Robust visual similarity retrieval in single model face databases

Yongseng Gao*, Yutao Qi

School of Microelectronic Engineering, Nathan Campus, Griffith University, QLD 4111, Australia

Received 20 May 2004; received in revised form 20 December 2004; accepted 20 December 2004

Abstract

In this paper, we introduce a novel visual similarity measuring technique to retrieve face images in photo album databases for law enforcement. Though much work is being done on face similarity matching techniques, little attention is given to the design of face matching schemes suitable for visual retrieval in single model databases where accuracy, robustness to scale and environmental changes, and computational efficiency are three important issues to be considered. This paper presents a robust face retrieval approach using structural and spatial point correspondence in which the directional corner points (DCPs) are generated for efficient face coding and retrieval. A complete investigation on the proposed method is conducted, which covers face retrieval under controlled/ideal condition, scale variations, environmental changes and subject actions. The system performance is compared with the performance of the eigenface method. It is an attractive finding that the proposed DCP retrieval technique has performed superior to the eigenface method in most of the comparison experiments. This research demonstrates that the proposed DCP approach provides a new way, which is both robust to scale and environmental changes, and efficient in computation, for retrieving human faces in single model databases.

© 2005 Pattern Recognition Society. Published by Elsevier Ltd. All rights reserved.

Keywords: Content-based image retrieval; Face database; Single model database; Face matching; Similarity measuring; Directional corner point

1. Introduction

In recent years, image retrieval by visual content from database has become a major research area due to the ever-increasing rate at which images are generated in many information systems [1]. A wide range of image data has established its relevance in the areas of medicine, industry, art and law enforcement. Most practices in querying and retrieving images are often based on queries by filenames, captions and keywords. This does not satisfy the requirements of modern image retrieval. Content-based image retrieval, which retrieves images by visual contents in an image, has attracted much attention in recent researches. Visual retrieval of face images, such as album search that helps police officers identify a suspect by searching a country or state photo database, is of particular interest in law enforcement applications. Bach et al. [2] first attempted this problem using an interactive manner. Automatic similarity retrieval in a face database presents a significant challenge to researchers due to the small inter-class variation in the face database. Human faces are very similar in structure with minor differences from person to person. Furthermore, scale variations, appearance changes due to environmental changes (i.e., lighting condition changes) and subject actions (i.e., head pose variations and facial expressions) increase the intra-class variation. These factors further complicate the face retrieval task as one of the most difficult problems in image retrieval.

Computerized human face recognition has been an active research area for the last few decades because it has many practical applications, such as access control, bankcard verification, mug shot searching, security monitoring, and
surveillance systems. Different applications present different constraints and need different processing requirements. Though much work is being done on face similarity matching techniques [3–7], little attention is given to the design of face matching schemes suitable for visual retrieval in single model databases where accuracy, robustness to scale and environmental changes, and computational efficiency are three important issues to be considered. For image retrieval, the method employed needs to satisfy following conditions: (1) It must be suitable for retrieval from single model databases because the face databases in law enforcement applications contain only one photo per person. (2) It should be robust to scale and environmental changes. Face detection and localization is a prior stage for automatic recognition of faces. It is inevitable that a face is localized correctly with minor scale variation. Thus, it is an important requirement that a face matching algorithm is robust to scale variations. The robustness to environmental changes (i.e., lighting condition changes) is another critical issue for face image retrieval, such as photo album searching. When collecting the image data, subjects are usually required to be cooperative to have a fronto-parallel head pose with neutral expression such that the effect of appearance changes due to subject actions is minimized. However, the lighting conditions between the query image and the models are usually different because they are taken at different time and in different places. The environmental changes could thus be the key remaining problem in this situation. (3) It must be fast in computation.

