LETTER

An Efficient Filtering Method for Scalable Face Image Retrieval

Deokmin HAAM[†], *Nonmember*, **Hyeon-Gyu KIM**[††a)], *Member*, *and* **Myoung-Ho KIM**[†], *Nonmember*

SUMMARY This paper presents a filtering method for efficient face image retrieval over large volume of face databases. The proposed method employs a new face image descriptor, called a *cell-orientation vector* (COV). It has a simple form: a 72-dimensional vector of integers from 0 to 8. Despite of its simplicity, it achieves high accuracy and efficiency. Our experimental results show that the proposed method based on COVs provides better performance than a recent approach based on identity-based quantization in terms of both accuracy and efficiency.

key words: *face image retrieval, face detection, image filtering, inverted index, cell-orientation vector*

1. Introduction

Face image retrieval has a wide range of applications, including name-based face image search, face tagging in images and videos, face identification in criminal investigation, copyright enforcement, and so on [1]. In these applications, when a face image is given as a query, images containing the same person appearing in the query should be retrieved from a face database and reported to users.

In the image retrieval over large volume of databases, existing discriminative facial features [2]–[4] cannot be used properly. This is because these features are typically very high-dimensional, and thus they are not suitable for quantization and inverted indexing [5]. In other words, using such features requires a linear scan of the whole database. This is definitely prohibitive for scalable image retrieval.

A straightforward approach for scalable image retrieval is to use the *bag-of-visual-words* representation that has been adopted in many image retrieval systems [6]–[8]. In this approach, an image is treated as a document containing visual words, each of which is represented as a vector capturing local features in the image. Then, an inverted index can be built over a set of images using the visual word vectors. With the index, image retrieval can be performed in a timely manner, as in the information retrieval (IR) approach.

For example, suppose a face image is given as a query. Then, visual word vectors are first extracted from the query image. With the vectors, top-k images containing similar visual words are found from an inverted index. To find an exact match, further similarity checking can be performed over the candidate image set, using a more accurate detection method (adopting a machine learning technique). In this case, the index plays a role of *filtering* images to find candidates quickly. On the other hand, the candidate images themselves can be reported as an answer in some applications. In any case, the indexing mechanism is essential for rapid retrieval over large-scale image databases.

More recent study [5], [9] discussed a method to improve retrieval accuracy by considering spatial information when constructing visual words. As an extreme case, visual words capturing the nose of a person can match the words capturing the mouth of another person in the above approach. To avoid such inaccuracy, similarity comparison can be performed only for vectors capturing the same geometric regions in any two images.

In this paper, we propose a more accurate filtering method for fast retrieval of target face images from large-scale databases. The goal of the proposed method is to find a set of candidates which contains faces of the same person appearing in the query image, while keeping the size of the set as small as possible. Note that higher filtering accuracy leads to faster retrieval of query answers. This is because, as filtering accuracy increases, the size of a candidate image set (i.e., k) is reduced, and time to scan the set and find an exact match can be saved from it.

For this purpose, a new face image descriptor, called a *cell-orientation vector* (COV), is introduced in the proposed method. It is in the form of a 72-dimensional vector of intergers from 0 to 8. Here, each dimension corresponds to a predefined subregion of a face image. It achieves good performance in terms of both accuracy and efficiency, which we observed through our experiments. In what follows, we discuss the process to compute the COV from a query image and to perform image filtering based on the COVs.

2. Proposed Method

2.1 COV Generation

Given a face image, the proposed method first extracts visual features from square regions around two eyes, which we call *eye regions*. To detect eye regions accurately from the image, some *preprocessing* steps are required because the face images can be captured with various angles and sizes.

Manuscript received July 30, 2014.
Manuscript revised October 12, 2014.
Manuscript publicized December 11, 2014.

