Shadow compensation in 2D images for face recognition

Sang-Il Choi*, Chunghoon Kim, Chong-Ho Choi

School of Electrical Engineering and Computer Science, Seoul National University, #047, San 56-1, Sillim-dong, Gwanak-gu, Seoul 151-744, Korea

Received 4 July 2006; received in revised form 3 November 2006; accepted 24 November 2006

Abstract

Illumination variation that occurs on face images degrades the performance of face recognition. In this paper, we propose a novel approach to handling illumination variation for face recognition. Since most human faces are similar in shape, we can find the shadow characteristics, which the illumination variation makes on the faces depending on the direction of light. By using these characteristics, we can compensate for the illumination variation on face images. The proposed method is simple and requires much less computational effort than the other methods based on 3D models, and at the same time, provides a comparable recognition rate.

Keywords: Face recognition; Illumination variation; Shadow compensation; Linear discriminant analysis

1. Introduction

Recently, there has been extensive research on face recognition following its increased application in various fields. As a result, numerous algorithms based on Eigenface [1], Fisherface [2] and independent component analysis (ICA) [3], have been developed, which are known to perform relatively well under ideal circumstances. However, there still remain many problems that must be overcome to develop a robust face recognition system that works well under various circumstances such as illumination variation.

In order to overcome the problems due to illumination variation, several algorithms have been introduced. Shashua [4] presented a 3D linear subspace approach and Batur and Hayes [5] proposed a segmented 3D linear subspace approach. Georghiades et al. [6] presented a modeling of an illumination cone and Lee and Kriegman [7] proposed the 9D linear subspace approach by using nine images captured under nine lighting directions. Basri and Jacobs [8] represented lighting using spherical harmonics and described the effects of Lambertian reflectance as an analogy to convolution. Zhang and Samaras [9] presented a statistical method in dealing with illumination variation. However, all these methods use either a 3D face model or a special physical configuration, which requires much computational effort [10]. Although Shen et al. [11] proposed an algorithm which restores the shaded image based on a 2D image, it requires several image processing techniques and iteration procedures. Xie and Lam [12] proposed a method to eliminate the influence due to illumination variation using a 2D shape model, which separates the input image into a texture model and a shape model for retaining the shape information. Recently, they tried to alleviate the effect of uneven illumination using a local normalization technique [13,14]. Song et al. [15] solved the illumination variation based on a 2D image using a mirror image under the assumption of facial symmetry, and so it uses only half of the information on a face image.

In this paper, we propose a new approach for handling illumination variation. Generally, human faces are similar in shape in that they are comprised of two eyes, a nose and a mouth. Each of these components makes a shadow on a face, showing distinctive characteristics depending on the direction of light. By using such characteristics generated by the shadow, we can compensate for illumination variation on a face image caused by the shadow and obtain a compensated image that is similar to the image taken under frontal illumination. This image, which will be used for face recognition, will be referred to as the compensated image.
On the other hand, there are several approaches that use local images such as the eyes, nose, and mouth in face recognition. Brunelli and Poggio [16] used the geometric local features, and Pentland et al. [17] presented a modular eigenspace technique which incorporates the local images. Similarly, by applying the histogram equalization to each of the local images, we can extract useful features which are robust under various illumination conditions. These local features are used together with the global features obtained from the compensated image to improve the performance of face recognition as in Ref. [18]. Through experiments, we show that the proposed method works quite well and also improves the confidence of the face recognition system.

The proposed method has several advantages in comparison to the other algorithms in overcoming illumination variation. This method is based on 2D images whereas most of the other methods are based on 3D models. Therefore, it requires much less computational effort. Moreover, the direction of light can be easily determined by using the binary image transformed from a face image. Also, it requires only one average shadow image for all individuals to obtain the compensated images, depending on the direction of light, which makes the compensation procedure very simple. The compensated image can be used without making any modification to other face recognition algorithms based on 2D images.

The rest of this paper is organized as follows. Section 2 explains how to compensate the shadow in a face image and obtain the additional local images. Section 3 consist of the experimental results, followed by conclusion in Section 4.