One of the pioneering studies on automated face recognition was using geometrical features done by Kanade [8] in 1973. Their system achieved a peak performance of 75% recognition rate on a database of 20 people using two images per person, one as the model and the other as the test image. Goldstein et al. [9], and Kaya and Kobayashi [10] showed that a face recognition program provided with features extracted manually could perform recognition with apparently satisfactory result. Bruneli and Poggio [5] automatically extracted a set of geometrical features from the picture of a face, such as nose width and length, mouth position and chin shape. There were 35 features extracted to form a 35-dimensional vector. The recognition was then performed with a Bayes classifier. They reported a recognition rate of 90% on a database of 47 people. Cox et al. [11] introduced a mixture-distance technique, which achieved 95% recognition rate on a query database of 685 individuals. Each face was represented by 30 manually extracted distances. Manjunath et al. [12] used Gabor wavelet decomposition to detect feature points for each face image, which greatly reduced the storage requirement for the database. Typically 35–45 feature points per face were generated. The matching process utilized the information present in a topological graphic representation of the feature points. After compensating for different centroid location, two cost values, the topological cost and similarity cost, were evaluated. The recognition accuracy in terms of the best match to the right person was 86% and 94% of the cases that the correct person’s face was in the top three candidate matches. Geometrical feature matching based on precisely measured distances between features may be most useful for finding possible matches in a large database such as mug shot album. Eigenface [6] is one of the most well-known face matching approaches. It is also known as principal component analysis (PCA). Kirby and Sirovich [13] initially used PCA to efficiently represent pictures of human faces. Turk and Pentland introduced eigenfaces [6] and eigenspaces [14] for face detection and identification. Any face image could be approximately reconstructed by a collection of weights for the face and a standard face picture, i.e. eigenpicture. The weights describing each face are used for similarity matching. The eigenface approach appears to be a fast and practical method for face image retrieval. However, the basis images developed by PCA depend only on second-order images statistics [15]. Bartlett et al. [15] used independent component analysis (ICA), which is sensitive to higher-order images statistics, in their face recognition system. Better performances were reported under various testing conditions. Neural networks [4,16–18] have been successfully applied to face recognition to tackle the problem of appearance changes and scale variations. Lawrence et al. [4] proposed a hybrid neural network, which combined local image sampling, a self-organizing map (SOM) neural network, and a convolutional neural network. The SOM provides a quantization of the image samples into a topological space. The convolutional network extracts successively larger features in a hierarchical set of layers, and provides partial invariance to translation, rotation, scale and deformation. The authors reported 96.2% correct recognition with the ORL database of 400 images of 40 individuals. The classification time was less than 0.5 s, but the training time was as long as 4 h. Lin et al. [17] used probabilistic decision-based neural network (PDBNN), which inherited the modular structure from its predecessor, a decision-based neural network (DBNN) [18]. A hierarchical neural network structure with non-linear basis functions and a competitive credit-assignment scheme was adopted. PDBNN-based biometric identification system has the merits of both neural networks and statistical approaches, and its distributed computing principle is relatively easy to implement on a parallel computer. In Ref. [17], it was reported that PDBNN face recognizer had the capability of recognizing up to 200 people, and could achieve up to 96% correct recognition rate in approximately 1 s. However, when the number of persons increases, the computing expense will become more demanding. For face verification, a multiresolution pyramid structure called Cresceptron [16] was developed. In general, neural network approaches encounter problems when the number of classes (i.e., individuals) increases. Moreover, they are not suitable for single model image retrieval tasks because multiple model images per person are required for training the systems to optimal parameter setting. The line edge map (LEM) matching technique [19] was recently used to deal with the difficulty of
face recognition under lighting condition changes. Though the LEM technique is relatively tolerant to lighting condition changes and suitable for retrieval in single model databases, its computational expense is still too high for image retrieval of face databases.

This paper proposes a new face description and similarity measuring technique for visual similarity retrieval in single model face databases, which is relatively robust to scale and environmental changes, and efficient in computation. Instead of using local operation of isolated points, the proposed method employs directional corner point (DCP) matching in which directional information showing connectivity to its neighbors is utilized in the point correspondence. Unlike neural network approaches that require multiple model faces per person to train the system to optimal setting, the DCP method is suitable for single model face database retrieval and is fast in computation. A complete feasibility investigation and evaluation for the proposed face retrieval method is conducted using two standard and public available face databases, which covers all conditions of human face retrieval. They are face retrievals under controlled/ideal condition, scale variations, environmental changes and subject actions. The system performance is compared with the performance of the eigenface method. This research demonstrates that the proposed DCP approach provides a new solution for visual similarity retrieval in single model face databases.

In the following, Section 2 presents a lighting-insensitive face descriptor, which incorporates structural information with spatial features. In Section 3, an efficient warping algorithm for visual similarity retrieval is proposed. Particular attention is given to the discriminative power to distinguish similar objects. In Section 4, the proposed system is extensively examined with various experimental conditions, and compared with a benchmark system. Finally, the paper concludes in Section 5.