[†]The authors are with the Department of Computer Science, KAIST, 373–1 Guseong-Dong, Yuseong-Gu, Daejeon, 305–701 Republic of Korea.
[††]The author is with the Department of Computer Engineering, Sahmyook University, Hwarangro 815, Nowon-gu, Seoul, 139–742 Republic of Korea.
a) E-mail: hgkim@syu.ac.kr (Corresponding author)

DOI: 10.1587/transinf.2014EDL8156

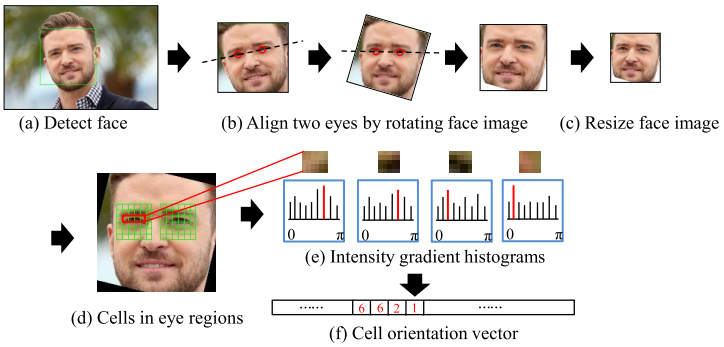

Fig. 1 COV generation process

For example, consider the image in Fig. 1 which illustrates the COV generation process. Given the image, a face region is first extracted from it using a face detector based on Haar-like features [10]. Since the detected face may vary in angle and size, it is required to be normalized. For this purpose, locations of two eye points are calculated. If the points are not horizontally placed, the image is rotated. Then, it is fit to a predefined image size, i.e., 160×160 pixels in our approach. Figures 1 (a) to (c) show these preprocessing steps to get a normalized face image.

From the normalized image, eye regions are extracted, as shown in Fig. 1 (d). In the proposed method, an eye region is defined as a square matrix consisting of 6×6 subregions, each of which is called a *cell*. Each cell again consists of 8×8 pixels.

As a feature for representing a cell, we use gradients of pixels in the cell. The gradient of a pixel at $(i,\ j)$ position is $\nabla f(i,\ j) = (\frac{\delta f(i,\ j)}{\delta x}, \frac{\delta f(i,\ j)}{\delta y})$, where $\frac{\delta f(i,\ j)}{\delta x}$ $(\frac{\delta f(i,\ j)}{\delta y})$ is the gradient of the pixel at $(i,\ j)$ in x (y) direction. We compute the gradient by using simple 1-D $[-1,\ 0,\ 1]$ masks as follows: $\frac{\delta f(i,\ j)}{\delta x} = \frac{I(i+1,\ j)-I(i-1,\ j)}{2}$ and $\frac{\delta f(i,\ j)}{\delta y} = \frac{I(i,\ j+1)-I(i,\ j-1)}{2}$, where $I(i,\ j)$ is the intensity of the pixel at $(i,\ j)$. Based on the gradient, the gradient orientation is computed by $\tan^{-1}(\frac{\delta f(i,\ j)}{\delta y} / \frac{\delta f(i,\ j)}{\delta x})$, and its magnitude is computed by $\sqrt{(\frac{\delta f(i,\ j)}{\delta x})^2 + (\frac{\delta f(i,\ j)}{\delta y})^2}$. Then, each pixel is assigned a gradient with $(r,\ \theta)$, where r and θ denote its magnitude and orientation ($0^\circ \leq \theta \leq 180^\circ$), respectively. The orientation θ is again mapped to an orientation index i, such that $i\pi/9$ is the closest one to θ in $\{0\pi/9, 1\pi/9, \ldots, 8\pi/9\}$, as discussed in HOG descriptor [11].

Based on the gradients of pixels, an *intensity gradient histogram* is built for each cell, as shown in Fig. 1 (e). The histogram consists of 9 bins, where the i-th bin has the sum of gradient magnitudes of pixels in the cell whose orientation index is i. Let $p_j(r,\ \theta)$ denote the j-th pixel with gradient $(r,\ \theta)$. Suppose there are n pixels in a cell. Then, the i-th bin of the histogram, denoted by h_i, is defined as follows.

$$h_i = \sum_{j=0}^{n} x, \text{ where } x = \left\{ \begin{array}{ll} r & \text{if } \theta = i \\ 0 & \text{otherwise} \end{array} \right\} \text{ in } p_j(r, \theta)$$

From the histogram, an orientation index with the largest h_i is chosen as a feature to represent the cell. The selected feature reflects the prominent tendency of directional change in intensity. Since 72 cells are extracted from an image in the proposed method, 72 orientation indexes are available after finishing the computation. These values are finally encoded into a COV, which is a 72-dimensional vector of intergers from 0 to 8. Figure 1 (f) shows an example of the COV generated from features in Fig. 1 (e).