2. Proposed algorithm

2.1. Determining the direction of light

To obtain the information about shadow characteristics, we need to know the direction of light. We denote the gray-level intensities of a face image (see Fig. 1(a)) of \( I(\text{height}) \times J(\text{width}) \) pixels as \( I_c \in \mathbb{R}^{I \times J} \). (We used \( I = 120 \) and \( J = 100 \) in the experiments.) The subscript \( c \) denotes the category where the direction of light belongs. Now we study how to determine the category \( c \) when \( c \) is unknown. In order to reduce the influence of the background on a face image, we take a square mask of 80 \( \times \) 80 (pixels) that only covers part of a face image of 120 \( \times \) 100 (see Fig. 1(b)). Let \( a_c \) denote the average value of the gray-level intensity of all the pixels in a face image of \( I_c \), i.e.,

\[
a_c = \frac{1}{IJ} \sum_{i=1}^{I} \sum_{j=1}^{J} I_c(i, j).
\]

We modify the values of \( I_c(i, j) \) in the gray scale of [0, 255] as follows:

\[
I'_c(i, j) = \begin{cases} 
0 & \text{if } I_c(i, j) < a_c, \\
255 & \text{if } I_c(i, j) \geq a_c.
\end{cases}
\]

The value for \( a_c \) varies depending on the face image because the skin color and shadow level in each image are all different. Fig. 1(b) shows some examples of \( I'_c(i, j) \) obtained from Eq. (1). Since the nose is the most prominent feature when forming the shadow on a face, the brightness of an image can be divided approximately into the left and right sides with respect to the nose. Let \( g_R \) and \( g_L \) be the average gray-level intensities of the right half (\( R \)) or left half (\( L \)) of an image, respectively, i.e.,

\[
g_R = \frac{2}{IJ} \sum_{i=1}^{I} \sum_{j=(J/2)+1}^{J} I'_c(i, j),
\]

\[
g_L = \frac{2}{IJ} \sum_{i=1}^{I} \sum_{j=1}^{J/2} I'_c(i, j).
\]

Then, we determine the category \( c \) with some constant \( t_k \), as shown below:

\[
c = \begin{cases} 
R_k & \text{if } t_k \leq g_R - g_L < t_{k+1}, \quad k = 1, 2, 3, \\
F & \text{if } |g_R - g_L| < t_1, \\
L_k & \text{if } -t_{k+1} \leq g_R - g_L < -t_k, \quad k = 1, 2, 3.
\end{cases}
\]

We set \( t_1, t_2, t_3 \) and \( t_4 \) as 25, 50, 75, 255, respectively, in the gray scale of [0,255]. These values were determined after some

![Fig. 1. (a) Images under various illuminations; (b) corresponding images obtained from Eq. (1).](image-url)
work with the Yale B database [6]. The illumination variation in a face image in the Yale B database is caused by the variation of light source direction both in azimuth and elevation, but we take only the azimuth into account in compensating the illumination variation. Fig. 2 shows the distribution of the angle between the horizontal direction of light and the frontal direction in each category. The vertical axis represents the angle of light source direction with respect to the frontal direction and the horizontal axis represents the category. The positive value implies that the light source was to the right of the subject while negative value means that it was to the left. In the figure, each vertical bar denotes a standard deviation of angle in both directions. As can be seen in Fig. 2, the mean of the angles that belong to $R_k$ or $L_k$ increase linearly as $k$ increases, which implies that the direction of light can be determined pretty well by using the binary image. After determining the direction of light, we compensate illumination variation following the procedure explained in the next subsection.