2. Describing a face by DCP

Edges are the most fundamental features of objects in the 3-D world. The edges in an image reflect large local intensity changes due to the geometrical structure of the object, the characteristic of surface reflectance of the object, and the viewing direction. They have the advantages of less demand on storage space and less sensitive to illumination changes. However, edge maps utilize spatial information of an image but lack structural representation. For face image retrieval, the representation efficiency of edge curves needs to be further improved to cater for the speed requirement of database retrieval. On the other hand, we believe that the structural information indicating the connectivity of these edge points could enhance the face identity description for similarity matching. In this study, a new face feature descriptor, DCPs, is proposed to integrate the structural connectivity information with spatial features of a face image. After detecting the edge curves [20], a corner detection process [21] is applied to generate the DCPs of a face. A DCP \( P(x, y, \delta_1, \delta_2) \) is represented by its Cartesian coordinates \((x, y)\) and two-directional attributes \(\delta_1\) and \(\delta_2\) (Fig. 1). \(\delta_1\) is the angular value of “horn 1” that points to its anterior neighboring corner point \(M\). Similarly, \(\delta_2\) is the angular value of “horn 2” that points to its posterior neighboring corner point \(N\). \(\delta_1\) and \(\delta_2\) range from 0° to 360°. If a DCP is a start point of a curve, such as point \(M\) in Fig. 1, a null is assigned to \(\delta_1\). If a DCP is an end point of a curve, such as point \(L\) in Fig. 1, a null is assigned to \(\delta_2\). An example of human face DCPs is illustrated in Fig. 2. A DCP is either a corner point with two “horns” pointing to its two neighboring DCPs or a start/end point of the edge curve with a single “horn” pointing to its neighboring DCP. These “horns” provide isolated feature points with additional structural information about the connectivity to their neighbors. The DCP descriptor, using sparse points, further reduces the storage demand of an edge map and thus improves the computational efficiency to meet the high-speed requirement in visual retrieval of face databases. On the other hand, the structural attributes on the points enhance the discriminative power of the descriptor to improve the retrieval accuracy. The DCP descriptor is expected to be less sensitive to illumination changes due to the fact that it is a feature derived from low-level illumination-insensitive edge map representation.

![Fig. 1. An illustration of DCPs.](image-url)
3. DCP correspondence

Based on above image coding, a face is represented by a DCP descriptor that is a set of DCPs in a 4-D feature space of location and directions. The face retrieval process locates the face in the query image, generates the DCP descriptor of the query face and calculates the differences between the query DCP descriptor and the model descriptors in the database. The model in the database with minimum difference is considered as the correct return. Here, we propose a warping process to establish the correspondence between two DCP descriptors. The difference between two faces is measured by a cost function of global warping.

3.1. Point-to-point warping

Let \( A(x^A, y^A, \delta^A_1, \delta^A_2) \) and \( B(x^B, y^B, \delta^B_1, \delta^B_2) \) be two DCPs. A 3-step warping process, which consists of translation, rotation and open/close operations, is used to establish the correspondence from \( A \) to \( B \).

1. **Translation operation**: A translation operation from \( A \) to \( B \), denoted as \( T(A \rightarrow B) \), moves \( A \) to the location of \( B \) such that \( x^A = x^B \) and \( y^A = y^B \) (Figs. 3(a) and (b)). The cost function for a translation operation from \( A \) to \( B \) is defined as

\[
C[T(A \rightarrow B)] = \sqrt{(x^A - x^B)^2 + (y^A - y^B)^2}.
\]

2. **Rotation operation**: A rotation operation from \( A \) to \( B \), denoted as \( R(A \rightarrow B) \), rotates \( A \) anticlockwise by \( \theta \) degree until the right “horn” of \( A \) coincides with the right “horn” of \( B \) (Figs. 3(b) and (c)). The “horn” of a double-horn DCP is defined as the right “horn” if one can rotate that “horn” anticlockwise to the other with an angle less than \( 180^\circ \). The cost function for a rotation operation from \( A \) to \( B \) is defined as

\[
C[R(A \rightarrow B)] = \begin{cases} 
\theta & \text{if } \theta \leq 180^\circ, \\
360^\circ - \theta & \text{if } 180^\circ < \theta \leq 360^\circ.
\end{cases}
\]

3. **Open/close operation**: An open (or close) operation from \( A \) to \( B \), denoted as \( O/C(A \rightarrow B) \), opens (or closes) the two “horns” of \( A \) by \( \psi \) degree until the two “horns” of \( A \) coincides with the corresponding “horns” of \( B \) if the intersecting angle between the two “horns” of \( A \) is smaller (or greater) than that of \( B \) (Figs. 3(c) and (d)). The cost function for an open/close operation from \( A \) to \( B \) is defined as

\[
C[O/C(A \rightarrow B)] = \psi.
\]