In this paper, we use only features around eyes when generating COVs, and the proposed method has good filtering performance as will be shown in the experiments. To improve filtering accuracy of the proposed method, we plan to extend COVs by considering more facial components, e.g., nose and mouth.

2.2 Face Image Filtering Based on COVs

In the proposed method, COVs are used to check the degree of relevance between two face images. Let v_i^α denote the dimension-i of the COV extracted from the α-th image in a database. In our approach, the degree of relevance between any two α-th and β-th images, denoted $f_{\alpha,\beta}$, is defined as follows.

$$f_{\alpha,\beta} = \sum_{i=0}^{71} x, \text{ where } x = \left\{ \begin{array}{ll} 1 & \text{if } v_i^\alpha = v_i^\beta \\ 0 & \text{otherwise} \end{array} \right.$$

Given a query image, our goal is to find a small set of candidates, which contains faces of the same person appeared in the query image. To enable this, COVs of the database images should be computed in advance. Then, an inverted index must be built over the COVs to improve search performance. In our approach, the index consists of 72 arrays, where the i-th array corresponds to the dimension-i of the COV, as shown in Fig. 2. The size of an array is 9, whose index j corresponds to an orientation $j\pi/9$. Each array element has a list of corresponding images. More specifically, the j-th element of the i-th array has a list of IDs pointing to the images, whose COVs have j in the dimension-i, i.e., $j\pi/9$ is the promient gradient orietation of the i-th cell.

Using the index, image filtering is performed. Given a query image, its COV is first computed. Let v_i^q denote the

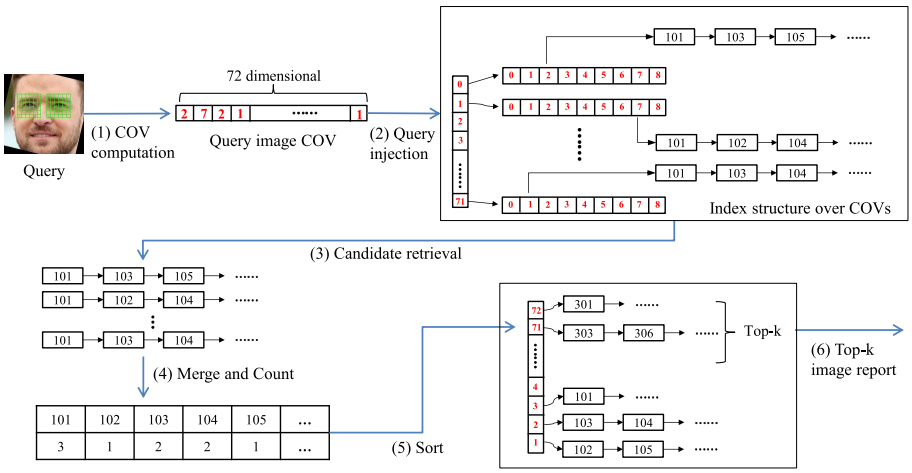

Fig. 2 Example of image filtering in the proposed method.

dimension-i of the query image COV, and A_i denote the i-th array in the inverted index. Then, for each i ($0 \leq i \leq 71$), v_i^q is inputted to the index as a query word. Suppose that the value of v_i^q is j. In this case, a list of $A_i[j]$ is returned.

After the queries are completed for all v_i^q, 72 lists are obtained. These lists are merged to compute the degree of relevance of the images, which are relevant to the query image. The merge is performed based on the image IDs in the lists. For an image ID, say α, we check how many times α occurs in the lists. This count is identical to the degree of relevance $f_{q,\alpha}$, where q means the query image. Based on the counts, images are sorted, and the top-k images with the highest relevancy are chosen as a set of candidates.

