IMAGE RETRIEVAL BASED ON BAG OF IMAGES

Jun Zhang and Lei Ye

School of Computer Science and Software Engineering,
University of Wollongong,
Wollongong, NSW, 2522 Australia

ABSTRACT

Conventional relevance feedback schemes may not be suitable to all practical applications of content-based image retrieval (CBIR), since most ordinary users would like to complete their search in a single interaction, especially on the web search. In this paper, we explore a new approach to improve the retrieval performance based on a new concept, bag of images, rather than relevance feedback. We consider that image collection comprises of image bags instead of independent individual images. Each image bag includes some relevant images with the same perceptual meaning. A theoretical case study demonstrates that image retrieval can benefit from the new concept. A number of experimental results show that the CBIR scheme based on bag of images can improve the retrieval performance dramatically.

Index Terms— Content-based image retrieval, similarity measure, bag of images, information retrieval

1. INTRODUCTION

Content-based image retrieval (CBIR) is a technology to search for relevant images to a user's query from an image collection [1, 2]. Conventional CBIR schemes employing relevance feedback have achieved certain success [3, 4]. The idea of relevance feedback is to involve the user in the retrieval process so as to improve the final retrieval results. First, the user supplies an image as a query and the system returns an initial set of retrieved results. After that, the user labels some returned images as relevant or irrelevant and the system adjusts the retrieval parameters based on the user's feedback. Then, the system displays the revised retrieval results. Relevance feedback can go through one or more iterations until the user is satisfied with the results. However, relevance feedback may not be suitable to some practical applications of CBIR, e.g., the web search. Actually on the web, few people use advanced search interfaces and most would like to complete their search in a single interaction [5]. The reason may have two aspects: it's hard for ordinary users to understand relevance feedback, and relevance feedback is mainly a recall enhancing strategy while web search users are only concerned with the precision in the first several pages. In this paper, we concentrate on developing new retrieval schemes to improve the retrieval performance without relevance feedback.

Image classification is a technology to classify relevant images which is related to image retrieval [6, 7]. Normally, one implicit assumption of image classification is that all image categories in a collection have been built up and several representative examples for each category are available. The goal of image classification is to assign a new image to an existing image category. The image classification techniques also can be used for image retrieval task. First, the user supplies an image as a query and the system assigns it to an existing image category. Then, all images in the predicted category will be returned as retrieval results. In image classification, we can design a specific similarity metric for each image category. By comparing the similarities between the query image and all image categories, we can classify the query image with high precision and get good retrieval results. However, it's difficult, if not impossible, to build up all image categories and provide some representative examples for each category. The reason is that in the world there is a large number of categories between which humans are able to distinguish [8].

In this paper, we propose a new concept, bag of images, for CBIR schemes. The basic idea is to relax the assumption of image classification. We consider that image collection comprises of image bags instead of well organized image categories. Each image bag includes some relevant images which have a same perceptual meaning. The image bags are constructed before image retrieval, e.g., some relevant images found on a web page can constitute an image bag. If the image collection is derived from large amounts of small organized collections, some neighboring images can constitute an image bag. Furthermore, a user's query is an image bag, named query image bag. Under this circumstance, all image bags in the image collection will be sorted in accordance with their similarities to the query image bag. Then, the image bags relevant to the user's query will be close to the query image bag while the irrelevant image bags will be further away from it. It is illustrated theoretically that the new concept can benefit image retrieval. A number of experimental results show that CBIR schemes based on bag of images can improve the retrieval performance dramatically.

978-1-4244-5654-3/09/$26.00 ©2009 IEEE

2. BAG OF IMAGES

In this section, we first illustrate the difference among three image retrieval models, bag of images based retrieval model, relevance feedback based retrieval model and image classification based retrieval model. Then, the important question, why the new concept, bag of images, can benefit image retrieval, will be investigated through a case study.

