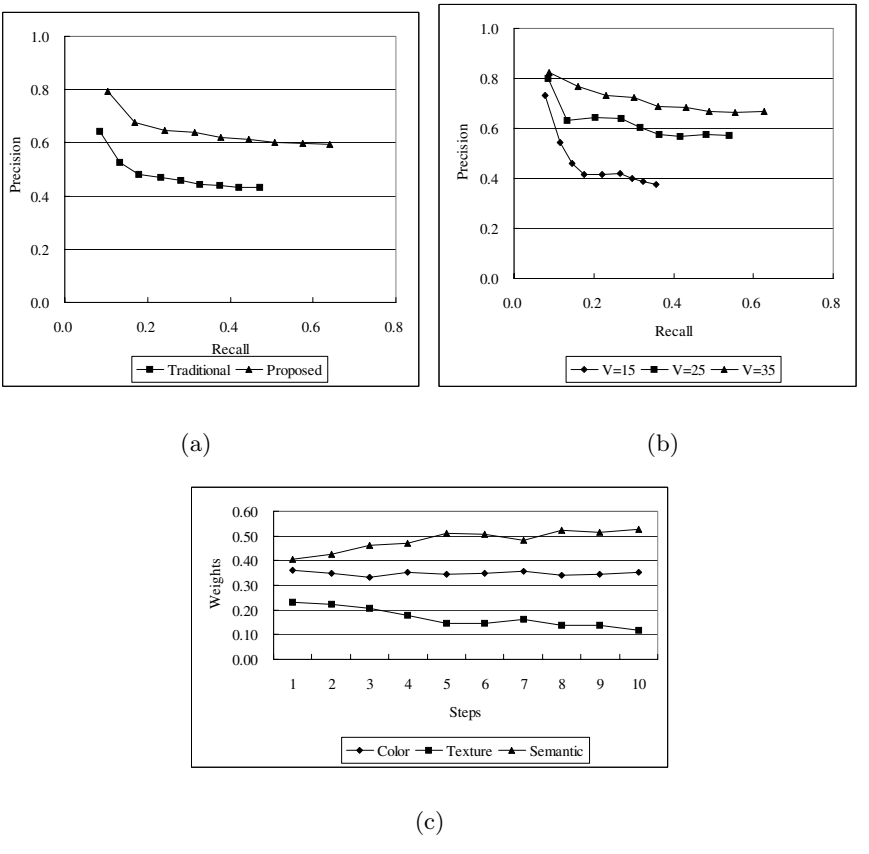

Fig. 3. (a) Precision-recall graphs of the traditional and the proposed RF methods, (b) Transition of precision-recall graph of the proposed method as the relevance feedback information keeps being added into the semantic matrix, and (c) Variation of three weights during a search session.

Figure 3(b) shows how the precision-recall graph of the proposed method changes as the semantic matrix grows. The symbol V in the legend indicates the average amount of relevance information contained in a row vector in the semantic matrix. The greater the value of V, the greater the accumulated infor-

mation in the semantic matrix. As we expect, the search performance continues to improve as the value V increases.

Figure 3(c) is another chart to demonstrate the transition of three weights during a search session. As the search session goes on, the weight of the semantic feature tends to increase while the weights of color or texture decrease or remain unchanged. This observation tells that the semantic feature plays a critical role in finding relevant images quickly. False positive rates can be also dramatically reduced thanks to such favorable behavior of the semantic feature.

4 Conclusions

Relevance feedback (RF) is a mechanism introduced earlier to exploit a user's perceptual feedback in image retrieval. Though RF is an intriguing concept for interactive image retrieval, it has one serious drawback, which is that RF ignores valuable feedback information generated from user interactions during search sessions.

In this paper, we propose a novel RF mechanism strengthened with a capability to store and reuse the relevance feedback information effectively. Specifically, the proposed method constructs a semantic space for a large collection of images by accumulating human perceptual responses over time through relevance feedback, and dynamically combines the accumulated high-level relevance information with low-level features to further improve the retrieval effectiveness.

Experimental results show that the retrieval performance of the proposed method is greatly enhanced compared with traditional RF methods and gets better and better as time passes. Furthermore, it is discovered that the semantic feature plays a critical role in finding relevant images quickly and reducing the false positive rates

References

1. Yong Rui, Thomas S. Huang, Michael Ortega, and Sharad Mehrotra, "Relevance Feedback: A Power Tool in Interactive Content-Based Image Retrieval", in IEEE Tran on Circuits and Systems for Video Technology, Special Issue on Segmentation, Description, and Retrieval of Video Content, pp. 644-655, Vol. 8, No. 5, Sept, 1998.
2. I. J. Cox, M. L. Miller, T. P. Minka, T. V. Papathomas, and P. N. Yianilos, "The Bayesian image retrieval system, PicHunter:theory, implementation, and psychophysical experiments," in IEEE Transaction on Image Processing, Vol. 9, pp. 20-37, Jan 2000.
3. H. Muller, W. Muller, S. Marchand-Maillet, and T. Pun, "Strategies for positive and negative relevance feedback in image retrieval," in Proc. of IEEE Conference on Pattern Recognition, Vol. 1, pp. 1043-1046, Sep 2000.
4. Jing Huang, S. Ravi Kumar, Mandar Mitra, Wei-Jing Zhu and Ramin Zabih. "Image Indexing Using Color Correlograms," in IEEE Conference on Computer Vision and Pattern Recognition, pp. 762-768, June 1997.
5. Seong-O Shim and Tae-Sun Choi, "Edge color histogram for image retrieval", 2002 Int'l conf. on Image processing, pp. 957-960, vol. 3, Jun.2002.