Figure 2 shows an example of the filtering process in the proposed method, which consists of 6 major steps. Given a query image, its COV is first computed. Then, it is injected into the inverted index to retrieve relevant images. In the example, $v_0^q = 2$, $v_1^q = 7$, and so on. Thus, 72 lists pointed by $A_0[2]$, $A_1[7]$, ..., $A_{71}[1]$ are returned. For each relevant image, i.e., image in the lists, the number of its occurrences is counted. Based on the counts, images are sorted in the decreasing order of their occurrences.

For efficiency, an array can be used for the sort. Since an image can occur up to 72 times in the lists, the array can be organized to have 72 elements, where the i-th element has a list of images whose occurrences are equal to i in the lists. Using the array, sorting can be performed in a linear time (by array indexing). After sorting is finished, top-k images are reported.

3. Experimental Results

For the experiments, about 75,000 face images were gathered from the Web and public databases, including Averdeen face database, The MUCT Face Database, nottfaces-originals, and Utrecht ECVP [12]. From the public databases, 1,000 images were randomly chosen as query images, while others were served as a target database for query.

Based on the database, we first checked the filtering accuracy of the proposed method. For this purpose, it was compared with a recent approach based on identity-based quantization [5]. The IB-based method was chosen because, to the best of our knowledge, it is only the recent approach that addresses the issue of scalable face image retrieval. To check performance of the IB-based method, we constructed a vocabulary where its images were randomly chosen from the database. The set consisted of 250 different persons, each of which had 50 face images. This setting is similar to the configuration of experiments used to show performance of the IB-based method in identity-based quantization [5]. For the comparison, we varied k from 500 to 4,000, where k is the size of a set with candidate images, then observed whether a query image is included in the set. Our experiments were conducted on an Intel Core2 Duo 3.0 GHz machine, running Windows 7 with 4 GB memory.

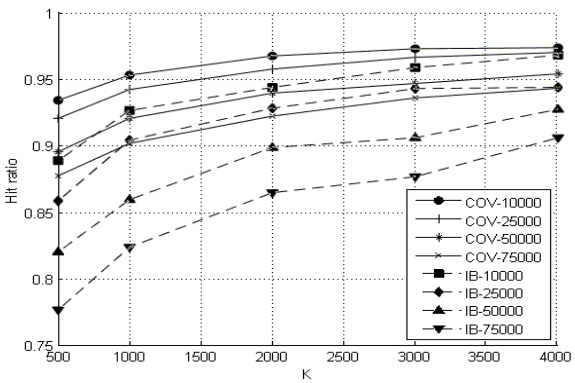

Fig. 3 Query hit ratio: COV-based vs. IB-based approaches.

Figure 3 shows its results, where the proposed and the existing methods are denoted as *COV* and *IB*, respectively. The number behind the COV or IB denotes the size of the target database. For each k, we examined the four cases by changing the size from 10,000 to 75,000. The performance of the two approaches was represented as a percentage of *query hits*; a query is *hit* if its answer is included in the candidate image set. In the figure, the percentage is denoted as

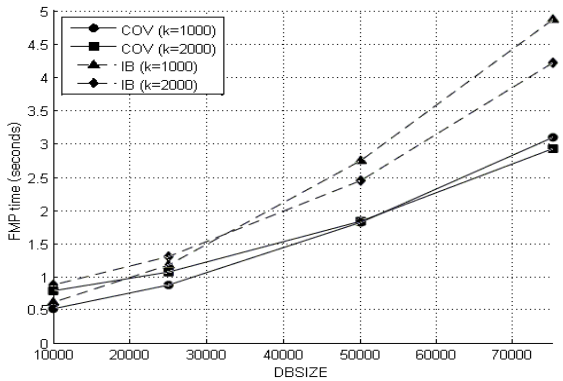

Fig. 4 FMP time: COV-based vs. IB-based approaches.

hit ratio.

In both approaches, hit ratios increased as k increased. This is because the probability that an answer is included in the candidate set increases as the size of candidate set, i.e., k, increases. On the other hand, their hit ratios were inpropoertional to the size of the database. This is due to that, as more images are included in the database, the number of images which are similar to the query but not answers also increases. Such situation gives negative influence on accuracy of image filtering in both methods.

