A Recursive Matching Method for Content-based Image Retrieval

Hyunwoo Kim, Tae-Kyun Kim, Wonjun Hwang, Seok Cheol Kee
Human Computer Interaction Lab., Samsung A.I.T., Republic of Korea
hwkim@sait.samsung.co.kr

Abstract

This paper presents an image retrieval method using a recursive matching scheme. In content-based image retrieval, a feature of an image cannot be perfectly extracted without inclusion of environmental noises like lighting and pose variation because of the imperfection of the extraction tools. Therefore, a feature extracted by a visual tool is the degraded version of the content information by environmental noise. To remove the effect by the environmental noise in image retrieval, we reuse the first-ranked retrieval image as an extra query image. The procedure are repeated within a fixed buffer of small size in a recursive way. Experimental results in the context of face retrieval show that the proposal method can improve image retrieval accuracy in the cost of the ignorable burden of the computational complexity.

Keywords: content-based image retrieval, recursive matching, face recognition

1. Introduction

Recently, content-based multimedia search has been focused in the aspect of overcoming the limitation of searching keywords in text-based retrieval and providing a friendly interface to human users. Specifically, the rapid Internet growth, the personalization of multimedia equipments, and the fast emerging of digital library are speeding up the need for content-based multimedia search.

Content-based image retrieval is to search the most visually similar images in ranked order by analyzing the information of image features like color, texture, shape, and face. In content-based image retrieval, feature vectors are extracted to describe image characteristics. First, all the gallery images are represented by their corresponding feature vectors (feature extraction), then when a query image is presented, the gallery images are retrieved in ranked order by computing similarities with the query image in the feature domain (image matching).

Many methods of feature extraction and image matching have been developed. In feature extraction, many various feature descriptions have been developed, specifically in the category of MPEG-7 [1]. In image matching, most researches have been focused on the reduction of searching time [2] and the fusion of multiple features [3].

However, based on our knowledge, the effort for improving matching accuracy in the image matching stage has not been tried yet. While other approaches have tried to reduce searching time or to develop efficient fusion methods of multiple features, our interest is on the issue how to improve the retrieval accuracy without significant burden for extra searching time. That is, assuming a set of ranked retrieved images with their corresponding similarity measures and feature vectors are already given, we are trying to improve the retrieval result recursively in terms of rank and the number of retrieved ground truth images.

Our idea is based on the observation that the first-ranked retrieval image, which is the most similar image with the query image, may be another version of the query image. That is, if feature extraction is good enough to search the image content, we can assume the difference between the query image and the first-ranked retrieval image is mostly due to environment variation, not the image content. So, we can use the first-ranked images as new query images in a recursive way: new queries are recursively found and evaluated and re-sorted until a criterion is met. The recursive process is called "the recursive matching" in this work.

Section 2 and 3 describe the system overview of the proposed image retrieval system and the recursive matching, respectively. Section 4 gives the experimental results with its application to face image retrieval and Section 5 concludes.

2. System Overview

The block diagram of an image retrieval system with the recursive matching step is shown in Figure 1. It consists of two parts: server and client applications. The server part extracts and stores feature vectors from gallery images, and it returns the retrieved images when a query is given by the request of the client part. The client part with the recursive matching part improves the retrieval accuracy by refining the similarity measure.

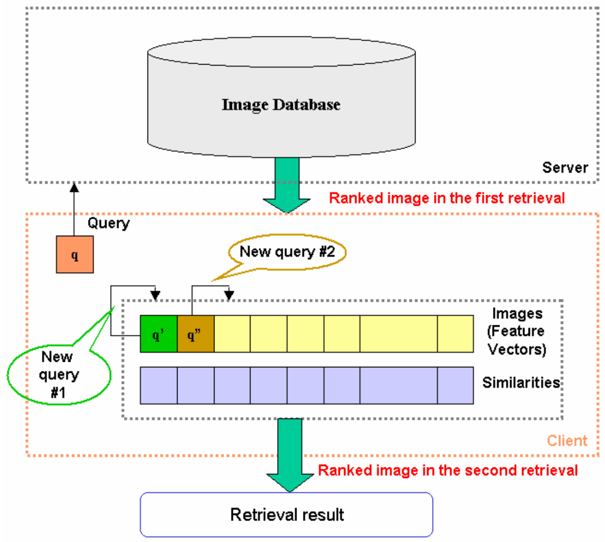


Figure 1. The block diagram of an image retrieval system.