2.2. Shadow compensation

Fig. 3(a) and (b) show the face images for two individuals from the Yale B database; one image is taken under the frontal illumination and the other is taken under the left side illumination. We denote a face image under the frontal illumination ($F$) as $I_{m,F}$, a face image under the left side illumination ($L_k$) as $I_{m,R_k,n}$, and a face image under the right side illumination ($R$) as $I_{m,L_k,n}$, where $m (=1, 2, \ldots, M)$ and $n (=1, 2, \ldots, N_c)$ in $I_{m,R_k,n}$ and $I_{m,L_k,n}$ denote the $n$th image of the $m$th person in an image database when the direction of light belongs to category $c$. The gray-level intensities $I_{m,R_k,n}(i,j)$ and $I_{m,L_k,n}(i,j)$ at pixel $(i,j)$, which have values ranging from 0 to 255, vary depending on the category, and are different from that of $I_{m,F}(i,j)$. We define the intensity difference between the images of $I_{m,F}$ and $I_{m,c,n}$ at each pixel $(i,j)$ as follows:

$$D_{m,c,n}(i, j) = I_{m,c,n}(i, j) - I_{m,c,n}(i, j),$$

$$i = 1, 2, \ldots, I; \quad j = 1, 2, \ldots, J.$$  \hspace{1cm} (2)

Fig. 3(c) shows $D_{m,L_k,n}$ normalized to have values ranging between 0 and 255, because some gray-level intensity values in the intensity difference may be negative. Since most human faces are similar in shape, we can assume that the shadows on facial images are also similar in shape when the direction of light belongs to the same category. However, the intensity difference $D_{m,c,n}$ of one person is insufficient to compensate for the intensity differences of another person’s images under various illumination conditions because $D_{m,c,n}$ contains not only the information on the category, but also the unique feature of the $n$th person. This is demonstrated by a simple example shown in Fig. 3(d). The intensity difference for each individual is obtained from $I_{m,F}$ and $I_{m,L_k,n}$, $m = 1, 2$ by using Eq. (2) in Fig. 3(a) and (b). We then synthesize images in Fig. 3(d) by adding $I_{1,L_k,n}$ of the first person and $D_{1,L_k,n}$ of the second person, and vice versa. As can be seen in Fig. 3(d), these images are quite different from their corresponding images under the frontal illumination in Fig. 3(a). Therefore, in order to compensate for the intensity difference due to illumination variation, we need to eliminate the influence of features that are due to individuals. For this, we define the average intensity difference for a fixed value of $c$ as follows:

$$D_{A,c}(i, j) = \frac{1}{MN_c} \sum_{m=1}^{M} \sum_{n=1}^{N_c} (I_{m,F}(i, j) - I_{m,c,n}(i, j)).$$

Note that there are no subscripts $m$ or $n$ in $D_{A,c}$. Since this average intensity difference represents the general characteristic of the shadow in a face image for the direction of light belonging to category $c$, it can be applied to any face image in order to compensate the shadow formed by the light belonging to the category $c$. The average intensity difference, which is shown in Fig. 4, was made from the images of 65 individuals in the
CMU-PIE illumination database [19]. $D_{A,c}$’s shown in Fig. 4 are represented in the same manner as in Fig. 3(c) because some of the values of $D_{A,c}$ are negative. With the average intensity difference, we can obtain the compensated image, $I_{m,c,n}^C$ of $I_{m,c,n}$ as

$$I_{m,c,n}^C(i, j) = I_{m,c,n}(i, j) + D_{A,c}(i, j).$$

(3)

Note that the compensated image for face recognition can be obtained with only one average intensity difference $D_{A,c}$ for each $c$. Fig. 5 shows the images in which shadows are compensated using Eq. (3).

2.3. Global features and local features

Along with the compensated image of a full face, local images such as the eyes, nose and mouth can provide additional features for face recognition [16]. Under illumination variation, these local features can be less sensitive in comparison to the global features. In many cases, the local images that correspond to the eyes, nose and mouth lie within the shadow region of a face. Thus, if the local images in the shadow region are separated from the global image and processed with the histogram equalization [20], the local images can be restored closely to the corresponding images that are not in the shadow region. Fig. 6(b) shows the images that were segmented from the image in Fig. 6(a), and their histogram equalized images are shown in Fig. 6(c). In Fig. 6(c), it is apparent that the dark parts on the left side of the image in Fig. 6(a) became much brighter. Next, each restored local image was put together to make local images of two eyes, a nose and a mouth as shown in Fig. 6(d). From these compensated global and local images, the global and local features (projection vectors) were obtained by the LDA-based subspace method [2,21], and the combined subspace were constructed with the projection vectors corresponding to large eigenvalues selected among the eigenvalues of each subspace [18]. Since we are primarily interested in shadow compensation, the details on the face recognition procedures will not be discussed in this paper. (For more details, see Refs. [2,18].)
3. Experimental results