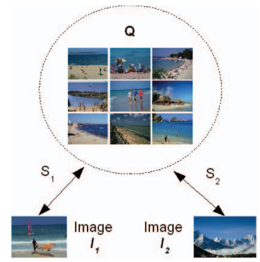
2.1. Retrieval models

Fig.1 shows the three different retrieval models. In the relevance feedback based image retrieval model, some query images will be provided by the user during multiple feedback iterations. In Fig.1(a) the query images are all positive although sometimes some negative examples also can be supplied. All images in the collection will be sorted in accordance with their relevance to the query images. The implicitly assumption is that the images in a collection are independent to each other. In the image classification based image retrieval model, the single query image will be classified into an existing image categories by its relevance to the representative examples of each category. The images in the category that the query image belongs to, will be returned as the retrieval results. A critical assumption is that image collection has been well organized and some representative images of each category are available. In the proposed bag of images based image retrieval model, the image collection consists of a large number of image bags. Each image bag includes several relevant images. After the user supplies a few images as a query image bag, the system will sort the image bags in accordance with their relevance to the query image bag. The top image bags will be returned as the retrieval results. This model is more flexible than other models. When the size of image bags is 1 and relevance feedback is available, this model changes to the conventional relevance feedback based image retrieval model. When all image categories are predefined and an image bag can represent an image category, this model changes to the image classification based image retrieval model. Normally, each image bag includes multiple relevant images. And two image bags may belong to the same image category or different image categories. The important feature is that the relevance information of images in an image bag can be used to improve the retrieval performance.

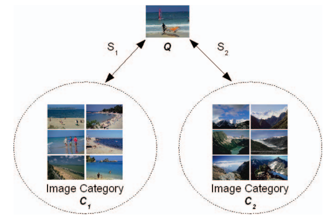
2.2. Case study

In this section, through a case study we explain, how image retrieval can benefit from the new concept, bag of images. Suppose there are two image categories, C_1 and C_2. Some images are selected from C_1 as a query, $Q = \{q_1, ..., q_n\}$. The dissimilarity between an image x to the query Q is denoted as

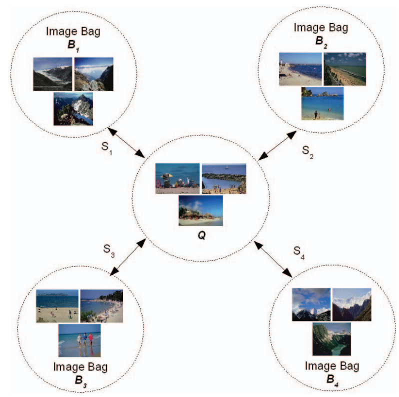
$$S_x = D(x, Q), \tag{1}$$



(a) Relevance feedback based



(b) Image classification based



(c) Bag of images based

Fig. 1. Three retrieval models

where $D()$ is a method for dissimilarity calculation. For example, in our experiments, CombSumScore [9] is applied to obtain the dissimilarity by combining the distances of multiple features and multiple images. We assume that dissimilarities of the images in a same category are independent and identically distributed (i.i.d.). Normal distributions are considered in this case study. We denote the probability density functions (pdf) [10] as

$$\begin{cases} p(S_x) \sim N\left(\mu_1, \sigma^2\right) \text{ for } C_1 \\ \\ p(S_x) \sim N\left(\mu_2, \sigma^2\right) \text{ for } C_2 \end{cases} . \tag{2}$$

For simplicity, we consider the two distributions have different means but the same standard variance. In conventional image retrieval schemes, all images in a collection will be sorted by their dissimilarities. Some irrelevant images with less dissimilarity may influence the retrieval performance. In this case, the probability that the affect of C_2 images to the retrieval performance is illustrated by the area highlighted by larger unshaded triangle in Fig.2.

Now we consider an image retrieval scheme based on bag of images. For an image bag B including m images, we compute the total dissimilarity between an image bag and the query image bag using the average rule,

$$S_B = \frac{1}{m} \sum_{x \in B} D(x, Q). \tag{3}$$

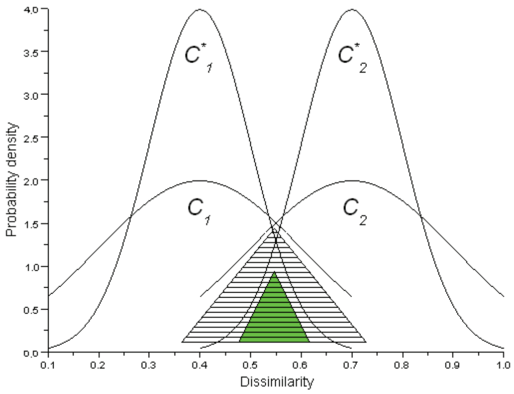


Fig. 2. Case study

Then, the probability density functions of image bags in two categories can be expressed as

$$\begin{cases} p^*(S_B) \sim N\left(\mu_1, \frac{\sigma^2}{m}\right) \text{ for } C_1^* \\ p^*(S_B) \sim N\left(\mu_2, \frac{\sigma^2}{m}\right) \text{ for } C_2^* \end{cases}. \tag{4}$$