Let \( A \rightarrow B \) denote a warping from \( A \) to \( B \). The cost function for warping \( A \) to \( B \) is defined as a combined cost of above three operations in Eq. (4).

\[
C(A \rightarrow B) = \sqrt{C^2[T(A \rightarrow B)] + f^2[C[R(A \rightarrow B)] + C[O/C(A \rightarrow B)]].}
\]

\( f(\cdot) \) is a non-linear function to map the angle to a scalar. It is desirable to ignore small angle variation, which is most likely segmentation error or intra-class variation, but penalize heavily on large deviation, which is most likely inter-class difference. In this study, a quadratic function

\[
f(\psi) = \frac{\psi^2}{W}
\]

is used, where \( W \) is the weight to be determined experimentally in Section 4.1.
By substituting $A(x^A, y^A, \delta^A_1, \delta^A_2)$ and $B(x^B, y^B, \delta^B_1, \delta^B_2)$ into Eq. (4), we have

$$C(A \rightarrow B) = \sqrt{(x^A - x^B)^2 + (y^A - y^B)^2 + f^2 [\min \{\Delta(\delta^A_1, \delta^B_1), \Delta(\delta^A_2, \delta^B_2), \Delta(\delta^A_1, \delta^B_2), \Delta(\delta^A_2, \delta^B_1)\}].} \quad (6)$$

where

$$\Delta(\delta^A_i, \delta^B_i) = \begin{cases} |\delta^A_i - \delta^B_i| & \text{if } |\delta^A_i - \delta^B_i| \leq 180^\circ, \\ 360^\circ - |\delta^A_i - \delta^B_i| & \text{if } 180^\circ < |\delta^A_i - \delta^B_i| \leq 360^\circ, \\ \end{cases}$$

and $\delta^k_{i, j} \in [0^\circ, 360^\circ)$.

For warping between a single-horn DCP and a double-horn DCP, one of the four $\delta^k_{i, j} \in [0^\circ, 360^\circ)$ is null which represents an indefinite direction. Thus $\Delta(\text{null}, \delta^B_j)_{j=1,2}$ or $\Delta(\delta^A_i, \text{null})_{i=1,2}$ is required for the calculation of Eq. (6). Since $\Delta(\delta^A_i, \delta^B_j)_{i,j=1,2}$ ranges from $0^\circ$ to $180^\circ$ as defined in Eq. (7), the maximum value, $180^\circ$, is assigned to $\Delta(\text{null}, \delta^B_j)_{j=1,2}$ or $\Delta(\delta^A_i, \text{null})_{i=1,2}$ to penalize warping a single-horn DCP to a double-horn DCP, or vice versa. It is desirable to prohibit a warping between two DCPs of different types. The cost function for warping between a single-horn DCP and a double-horn DCP is the same as Eq. (6).

For warping between two single-horn DCPs, a rotation operation from $A$ to $B$ rotates $A$ anticlockwise by $\theta$ degree until the single “horn” of $A$ coincides with the single “horn” of $B$. The cost function for a rotation operation between two single “horn” DCPs is of the same form as Eq. (2). It can be rewritten as Eq. (8) by substituting $A(x^A, y^A, \delta^A_1, \delta^A_2)$ and $B(x^B, y^B, \delta^B_1, \delta^B_2)$ into Eq. (2).

$$C[R(A \rightarrow B)] = \Delta(\delta^A_{\text{null}}, \delta^B_{\text{null}}), \quad (8)$$

where $\delta^A_{\text{null}}$ and $\delta^B_{\text{null}}$ are of non-null values which indicate the directions of the single-horns of $A$ and $B$, respectively.

The cost function for an open/close operation in this case is defined the same as $C[R(A \rightarrow B)]$ such that the value of $C[R(A \rightarrow B)] + C[O(A \rightarrow B)]$ is within the same range of $[0^\circ, 360^\circ)$ in all of the three warping cases (i.e., double-horn to double-horn warping, single-horn to double-horn warping or vice versa, and single-horn to single-horn warping). Thus the cost function for warping a single-horn DCP ($A$) to another single-horn DCP ($B$) is defined as

$$C(A \rightarrow B) = \sqrt{(x^A - x^B)^2 + (y^A - y^B)^2 + f^2 [2 \times \Delta(\delta^A_{\text{null}}, \delta^B_{\text{null}})]}. \quad (9)$$

### 3.2. Set-to-set correspondence

In that a DCP is modeled as a point in the 4-D feature space of location and directions, the representation of a generic face results to be a set of points in this space.

