This paper is published in Proc. of 2002 Int'l Conf. on Image Processing, New York, Sep., 2002.

A GRAPHIC-THEORETIC MODEL FOR INCREMENTAL RELEVANCE FEEDBACK IN IMAGE RETRIEVAL

Yueting Zhuang[1], Jun Yang[1,2], Qing Li[2], Yunhe Pan[1]

[1]Department of Computer Science, Zhejiang University, Hangzhou, China
yzhuang@cs.zju.edu.cn; panyh@sun.zju.edu.cn
[2]Department of Computer Engineering and Information Technology,
City University of Hong Kong, HKSAR, China, {itjyang, itqli}@cityu.edu.hk

ABSTRACT

Many traditional relevance feedback approaches for CBIR can only achieve limited short-term performance improvement without benefiting long-term performance. To remedy this limitation, we propose a graphic-theoretic model for incremental relevance feedback in image retrieval. Firstly, a *two-layered graph model* is introduced that describes the correlations between images. A learning strategy is then suggested to enrich the graph model with semantic correlations between images derived from user feedbacks. Based on the graph model, we propose *link analysis* approach for image retrieval and relevance feedback. Experiments conducted on real-world images have demonstrated the advantage of our approach over traditional approaches in both short-term and long-term performance.

1. INTRODUCTION

Content-based image retrieval (CBIR) has received extensive study in recent years. In a typical paradigm of CBIR systems, the user submits a sample image as the query and a set of visually similar images are retrieved based on low-level image features. Therefore, the retrieval performance (usually in terms of precision and recall) of a CBIR system is severely limited when the sample image does not describe the user's need precisely. To overcome this limitation, many CBIR systems [1,3,4,5,6,7] have applied the relevance feedback technique, which improves the retrieval performance by adjusting the original query based on the relevant and irrelevant image examples designated by users.

Although the current feedback technique has been proved effective in boosting the retrieval performance, there is still room for improvement. On one hand, many feedback approaches seek to find query vector and similarity metric that best describes the desired images in the feature space. Therefore, if the desired images cannot be sufficiently described by low-level features, they fail to return many relevant results even with a large number of feedbacks. On the other hand, most feedback approaches, except a few [3,4,5], do not have a learning mechanism to memorize the feedbacks conducted previously and reuse them in favor of future queries. If we define a *retrieval session* as a user query and its subsequent feedback process, most approaches can only improve the retrieval results within a single session (short-term performance), without achieving better performance across different sessions (long-term performance). The existing learning mechanisms either require sophisticated computational models [3,5], or involve the use of keywords [4].

To tackle the aforementioned problems, we propose a graphic-theoretic model for incremental relevance feedback in image retrieval. The foundation of our approach is a two-layered graph model that describes the correlations between images using links. Motivated by link analysis, a widely used technique for Web information retrieval, we conduct image retrieval and relevance feedback by analyzing the link structure of the graph model. Furthermore, a learning strategy is suggested to derive semantic correlations between images from user feedbacks and incorporate them into the graph model, which promotes incrementally the long-term retrieval performance. The rest of this paper is organized as follows. In Section 2, we review briefly the existing feedback approaches. Section 3 describes the details of our approach. We present the experiment results in Section 4 and conclude the whole paper in Section 5.

2. RELATED WORK

Many traditional methods for relevance feedback focus on adjusting the query vector and/or improving the similarity metric for better description of the desired images. For example, the MARS system [7] has implemented both ideas. On one hand, the weight for different features in the similarity metric is adjusted based on user feedbacks, so that more weight is placed on the features that are characteristic of the desired images. On the other hand, the query vector is moved towards the relevant examples and away from the irrelevant examples in the feature space. The method proposed by Ishikawa et al. for the MindReader system [1] formulates a global optimization problem, the solution to which includes both the optimal similarity metric and the optimal query vector. Rui et al. [6] further improved this approach by proposing a hierarchical model that can accommodate various types of features.