The interference area is highlighted by shaded triangle (green) in Fig.2. In Fig.2, we can understand the change of probability density curves and the change of interference area. Obviously, the interference area becomes smaller when the concept of image bag is introduced into image retrieval scheme, so the retrieval performance can be improved effectively. From this case, we can draw a conclusion that image retrieval can benefit from the concept of image bag, since it is able to make relevant images to be close to the user's query and irrelevant images to be further away from it.

2.3. Discussion

In this section, we try to answer the other question, how to get the image bags in the practical applications of CBIR.

- On the web search, a web page may include some relevant images which can be used to construct an image bag. The relevance analysis can be based on the text around an image or the content of an image.
- If the image collection is derived from large amounts of small organized collections, some neighboring images with high relevance can constitute an image bag.
- When the CBIR technology is applied to search interesting pictures in a user's private photo collection, the pictures with consecutive numbers can constitute an image bag since these pictures may refer to the same visual concept with high probability.

In a word, the concept, bag of images, is presented to explore the hidden relationship among images in an image collection.

3. RETRIEVAL EXPERIMENTS

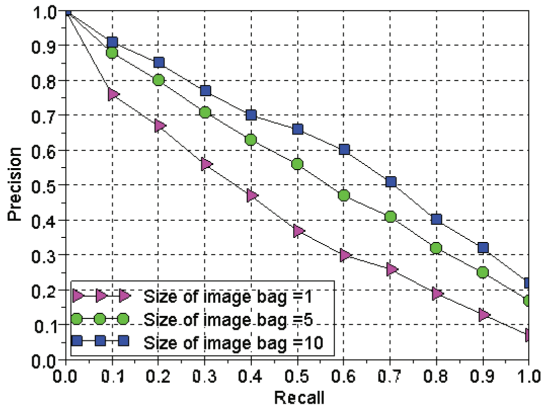
An experimental CBIR system is implemented, in which the input is some query images and no relevance feedback is available. All experiments are carried on a sub set of Corel image collection. There are 20 image categories and each image category includes about 100 images. The images in one category have an identical perceptual meaning, so the ground truth is based on the image category. In our experiments, the low-level features selected for image representation are color and texture. Five standardized MPEG-7 visual descriptors [11] are used in the experiments. The distance metrics recommended by MPEG-7 are used to measure the feature distances.

Since we concentrate on the small number of examples, each query consists of only 1 or 5 image examples. We use the CombSumScore scheme to obtain the dissimilarity between a collection image and the query, which is the best one in the aggregation schemes evaluated by Donald *et.al.* [9] for multiple features and multiple examples. Considering there are f feature descriptors and n query images, the CombSumScore scheme can be expressed as

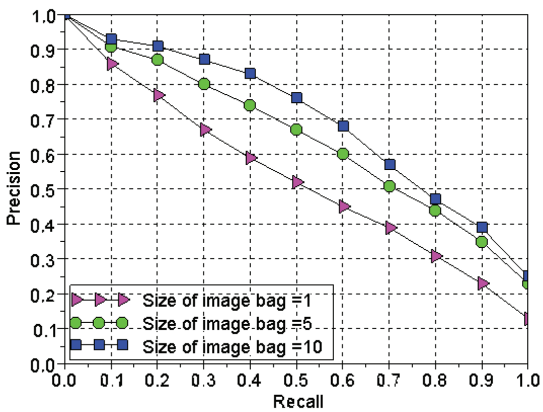
$$S_x = \sum_{i=1}^{n} \sum_{j=1}^{f} d_{ij}, \tag{5}$$

where d_{ij} is the distance of a collection image to the ith query image on the jth feature. To evaluate the usefulness of image bag, the size of image bag is manually set as 1, 5 and 10. The dissimilarity between an image bag and the query image bag is calculated using Eq.(3). When the size of image bag is 1, it reduces to the CombSumScore scheme. The retrieval performance in terms of average precision and recall on 300 random queries are reported. Precision is defined as the fraction of retrieved images that are relevant. Recall is defined as the fraction of relevant images that are retrieved [12].