The dissimilarity between two faces can be characterized by the warping cost between the two sets in the 4-D space, as shown in Fig. 4.

Given two finite DCP sets $Q(A_1, A_2, \ldots, A_p)$ representing a query face and $M(B_1, B_2, \ldots, B_q)$ representing a model in the face database. $p$ and $q$ are the numbers of DCPs in $Q$ and $M$. A DCP set to set warping process is proposed to establish every DCP correspondence between the two DCP sets by minimizing the global warping cost. For each DCP $A_i$ in $Q$, its corresponding DCP $B_j$ in $M$ is identified as the one with minimum warping cost from $A_i$ to $B_j$ among all $B_j \in M$. The cost for establishing the paring for $A_i$ can be calculated by

$$\min_{B_j \in M} C(A_i \rightarrow B_j). \quad (10)$$

The cost for warping the whole set $Q$ to set $M$ (i.e., establishing paring for all DCPs in $Q$), denoted as $Q \Rightarrow M$, is defined as

$$C(Q \Rightarrow M) = \frac{1}{p} \sum_{A_i \in Q} \min_{B_j \in M} C(A_i \rightarrow B_j). \quad (11)$$

Finally, the dissimilarity between $Q$ and $M$ is defined as the maximum value of the two minimum costs to establish

![Fig. 4. An example pair of DCP sets (in red and green, respectively) from different face images of the same person.](image-url)
bilateral correspondences from $Q$ to $M$ and vice versa.

$$D(Q, M) = \max[C(Q \Rightarrow M), C(M \Rightarrow Q)].$$ (12)

For a given query face, the face retrieval process calculates above dissimilarity between the query face and each model in the database. The model in the database with minimum dissimilarity is considered as the correct return.

4. Experimental results

A complete system performance examination that covers all aspects of face retrieval was conducted. The following issues for face database retrieval are investigated.

1. Retrieval accuracy under controlled condition.
2. Sensitivity to scale variations.
3. Sensitivity to environmental changes. When collecting the image data in law enforcement, the subjects are usually asked to have a neutral expression and fronto-parallel face images are taken such that the effect of subject actions is minimized. However, the lighting conditions between the query image and the models are usually different because they are taken at different time and in different places.
4. Computational speed.

In order to make the investigation complete, the sensitivities to subject actions are also tested though their effect has been minimized in the data collection stage. The system was compared with the eigenface method [6], which is widely used as the benchmark approach.

In this study, two well-known and publicly available face databases were tested because one database [22] contains faces under controlled condition, lighting condition changes (environmental changes) and expression changes (subject actions) and the other [24] contains faces under controlled condition and rotated faces (subject actions). The AR face database [22] from Purdue University was used to evaluate the system performances under controlled condition, scale variations, lighting condition changes and facial expression changes. The database consists of over 3200 color images of the frontal view faces of 126 people (70 men and 56 women). There are 26 different images per person, recorded in two different sessions with a two-week time interval, each session consisting of 13 images. However, some images were found lost or corrupted after downloading through Internet. One hundred and twelve sets of images (61 men and 51 women) can be used. No restrictions on wear (clothes, glasses, etc.), make-up, hairstyle, etc. were imposed to the participants. For the details on the collection of the images in the AR face database, readers can refer to Martinez and Benavente [23]. Since the number of face images per person in the AR database is larger than in many other available face databases and only one image is used for training purpose, we are able to test the proposed approach under a large variety of conditions. The database from the University of Bern [24] was used to examine the system performances under controlled condition and head pose variations. The database contains frontal views of 30 people with different head pose variations (Two fronto-parallel pose, two looking to the right, two looking to the left, two looking downwards and two looking upwards). In all the experiments, a preprocessing to locate the faces was applied. Original images were normalized (in scale and orientation) such that the two eyes were aligned roughly at the same position with a distance of 80 pixels. Then the facial areas were cropped into the final images for query and retrieval. Some automatic face detection and eye location algorithms can be found in Refs. [25–29]. A few examples of the cropped faces are shown in Fig. 5.

4.1. Determination of $W$

The effect of $W$ in Eq. (5) was investigated using the AR database. All the neutral expression faces under background

Fig. 5. Examples of cropped faces.
Fig. 6. Example pairs of face images in the AR face database [22]. One is used as the model in the single model database; the other is used as the query image.