A commonality of the methods mentioned above is that for a given query they seek to find a close region (subspace) in the feature space that covers the maximum number of good results. The center of the region is defined by the optimal query vector and its shape is determined by the optimal similarity metric. Obviously, this maximum number is subject to the distribution of good results in the feature space. Unfortunately, the good results for many queries, which are defined at the semantic level, conform to a sparse or irregular distribution in the feature space. In this case, no matter how the region is optimized, the number of good results it covers is limited and thus the retrieval performance is low. Furthermore, since the query vector and the similarity metric are optimized for a specific retrieval session and are discarded when the session is finished, the historical feedbacks cannot be reused to process future queries, which have to be started from scratch. Therefore, the long-term performance remains unchanged even after a great number of feedbacks.

A few learning mechanisms have been proposed for relevance feedback. For example, Minka et al.'s system [5] precomputes many possible groupings of images based on "a society of models" and learns the "bias" towards these groupings from relevant/irrelevant examples to facilitate future queries. Lee

et al. [3] proposed a method to capture the semantic correlations between images from feedbacks and embed them into the system by splitting/merging image clusters, based on which image retrieval is conducted. Both approaches employ complicated mathematical model. In contrast, *iFind* system [4] adopts a simple keyword propagation mechanism that learns the keyword annotation of images from user feedbacks. To make it work, however, the query must be formulated by keywords.

Similar to CBIR approaches, our approach is also based on low-level image features. Nevertheless, because of the novel model used, it overcomes in a certain degree the inherent limitation of traditional feedback methods by achieving superior short-term performance and promoting long-term performance. Compared with the existing learning mechanisms, our approach is simpler and computationally more efficient than [3] and [5], and does not involve keywords in the retrieval process.

3. OUR APPROACH

3.1. Two-Layered Graph Model

In our approach, an image is described through its correlations with other images in a two-layered graph model shown in Fig.1. It consists of two superimposed layers, the semantic layer and visual layer. Each layer is an undirected graph, in which each node represents an image in the database, and each link between two nodes represents the correlation between two corresponding images, with an associated weight to indicate the strength of the correlation. The nodes of the two layers correspond to the same set of images, but their links have different interpretations. A link in the semantic layer (semantic link) reveals the correlation between two images defined from a high-level semantic perspective, while a link in the visual layer (visual link) denotes the visual similarity between them defined on low-level features.

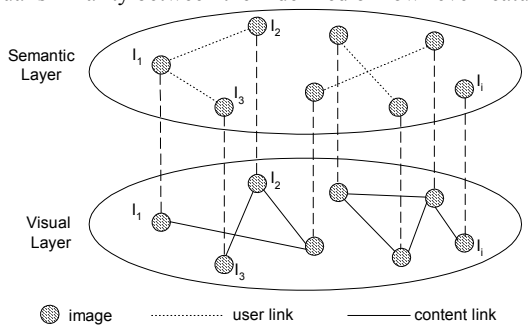

Fig.1: The two-layered graph model

The graph model provides the foundation based on which image retrieval is conducted. Therefore, the richness and quality of links are essential to the retrieval performance. In our approach, visual links are obtained in an offline manner: When an image is registered into the database, its similarity with every existing image is calculated based on low-level features and normalized to [0,1]. Visual links are created between two images that have a similarity above a threshold, with the link weight set to the similarity. The threshold is defined with a relatively high value in order to prevent creating noisy links between images with low similarity. For image retrieval, semantic links are of greater importance than visual links, since people tend to judge the similarity between two images from a semantic perspective. Due to the difficulty of automatic image understanding, semantic links are derived from user interactions in an online manner, using the learning strategy described below.

3.2. Learning Strategy

A straightforward way to create semantic links is to identify manually all the semantically relevant images, which is laborious and time-consuming. Alternatively, we suggest a machine learning strategy to derive the semantic links automatically from the information contained in user feedbacks. The idea is simple and intuitive: When a user submits a sample image as the initial query and labels some of the retrieved images as relevant or irrelevant, semantic links are created between the sample image and each relevant example, while the existing semantic links between the sample and any irrelevant example are removed. An algorithmic description of this strategy is given below:

1. *Collect the sample image I_S, the set of relevant examples I_R, and the set of irrelevant examples I_N.*
2. *For each $I_i \in I_R$, if there is no semantic link between I_i and I_S, create a semantic link between them with the initial weight set to 1. Otherwise, increase the weight of the semantic link between them with an increment of 1.*
3. *For each $I_i \in I_N$, if there is a semantic link between I_i and I_S, divide its weight by a factor of 4. If the resulting weight is below 1, remove that link.*

As the system interacts with various users over sessions, semantic links are created and their weights are tuned, which gradually reflect the users' perception of the semantic correlations between images. These semantic links will be utilized to process queries that come afterwards. In this regard, our learning strategy provides a mechanism to memorize, accumulate, and reuse the historical feedback information in favor of long-term performance.