Fig.3(a) and (b) show the retrieval performances of single query image and 5 query images, respectively. The results demonstrate the retrieval performance can be improved dramatically when the size of image bag increases. The reason is, the larger the size of image bag is, it is easier to discriminate relevant image bag and irrelevant image bag. In the case of single query image, when the size of image bag is 5, the average retrieval performance can be improved by over 15%, compared to the conventional CombSumScore scheme. When the size increases to 10, the retrieval performance can be further improved by about 5%. In the case of 5 query images, when the size of image bag is 5, the average retrieval performance can be improved by over 10%, compared to the conventional CombSumScore scheme. When the size increases to 10, the retrieval performance can be further improved by about 5%. We can find that the retrieval performance can be improved effectively even though the size of image bag is small, which is very suitable to the practical applications of CBIR.



(a) Single query image



(b) 5 query images

Fig. 3. Retrieval performance

4. CONCLUSIONS

In this paper, we aimed to explore the approaches to improve the image retrieval performance without relevance feedback. To achieve this goal, a new concept, bag of images, was introduced into CBIR schemes. We considered that image collection comprises of image bags instead of independent individual images. Each image bag includes some relevant images which have an identical a same perceptual meaning. A theoretical case study demonstrated that image retrieval can benefit from the new concept since image bag is able to make relevant images to be close to the user's query and irrelevant images to be further away from it. Although a simple method is applied to measure the dissimilarity between two image bags, the experimental results show that the CBIR scheme based on bag of images can improve the retrieval performance dramatically. The future work will focus on two directions, one is investigating the automatic technologies to construct image bags and the other is developing more effective methods to measure the dissimilarity between two image bags.

5. REFERENCES

[1] A. W. M. Smeulders, M. Worring, S. Santini, A. Gupta, and R. Jain, "Content-based image retrieval at the end of the early years," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 12, pp. 1349–1380, Dec. 2000.

[2] M. S. Lew, N. Sebe, C. Djeraba, and R. Jain, "Content-based multimedia information retrieval: State of the art and challenges," *ACM Transactions on Multimedia Computing, Communications, and Applications*, vol. 2, no. 1, pp. 1–19, February 2006.

[3] Y. Rui, T. S. Huang, M. Ortega, and S. Mehrotra, "Relevance feedback: A power tool for interactive content-based image retrieval," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 8, no. 5, pp. 644–655, Sep. 1998.

[4] X. S. Zhou and T. S. Huang, "Relevance feedback in image retrieval: A comprehensive review," *Multimedia Systems*, vol. 8, no. 6, pp. 536–544, Apr 2003.

[5] C. D. Manning, P. Raghavan, and H. Schtze, *Introduction to Information Retrieval*. Cambridge University Press, 2008.

[6] A. Vailaya, M. A. T. Figueiredo, A. K. Jain, and H.-J. Zhang, "Image classification for content-based indexing," *IEEE Trans. Image Process.*, vol. 10, no. 1, pp. 117–130, Jan. 2001.

[7] A. Frome, Y. Singer, and J. Malik, "Image retrieval and classification using local distance functions," in *Advances in Neural Information Processing Systems*, 2006.

[8] I. Biederman, "Recognition-by-components: A theory of human image understanding," *Psychological Review*, vol. 94, no. 2, pp. 115–147, 1987.

[9] K. M. Donald and A. F. Smeaton, "A comparison of score, rank and probability-based fusion methods for video shot retrieval," in *Proc. of IEEE Int. Conf. on Image And Video Retrieval*, Singapore, Jul. 2005, pp. 61–70.

[10] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification*. New York: Wiley, 2001.

[11] B. S. Manjunath, J. R. Ohm, V. V. Vasudevan, and A. Yamada, "Color and texture descriptors," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 11, no. 6, pp. 703–715, Jun. 2001.

[12] H. Muller, W. Muller, D. M. Squire, S. Marchand-Maillet, and T. Pun, "Performance evaluation in content-based image retrieval: overview and proposals," *Pattern Recognition Letters*, vol. 22, no. 5, pp. 593–601, Apr. 2001.