Fig. 7. The effect of \( W \) on retrieval rate.

lighting condition taken in the first session were used to construct the model database. The neutral expression faces under background lighting condition taken in the second session were used as query images. Sample query images and models in the database are illustrated in Fig. 6. The retrieval rate is plotted against the \( W \) in Fig. 7. \( W \) was found easily tuned since the retrieval rate remained higher than 93.75% when \( W \) ranged from 500 to 1200 (Fig. 7). The algorithm could only achieve a 2.68% retrieval rate when \( W = 1 \). It improved quickly with the increase of \( W \) and reached the optimal value of 94.64% when \( W = 800 \) and remained unchanged till 950. For all the other experiments in this study, \( W \) was set as 900. Readers can use this experimental parameter determination method to achieve optimal setting in other applications under different conditions.

### 4.2. Face retrieval under normal condition

The face images under controlled condition in the database from the University of Bern [24] were also used to evaluate the performance of the proposed approach. The retrieval rates are summarized in Table 1. The DCP experiments were conducted under three retrieval conditions, namely, top 1, top 5 and top 10 retrievals. In image retrieval, the performance of a system is not only reflected by the top 1 (or rank 1) retrieval rate but also by the top \( N \) retrieval rate, i.e., whether the correct image is among the best \( N \) retrieved images if it is not the best matched image. In the top 1 retrieval, a correct retrieval was counted when the best returned face in the model database was from the same person of the query face. In the top 5 or the top 10 retrieval, a correct retrieval was counted when the face image from the same person of the query face was among the best 5 or 10 returned faces in the model database, respectively. It is found that DCP approach performed better than (or equally well as) the eigenface method. The DCP and eigenface approaches both achieved 100% accuracy for retrieving faces in the Bern database. However, the DCP method significantly outperformed the eigenface method on the AR face database. Detailed eigenface data are tabulated in Table 2.

The performance of the eigenface approach depends on the number of eigenvectors. If this number is too small, important information about the identity is likely to be lost. If it is too high, the weights corresponding to small eigenvalues might be noises. The number of eigenvectors is limited by the rank of the training set matrix. One hundred and twelve is its upper bound in the experiment on the AR database and thus 78.57% is the best performance that the eigenface

<table>
<thead>
<tr>
<th>Method</th>
<th>Retrieval rate (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bern database</td>
<td></td>
</tr>
<tr>
<td>Eigenface</td>
<td>100</td>
</tr>
<tr>
<td>DCP</td>
<td>100</td>
</tr>
<tr>
<td>AR database</td>
<td></td>
</tr>
<tr>
<td>Eigenface</td>
<td>55.4</td>
</tr>
<tr>
<td>DCP (top 1 retrieval)</td>
<td>94.64</td>
</tr>
<tr>
<td>DCP (top 5 retrieval)</td>
<td>99.11</td>
</tr>
<tr>
<td>DCP (top 10 retrieval)</td>
<td>100</td>
</tr>
</tbody>
</table>

Table 1 Retrieval results of the proposed DCP and the eigenface (20-eigenvectors) methods.
approach can achieve here. One way to interpret this is that the eigenface approach works well as long as the test image is “similar” to the ensemble of images used in the calculation of eigenface [30]. And the training set should include a number of images for each person with some variations [6] to obtain a better performance. Here, only one model image per person was used for training. Another reason is the differences between two face images from the same person are larger in the AR database than that in the database from the University of Bern. In particular, the illuminations of the query images and the models are slightly different (see the last pair in Fig. 6) because the query images were taken after two weeks of taking the models.

4.3. Sensitivity to scale variations

The sensitivity to scale variations is an important issue that has not attracted much attention to date. Face detection and localization is a prior stage for face recognition. It is inevitable that a face is localized correctly with minor scale variation [31]. Recently, Martinez [32] highlighted and investigated the effect of imprecise localization problem on face recognition accuracy. It should be noticed that the localization problem does not mean a failure in face detection stage; rather, the face localization step succeeds, but that small-scale variation may cause mismatch in the identification process. Therefore, it is an important requirement that a face similarity measuring algorithm should be robust to scale variations.

A sensitivity analysis to scale variations was conducted using the AR database. The scale variations were generated by applying a random scaling factor, which was uniformly distributed within $[1-10\%,$ $1+10\%]$ to the query images. Four faces with different scales were generated from each query image. Thus, we had 448 query faces for the retrieval test with scale variations ranging from $-10\%$ to $+10\%$. The scales of the models were not changed.

The experimental results are tabulated in Table 3. The results show that the proposed DCP approach outperformed the eigenface approach by 23.4%, which means that it is much more robust to scale variations than the eigenface method. This is a very attractive property that can alleviate the difficulty of precisely locating faces in the prior face detection stage.