3.3. Retrieval and Relevance Feedback by Link Analysis

Our graph model has an analogy with the Web environment in which web pages are interconnected by hyperlinks. Many approaches have been proposed to retrieve web pages by analyzing the structure of hyperlinks [2], based on the assumption that each hyperlink suggests a relationship between two pages connected by it. The same assumption also holds in our case, since the links of the graph model reveal either semantic or visual correlations between images. Therefore, we apply the idea of link analysis for the purpose of image retrieval.

In our approach, the sample image can be either selected from the existing images in the database, or submitted by the user. In the latter case, the new image is immediately registered into the graph model, with all the necessary visual links created. Therefore, in both situations the query is started from a node in the graph model. Our retrieval algorithm is based on the notion of *similarity propagation*—the similarity (with the query) can be propagated from one image to its correlated images through either visual links or semantic links. To process a query, we firstly pump the initial similarity into the node representing the sample image, and then allow the similarity to flow through the links of the graph model, with the amount of flow modulated by the weights of links. The asymptotic pattern of the similarity distribution among the nodes defines the similarity of the corresponding images with the query.

Suppose the semantic layer and the visual layer are represented by their adjacency matrix M_u *and* M_C, in which each off-diagonal element m_{ij} contains the weight of the link between image I_i and I_j and all the diagonal elements are set to 0. If there is no link between I_i and I_j, m_{ij} is also set to 0. The propagation process is modeled over discrete steps $t=0,1,...,N$. We define $R(t)$ as the similarity vector at step t, with its element $r_i(t)$ being the similarity of image I_i to the query. In the initial vector $R(0)$, the element corresponding to the sample image is set to 1, while other elements are 0. Since there are two layers, the propagation is performed between them in an interleaved manner:

$$\begin{aligned} R(2t+1) &= [\alpha_u M_u + (1-\alpha_u)I]R(2t) \\ R(2t+2) &= [\alpha_c M_c + (1-\alpha_c)I]R(2t+1) \quad t = 0,1,\cdots,N \end{aligned} \tag{1}$$

where I is a identity matrix, α_u and α_c are parameters within [0,1], which determines the amount of similarity that flows along the links, with larger value corresponding to a larger flow of similarity. Because the user-perceived semantic correlation is more reliable than the visual correlation defined on low-level features, we intuitively set α_u =0.1 and α_c =0.01, such that a semantic link carries a larger flow of similarity than a visual link of the same weight. The number of propagation steps is set to 6, since the distribution of similarity reaches a stable pattern after 6 steps. The vector $R(t)$ after the propagation process gives the final similarity of each image to the query.

After the first batch of results is displayed, users can refine the results by designating relevant and irrelevant image examples. Based on these feedbacks, we firstly update the semantic links using the learning strategy given in Section 3.2. After that, the similarity of each candidate image is recalculated using similarity propagation based on the following notion: An image that is in the proximity of a relevant example and at the same time away from the proximity of irrelevant examples has a high probability to be a good result. Here the proximity of an image is defined as the images that connect with it via a few links (either visual link or semantic link) in the graph model. The feedback process is presented as follows:

1. *Collect the set of relevant examples I_R and set of irrelevant examples I_N.*
2. *Apply the learning strategy in section 3.2 to update the semantic links.*
3. *Initialize similarity vector $R_R(0)$, s.t. $r_i(0)=1$ if $I_i \in I_R$; otherwise, $r_i(0)=0$. Perform similarity propagation based on $R_R(0)$ using Eq.1 for M_R steps and get $R_R(M_R)$.*
4. *Initialize similarity vector $R_N(0)$, s.t. $r_i(0)=1$ if $I_i \in I_N$; otherwise, $r_i(0)=0$. Perform similarity propagation based on $R_N(0)$ using Eq.1 for M_N steps and get $R_N(M_N)$.*
5. *$R^* = R_R(M_R) - R_N(M_N)$*
6. *Display images ranked in descending order of their similarity indicated by R*.*