Here are the implementation details of the recursive matching part of the server. The implementation is done by a special buffer, which contains feature vectors and similarity scores of the retrieved images. Within the fixed buffer, the similarity scores are recursively refined without any communication with the server and other parts. Note that the computation in the buffer doesn't give too much burden to overload the computation and memory. Finally, the system provides the refined ranked images owing the recursive matching.

3. Recursive Matching

When an image is encoded as a feature vector for image retrieval, the mathematical representation of the image contains environmental variation like lighting or pose change, and its image characteristic itself at the same time. Therefore, when a query image is given, the retrieved images reflect the encoding error of the query image by the environmental variation. Some images are retrieved by the similar variation with the query image.

Assuming that most of the first ranked retrieval image is mainly found by the image characteristic, the first ranked image contains same characteristic and different environmental variation, that is, very similar characteristic of the query image chosen by users. That means that it can be used another query image, called "the new query". Considering the case the first-ranked retrieval image having different characteristics with the query image, the effect of the second retrieval by "the new query " is weakened by weighting the matching scores (distance from feature space, or feature distance) between the original score and the new score. The process can be done recursively, so it is called "the recursive matching" or "the recursive retrieval."

Here is the mathematical formulation and algorithm description for the recursive matching. For convenient, we will confuse an image with its corresponding feature vector, which is the mathematical representation of an image for image retrieval. Suppose that, given a query image q, we would like to retrieve K images in ordered rank from an image database $\{I_i \mid i=1,\cdots,N\}$ of size N.

In the first step, from the query image, we can have sorted image set $\{q_i \mid i=1,\cdots,M\}$ of size M from the image database, and they are stored into a buffer of size M. It includes their corresponding score array $\{s(i) \mid i=1,\cdots,M\}$, which is represented by $s(i)=D(q,q_i)$, where $D(a,b)$ denotes the feature distance between a and b. In the next step, the first-ranked retrieval image q_1 is selected as a new query q' and the matching and sorting procedure is repeated only within the buffer, not the whole database. As a result, we have the re-sorted image set $\{q'_i \mid i=1,\cdots,M\}$ and the corresponding score array $\{s'(i) \mid i=1,\cdots,M\}$. To reflect the matching by the new query, the score array is updated by

$$s'(i)=s(i)+w_1 \cdot D(q',q'_i)\,, \qquad (1)$$

where w_1 denotes a weighing constant. This procedure is performed recursively. In n-th step, we have a sorted data set $\{q_i^{(n-1)} \mid i=1,\cdots,M\}$ with corresponding score array $\{s^{(n-1)}(i) \mid i=1,\cdots,M\}$, where

$$s^{(n-1)}(i)=s^{(n-2)}(i)+w_{n-1} \cdot D(q^{(n-1)},q_i^{(n-1)})\,. (2)$$

After iterations, the final ranked images of size K are retrieved and it is return as an output. If the size of buffer is smaller than that of the requested retrieval images (that is, $M < K$), the images over buffer size are saved in the first iteration and they are attached at the end of the buffer output. In this work, we set $w_{n-1} = (w_1)^{n-1}$ and $w_1 = w$, that is, exponentially decreasing effect.

4. Application to Face Retrieval

We applied this retrieval to our face retrieval system to show the effectiveness of the proposed algorithm. In Figure 2, our face retrieval system is shown and it retrieves the most similar face images in order from specified image database based on given face description algorithm.