### Table 3

<table>
<thead>
<tr>
<th>Method</th>
<th>Top 1 (%)</th>
<th>Top 5 (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Eigenface (112-eigenvectors)</td>
<td>44.9</td>
<td>68.8</td>
</tr>
<tr>
<td>DCP method</td>
<td>68.30</td>
<td>73.22</td>
</tr>
</tbody>
</table>

4.4. Sensitivity to environmental changes

The robustness to environmental changes (i.e., lighting condition changes) is one of the critical issues for face image retrieval systems, such as photo album searching systems. When collecting the image data, subjects are usually required to be cooperative to have a neutral expression and fronto-parallel view faces are taken such that the effect of subject actions (i.e., head pose variations and facial expressions) is minimized. Usually, passport/IC photos are used to construct a model face database. Thus the environmental changes could be the only remaining appearance variability problem in face retrieval. In general, the query image and the model image of each person are taken at different time and in different places. Their lighting conditions are different and unknown. Moreover, the model images from different subjects are taken under different lighting conditions. It is impossible to get a query image under the same lighting condition as when all the model images in the photo database are taken. Hence, the retrieval algorithm has to be robust to the variability in appearance due to lighting condition changes.

The issue addressed in this section is whether the DCP representation is sufficient or how well it performs for retrieving faces under varying lighting conditions. The experiment was designed using face images taken under different lighting conditions from the AR database (Fig. 8). The faces in neutral expression with background illuminations were used as single models of the subjects. The images under three different lighting conditions were used as query images.

The experimental results on query images with three different lighting conditions are illustrated in Table 4 together with the retrieval results under the controlled condition as comparison benchmarks. In all the three experiments, the proposed DCP method significantly outperformed the eigenface approach. For the eigenface method, it has been suggested that the first three principal components are the primary components responding sensibly to lighting variations. The system error rate can thus be reduced by discarding these three most significant principal components [33]. Though the accuracies of the eigenface approach increased without using the first three eigenvectors, the DCP approach still significantly outperformed it when one light was on.

The variations of lighting condition did affect the system performance. Nevertheless, the DCP approach is much more tolerant to lighting condition changes than the
Table 4
Sensitivity to lighting condition changes

<table>
<thead>
<tr>
<th>Query faces</th>
<th>Eigenface</th>
<th>DCP method</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Top 1 retrieval (%)</td>
</tr>
<tr>
<td>Normal</td>
<td>20-eigenvectors</td>
<td>55.36%</td>
</tr>
<tr>
<td>condition</td>
<td>60-eigenvectors</td>
<td>71.43%</td>
</tr>
<tr>
<td>(Benchmark)</td>
<td>112-eigenvectors</td>
<td>78.57%</td>
</tr>
<tr>
<td></td>
<td>20-eigenvectors</td>
<td>6.25%</td>
</tr>
<tr>
<td></td>
<td>60-eigenvectors</td>
<td>9.82%</td>
</tr>
<tr>
<td>Left light</td>
<td>112-eigenvectors</td>
<td>9.82%</td>
</tr>
<tr>
<td>on</td>
<td>112-eigenvectors w/o</td>
<td>87.50</td>
</tr>
<tr>
<td>first 3 components</td>
<td>26.79%</td>
<td></td>
</tr>
<tr>
<td></td>
<td>20-eigenvectors</td>
<td>4.46%</td>
</tr>
<tr>
<td></td>
<td>60-eigenvectors</td>
<td>7.14%</td>
</tr>
<tr>
<td>Right light</td>
<td>112-eigenvectors</td>
<td>7.14%</td>
</tr>
<tr>
<td>on</td>
<td>112-eigenvectors w/o</td>
<td>89.29</td>
</tr>
<tr>
<td>first 3 components</td>
<td>49.11%</td>
<td></td>
</tr>
<tr>
<td></td>
<td>20-eigenvectors</td>
<td>1.79%</td>
</tr>
<tr>
<td></td>
<td>60-eigenvectors</td>
<td>2.68%</td>
</tr>
<tr>
<td>Both lights</td>
<td>112-eigenvectors</td>
<td>2.68%</td>
</tr>
<tr>
<td>on</td>
<td>112-eigenvectors w/o</td>
<td>61.61</td>
</tr>
<tr>
<td>first 3 components</td>
<td>64.29%</td>
<td></td>
</tr>
</tbody>
</table>

Fig. 8. Examples of model faces and query faces under different lighting condition changes.
eigenface method. The effect on retrieval rates when one light was on produced only 7.14% and 5.35% decreases in retrieval accuracy for the DCP approach. When both lights were on, the error rate became much higher than that of only one light on. This evidence shows that the DCP would still be affected by extreme lighting condition changes, such as over-illumination, though it is less sensitive to some extent. The over-illumination would cause strong specular reflection on the face skin (it is no longer a Lambertian surface). Therefore the shape information on faces would have been suppressed or lost, which would result in the increase of the error rate.