As illustrated in Fig.2, the above algorithm performs propagation using relevant and irrelevant examples as the seeds respectively. As a result, $R_R(M_R)$ defines the similarity of the candidate images to the relevant examples, while $R_N(M_N)$ defines their similarity to the irrelevant examples. Hence, the difference between $R_R(M_R)$ and $R_N(M_N)$ gives a good estimation of the overall similarity. Note that we set M_R to a larger value than M_N (currently, M_R=6 and M_N=4) based on the following observation: A good result can be visually very similar to and thus has a visual link with an irrelevant example, and in this case over-propagation from irrelevant examples may pull down the similarity of many good results.

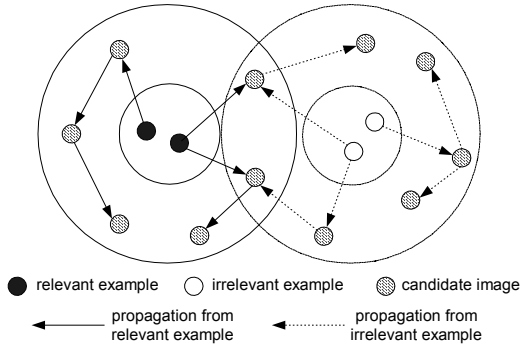

Fig.2: Relevance feedback by relevance propagation

The benefit of our approach to long-term performance is evident. Initially, since there is no semantic links between images, the retrieval is performed solely on the visual layer (low-level features). At this stage, our approach is no more than a CBIR approach. As user feedbacks are conducted, semantic links are gradually incorporated into the graph model. Better performance can be expected because not only the low-level features but also the user-perceived semantic correlations are exploited in the retrieval process.

4. EXPERIMENTAL RESULTS

To manifest the effectiveness of our approach, we have implemented a prototype system using this approach and conducted experiments on real-world images. The test data consists of 4,000 images selected from Corel Image Gallery. As the *ground truth*, the test images are classified into 40 topical categories by domain professionals, with exactly 100 images in each category. Since this classification is based on high-level concepts, the images within some categories (e.g., "city") display heterogeneous visual features. Images from the same category are considered relevant. The low-level features extracted to create visual links include 256-d HSV color histogram, 64-d Lab color coherence, and 32-d Tamura directionality. Euclidean distance is used as the similarity measure.

In our experiments, a query is formulated by a sample image randomly selected from the test data. For each query, the system retrieves the first 100 images that are ranked top by the retrieval algorithm. User feedbacks are automatically generated among these 100 images by the system according to the ground truth, i.e., images that belong to the same category as the sample are labeled as relevant and the rest are labeled as irrelevant. The feedback examples are fed into our feedback algorithm to refine the current results. Since the number of retrieved images is equal to the number of relevant images, precision and recall are the same and we use "retrieval accuracy" to refer to both of them. Using the above experimentation method, we studied both the short-term and long-term performance of our approach.

For short-term performance, we examined the change of retrieval accuracy in the feedback process of a single retrieval session. We generated 200 random queries (5 queries for each category) and conducted 10 rounds of feedback for each query,

with the average accuracy achieved at each round shown in Fig.3. For comparison, we implemented three traditional feedback methods, which are the query vector adjustment (QVA for short) and the similarity metric adjustment (SMA) proposed in MARS [7], as well as the hybrid approach in MindReader [1]. All these methods are based on the same set of low-level features. We use them to process the same 200 queries and plot their performance in Fig.3 together with that of our approach.