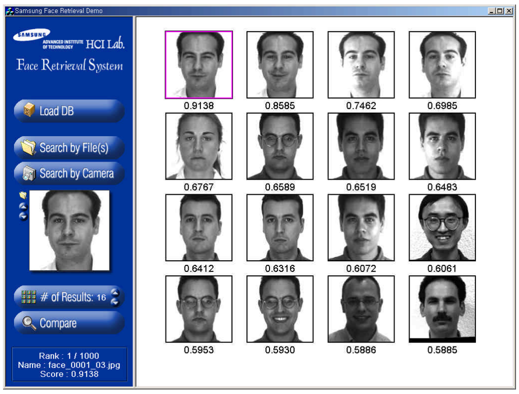


Figure 2. A face retrieval system.

4.1. Face Descriptor

For face retrieval, at first, a face image should be represented in a vector form to describe its identity information and the vector representation is called feature vector. To compare the similarity between two face images, the distance or correlation is computed between their corresponding feature vectors. Practically, a face retrieval system should the following requirement. That is, the dimension of extracted feature vectors and the computational complexity should be as small as possible, achieving good retrieval accuracy enough to find similar face images in huge database.

In [5], a component-based face descriptor using LDA (Linear Discriminant Analysis) and a simple pose classification were presented and we used this face descriptor for our experiments. The dimension of feature vector is 240 bits per each face image. As a similarity measure, Euclidean distance between feature vectors is used.

4.2. Experimental Conditions

As a dataset, the MPEG-7 face dataset is used. It consists of five databases: the extended version 1 MPEG-7 face database (E1), Altkom database (A2), MPEG-7 testset in XM2VTS database (M3), FERET database (F4), and MPEG-7 testset in the Banca database (B5). The details are described in Table 1**.**

Ref		
E1	The Extended version 1 MPEG-7 face database	635 persons (5 images per person exhibiting illumination and view variations)
A2	Altkom database	80 persons (15 images per person: 5views*3illuminations)
M3	MPEG-7 testset in XM2VTS database	295 persons (10 images per person: 5 views*2 different times (session 1&4)
F4	FERET database	875 persons 4000 images selected for the "background" at testing stage for both face image retrieval and personal identification
B5	MPEG-7 testset in the Banca database	52 persons (10 images per person: 4 * office, 4*outdoor,2*ideal; each image taken at different time).

Table 1. Face dataset.

The total number of images is 11845. 3655 images are used only for learning of LDA projections (training images), and 8190 images are used only for performance evaluation (test images). Among the test images, 4190 images are used as query images and 4000 images from F4 are used as obstruction images. Detailed list is presented in Table 2 and the ground truth images of all query images are known to evaluate the algorithm performance.

	DB	**Person**	**Image**	**Total**
Training Image	*A2*	40	15	600
	B5	-	-	-
	M1	317	5	1,585
	M3	147	10	1,470
	F4	-	-	-
Total		504		3,655

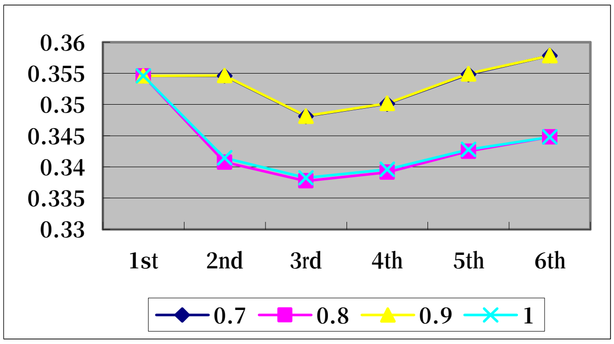
Test Image	**DB**	**Person**	**Image**	**Total**
	A2	40	15	600
	B5	52	10	520