Note that hit ratios of the proposed method are higher than those of the IB-based approach in all cases. Its hit ratios are kept higher than 0.9, only except the two test cases: COV-50000 and COV-75000 when $k = 500$. This is when the size of the candidate image set is less than or equal to 1% of the target database size.

Figure 4 compares the time to find an exact match from the database in the two approaches. The time is denoted as *FMP time*, an abbreviation of the *first matching proving time*. It consists of (1) feature extraction from the query (face detection time, facial components detection time and the feature extraction time), (2) time to make a candidate image set and compare the query with the candidates, and (3) time to scan a whole database when the filtering is failed.

In terms of the FMP time, the proposed method is about 30% faster than the IB-based approach, as shown in the figure. This result shows that the proposed method can be applied to various applications where rapid image retrieval is required, such as criminal investigation.

4. Conclusion

In this paper, we proposed a novel filtering method for efficient face image retrieval over large-scale databases. The proposed method is based on a new face image descriptor, called the cell orientation vector (COV). It has a simple form: a 72-dimensional vector of integers from 0 to 8. Despite its simplicity, it achieves high accuracy and efficiency. To verify this, experiments were conducted over about 75,000 face images gathered from the Web and public databases, and the proposed method was compared with a recent approach based on identity-based quantization over the database. In the results, the proposed method provided more accurate results than the IB-based approach in all test cases. In particular, its retrieval accuracy was kept higher than 90%, with the only exception being the case where the size of the candidate image set is less than or equal to 1% of the target database size. In addition, our method was about 30% faster than the IB-based approach in the exact image retrieval. These results show the superiority of the proposed method based on the COVs.

Acknowledgements

This research was supported by the MSIP (Ministry of Science, ICT and Future Planning) of Korea under the ITRC support program (NIPA-2013-H0301-13-4009), and the National Research Foundation of Korea grant funded by the Korea government (MEST) (No. 2012R1A2A2A01046694).

References

[1] A.F. Abate, M. Nappi, D. Riccio, and G. Sabatino, "2D and 3D face recognition: A survey," Pattern Recognit. Lett., vol.28, no.14, pp.1885–1906, 2007.
[2] O. Déniz, G. Bueno, J. Salido, and F. De la Torre, "Face recognition using histograms of oriented gradients," Pattern Recognit. Lett., vol.32, no.12, pp.1598–1603, 2011.
[3] G. Hua and A. Akbarzadeh, "A robust elastic and partial matching metric for face recognition," IEEE Int. Conf. on Computer Vision, pp.2082–2089, 2009.
[4] W. Biao, Y. Wenming, and L.I. Weifeng, "Two-stage block-based whitened principal component analysis with application to single sample face recognition," IEICE Trans. Inf. & Syst., vol.E95-D, no.3, pp.853–860, March 2012.
[5] Z. Wu, Q. Ke, J. Sun, and H.Y. Shum, "Scalable face image retrieval with identity-based quantization and multireference reranking," IEEE Trans. Pattern Anal. Mach. Intell., vol.33, no.10, pp.1991–2001, 2011.
[6] J. Sivic and A. Zisserman, "Video Google: A text retrieval approach to object matching in videos," Proc. IEEE Int. Conf. on Computer Vision, pp.1470–1477, 2003.
[7] P. Pengyi and S. Kamata, "Hilbert scan based bag-of-features for image retrieval," IEICE Trans. Inf. & Syst., vol.E94-D, no.6, pp.1260–1268, June 2011.
[8] D. Nister and H. Stewenius, "Scalable recognition with a vocabulary tree," Proc. IEEE Int. Conf. on Computer Vision and Pattern Recognition, vol.2, pp.2161–2168, 2006.
[9] B.C. Chen, Y.H. Kuo, Y.Y. Chen, K.Y. Chu, and W. Hsu, "Semi-supervised face image retrieval using sparse coding with identity constraint," ACM Int. Conf. on Multimedia, pp.1369–1372, 2011.
[10] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," IEEE Int. Conf. on Computer Vision and Pattern Recognition, vol.1, p.I-511, 2001.
[11] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," IEEE Int. Conf. on Computer Vision and Pattern Recognition, vol.1, pp.886–893, 2005.
[12] PICS Face Database, http://pics.stir.ac.uk