4.5. Computational efficiency

An experiment was conducted to evaluate the computational efficiency of the DCP matching using the AR face database [22]. The average computational time for one match is 57 ms. The experiment was conducted on a PC under Windows platform with 1.3 GHz CPU and 512 MB RAM. In a face retrieval system, searching is the most computationally expensive operation due to the large number of images available in the database. Therefore, it is a prerequisite of image retrieval systems to use efficient visual similarity matching algorithms. In most systems, face features are extracted/coded off-line from the original images, and stored in the face feature database. In querying process, the same features are extracted from the query face, and the features of the query image are compared with the features of each model image in the database. In practice, apart from adopting a fast face matching algorithm, pre-filtering operation [19] can be employed to further speed up the search by reducing the number of candidates.

4.6. Sensitivity to subject actions

Till now, the DCP approach has demonstrated very attractive capability on retrieval accuracy, robustness to scale and environmental changes, and computational efficiency. For a complete investigation covering all conditions of face similarity measuring, it is also interesting to know the remaining sensitivity problem due to subject actions (i.e., facial expressions and head pose variations), though they are of less importance in database retrieval systems as mentioned earlier. Similar experiments were conducted to evaluate the effects of different facial expressions (smile, anger and scream) using the AR database. The experimental results were summarized in Table 5. The smile expression caused recognition rate to drop by 31.25% as compared to neutral expression in Table 1 while the anger expression caused only 2.89% drop of the rate. This was not unexpected because the anger expression produced less physical variation from neutral expression than the expression of smile. The scream expression could be the extreme case of deformation among various human facial expressions, i.e., most facial features were distorted. The eigenface approach was found less sensitive to significant facial expression changes such as smile and screaming.

The face database [24] was also used to evaluate the system performance on query face images with different head poses. One fronto-parallel face per person was used as the model face. The system was tested using the 8 poses looking to the left, right, up and down for each person. There were 240 query images in total. The retrieval results are summarized in Table 6. It can be observed that head pose variations degraded the recognition rate of all the investigated methods, but DCP was more robust to head pose variations than the eigenface approach. Virtual view technique [34] and head pose recovery technique [35] can be employed to further improve the system performance of retrieving a face under pose variation.
5. Conclusion

In this paper, we present a novel visual similarity measuring technique to retrieve face images in photo album databases for law enforcement. The proposed DCP is particularly designed to address the problem of image querying and retrieval from single model face databases where accuracy, robustness to scale and environmental changes, and computational efficiency are three important issues to be considered. In order to meet the efficiency requirements of an image descriptor in high-dimensional image spaces, we extract DCPs from edge maps to further reduce the storage demand and the computational expense of edge map matching, while keeping the advantage of insensitivity to illumination changes. On the other hand, the directional attributes are introduced to enhance the discriminative capability to cater for high accuracy requirement.

The algorithm has been evaluated using two well known and public available face databases of over 3000 images and compared with the eigenface approach, one of the best face retrieval techniques. It is a very encouraging finding that the proposed DCP approach performed superior to the eigenface approach in most of the comparison experiments. DCP correctly returned 100% and 94.64% of the queries on the face databases [24] and [22] respectively. The DCP approach significantly outperformed the eigenface method by 23.4% in retrieval rate by querying a face with scale variation. The effect on retrieval rates when one light was on degraded only by 7.14% and 5.35%. This study demonstrates that the proposed DCP method provides a new solution, which is both robust to scale and environmental changes, and efficient in computation, for retrieving human faces in single model databases.

Acknowledgements

This work is partially supported by the Australian Research Council (ARC) Discovery Grant DP0451091.

References


About the Author—YONGSHENG GAO received the B.Sc. and M.Sc. degrees in Electronic Engineering from Zhejiang University, China, in 1985 and 1988, respectively, and the Ph.D. Degree in computer engineering from Nanyang Technological University, Singapore. Currently, he is a Senior Lecturer with the School of Microelectronic Engineering, Faculty of Engineering and Information Technology, Griffith University, Australia. His research interests include face recognition, biometrics, image retrieval, computer vision, and pattern recognition. He is a member of the IEEE.