	M1	318	5	1,590
	M3	148	10	1,480
	F4	-	-	4,000
Total		558		8,190

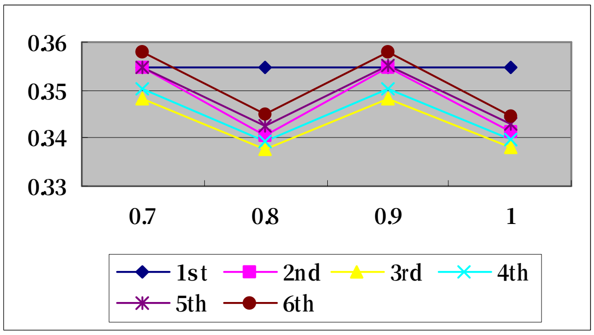
Table 2. Experimental condition.

4.3. Experimental Results

The retrieval accuracy is measured in terms of ANMRR (Average Normalized Modified Retrieval Rate) specified in [1], which evaluates two criteria simultaneously: how much of the correct images is retrieved, and how high they are ranked among the retrievals. It ranges from 0 to 1 and smaller value means better accuracy.



(a)



(b)

Figure 3. Experimental result. (a) The accuracy plot with respect to the weighting parameters. (b) The accuracy plot with respect to the number of iterations.

In the experiments, the size of dataset to be searched, N, are 8190, and the buffer size for the recursive matching, M, is 60. When we used the quick sort algorithm, the increased searching time is only around 0.007 (= $\frac{60\log 60}{8190\log 8190} \times 2$) times of the conventional algorithm (the multiplied figure 2 denotes the number of iterations for our best accuracy case).

Finally, the experimental results are presented in Figure 3. Note that the 1st iteration in this figure means the conventional retrieval without the recursive matching. Figure 3(a) is the accuracy plot of with respect to weighting parameters, $w = \{0.7, 0.8, 0.9, 1.0\}$. With regardless of weighting parameters, the 2nd iterations gave the best accuracy for this dataset. The degradation over the 3rd iterations seems to be due to the possible guess that the retrieved images are too far from the given query images. Table 3(b) is the accuracy plot with respect to the number of iterations, and it shows that at the weighting parameters 0.8 and 1.0 results in the best accuracy. The best accuracy was at the 3rd iterations (additional 2 iterations) with weighting parameter 0.8. Compared with the conventional retrieval (1st retrieval) the improvement was 0.0168 in terms of ANMRR (from 0.3546 to 0.3378).

5. Concluding Remarks

In this paper, we presented the recursive matching giving retrieval improvement with ignorable computational cost. It was applied to face image retrieval and the experimental results showed the feasibility. The future work is to extend the algorithm to other image retrieval applications. Importantly, more theoretical analysis should be continued.

References

[1] B. S. Manjunath, Philippe Salembier, and Thomas Sikora, "*Introduction to MPEG-7: Multimedia Content Description Interface*," John Wiley & Sons Ltd., 2002.

[2] W. Hwang and J. Weng, "*Hierarchical Discriminant Regression*," IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 22, no. 11, November 2000, pp. 1277-1293.

[3] C. Breitneder, and H. Eidenberger, "*Automatic Query Generation for Content-Based Image Retrieval*", IEEE Conference on Multimedia and Expo, Jul. 30--Aug. 2, 2000, vol. 2, pp. 705-708.

[4] N. Doulamis, and A. Doulamis, "*Optimal recursive similarity measure estimation for interactive content-based image retrieval*," Proceedings. Of International Conference on Image Processing. 2002, Volume: 1, 2002. Page(s): 972 –975

[5] Tae-Kyun Kim, Hyunwoo Kim, Wonjun Hwang, Seok Cheol Kee, and Jong Ha Lee, "Component-based LDA Face Descriptor for Image Retrieval," Proc. of the 13th British Machine Vision Conference, Cardiff, UK, 2002.