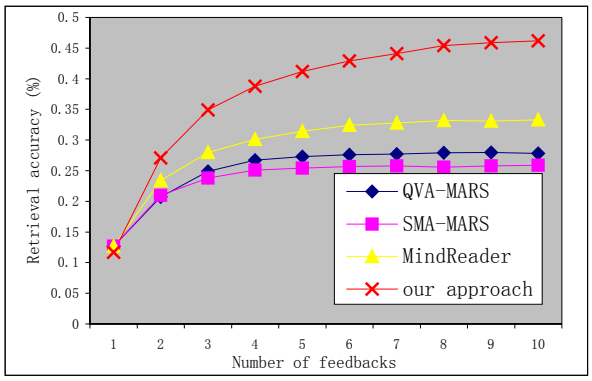

Fig.3: Comparison on short-term performance

As can be seen, all the methods are at the same performance level before any feedback is done. This is because they rely on the same set of low-level features, though our retrieval algorithm is slightly different. As feedback proceeds, our approach outperforms significantly the three traditional feedback methods, while the difference between those three is relatively small. This observation can be explained as follows: As discussed in Section 2, the performance of traditional feedback methods is bounded by an "optimal subspace" and is therefore subject to the distribution of good results in the feature space. Although our approach also relies on low-level features, it can transcend this bound through semantic links, which can connect two relevant images that are far away from each other in the feature space. Thus, more relevant images that are not visually similar to the sample can be found by following the semantic links as bridges.

Long-term performance is examined as follows: For each category, we applied a succession of retrieval sessions, each consisting of a random query followed by a single round of feedback. Since the feedback in each session causes some semantic links to be incorporated into the graph model, and these links are exploited for the subsequent sessions, the change of the retrieval accuracy over different sessions reflects the long-term performance. We conduct this experiment on all the categories and show the change of average accuracy in Tab.1. As we can see, the accuracy improves substantially over sessions, reaching 45% after 12 sessions. Given that only a single round of feedback is conducted in each session, our approach is very effective in promoting long-term performance. In comparison, the three traditional methods in the last experiment have no learning mechanism to enhance their long-term performance. In our preliminary experiments, we have not compared our method with the learning mechanisms proposed by Minka et al. [5] and Lee et al. [3] due to their complexity and the limited time. (The first approach requires a society of models for image segmentation, and the second one uses neural network model).

Tab.1: Improvement of retrieval accuracy over sessions

Sessions	1	2	3	4	5	6
Accuracy(%)	13.1	26.7	32.1	35.5	37.7	39.3
Session	7	8	9	10	11	12
Accuracy(%)	40.5	41.8	43.0	43.9	44.6	45.1

5. CONCLUSION

In this paper, we have proposed a graphic-theoretic model for incremental relevance feedback in image retrieval. A two-layered graph model is introduced to describe the correlations between images. Based on it, image retrieval and relevance feedback are performed by link analysis approach. A learning strategy is suggested to enrich the graph model with user-perceived semantic correlations between images. Experiment results have verified the advantage of our approach over traditional CBIR feedback methods in both short-term and long-term performance.

Using the graphic-theoretic model has a 2-sided effect on the efficiency of image retrieval. On one side, the search space is reduced from the entire database to a small locality of the graph model. On the other side, however, the number of links grows exponentially with the number of images, making the storage and access of links extremely expensive. Our future work includes developing a strategy for efficient storage and access of links, as well as a method to identify and eliminate noisy links. Moreover, we will investigate the potentials of the graph model and the link analysis approach in supporting other popular functionalities of an image system, such as classification, navigation, and browsing.

ACKNOWLEDGMENTS

The work described in this paper was supported, substantially, by a grant from CityU (Project No. 7100196), partially by a grant from the Research Grants Council of the Hong Kong Special Administrative Region, China [Project No. CityU 1119/99E], and partially by a grant from the Doctorate Research Foundation of the State Education Commission of China.

REFERENCE

1. Y. Ishikawa, et al., "Mindreader: Query databases through multiple examples," *Proc. of the 24th VLDB Conf.*, New York, pp. 218-227, 1998.
2. J.M. Kleinberg, "Authoritative Sources in a Hyperlinked Environment." J. of the ACM, 46(5):604-632, 1999.
3. C.S. Lee, W.Y. Ma, H.J. Zhang, "Information Embedding Based on User's Relevance Feedback for Image Retrieval," Technical Report, HP Labs, 1998.
4. Y. Lu, et al. "A Unified Framework for Semantics and Feature Based Relevance Feedback in Image Retrieval Systems," *Proc. of ACM Multimedia*, pp. 31-38, 2000.
5. T.P. Minka et al., "Interactive learning with a 'society of models'," *Proc. of IEEE CVPR*, pp. 447-452, 1996.
6. Y. Rui and T.S. Huang, "Optimizing Learning in Image Retrieval," *Proc. of IEEE CVPR*, pp. 236-243, 2000.
7. 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.