We applied the proposed method on the Yale B database and the CMU-PIE illumination database according to the face recognition procedure in Ref. [18]. The center of each eye was manually located and the eyes were rotated to be aligned horizontally as in Ref. [22]. Each face image was cropped and rescaled so that the center of each eye is placed at its fixed point in an image of 120 × 100 (pixels). Each region corresponding to the eyes, nose, and mouth was cropped from a predetermined area in the rescaled image. The resolutions of the global image and the local images of an eye, one half of the nose, and one half of the mouth were 120 × 100, 30 × 40, 70 × 15, and 30 × 30 (pixels), respectively, as shown in Fig. 6. After the histogram equalization of these images, we compensated these images as described in Section 2. The features were extracted from each of the global image and local images, and these features with the $L_2$ metric were used for face recognition following the procedure in Ref. [18].

3.1. Yale B database

The Yale B database contains images of 10 individuals in nine poses and 64 illuminations per pose. We used 45 face images for each subject in the frontal pose (YaleB/Pose00) which were further subdivided into four subsets (subset $i$, $i = 1, 2, 3, 4$) depending on the direction of light as in Ref. [21]. The index of the subset increases as the light source moves away from the front in taking the pictures. The images in subsets 1 and 2 were selected as a training set and the other images in subsets 3 and 4 were used as the test set. Table 1 shows the recognition rates, which are based on the raw images ($I^{raw}$), the images after the histogram equalization ($I^{hist}$), the shadow compensated images ($I^C$), and the local images in addition to the compensated images ($I^{CL}$). The histogram equalization alone gives a recognition rate of 94.0% while the shadow compensated images show a 3.1% increase. When the local features were added, we could observe an additional 2.6% increase for all the subsets. Note that the performance of the proposed method is most significant for subset 4, where images are severely affected by shadow. Fig. 7(a) shows the recognition rate for a different number of features. The proposed method gives a recognition rate of 99.6% with 36 features. As can be seen from Table 2, the proposed method is better than all the other methods including the methods based on the 3D models, except the cone-cast [6] and the gradient angle [24]. Although the method in Ref. [6] gives a recognition rate of 100%, it requires much more computational effort due to a large number of extreme rays that make up the illumination cones. For example, there are $O(n^2)$ extreme rays, where $n$ is the number of pixels for a convex Lambertian surface. Lee et al. [7] achieved a recognition rate of 99.1%, but some of the 3D information, such as albedos and
for the proposed method, while it is 95% for the local normalization method in Ref. [14], which applies a local normalization technique on each pixel of an image, gives a recognition rate of 98.9%.

In addition, we computed the relative distance \( d_{rel} = d_2/d_1 \), where \( d_1 \) and \( d_2 \) are the distances of the first and second nearest neighbors of a probe image, respectively. \( d_{rel} \) shows the robustness of the face recognition, and \( \log_{10}(d_2/d_1) \) is called the confidence measure [25]. Fig. 8(a) shows the probability distributions of the correct and incorrect recognition results depending on the relative distance when using \( I_{CL} \) images. From this figure, we can see that \( d_{rel} \) is distributed between 1 and 1.07 in the case of correct recognition results, while it is distributed mostly above 1.07 in the case of correct recognition results. The mean of \( d_{rel} \) for the probe images, which are correctly classified, was found to be 4.74 for \( I_{hist} \) images, and 5.95 for \( I_{C} \) images. A higher mean value of \( d_{rel} \) indicates that the recognition result is more reliable. It means that the compensation procedure improves the reliability of the decision. We can improve the reliability of a decision made on face recognition by accepting the result when \( d_{rel} \) is greater than a sufficiently large value \( T \) and rejecting the results otherwise. Fig. 9(a) shows the correct recognition rate versus the rejection rate for various stages of compensation. As illustrated in the figure, the recognition rates improve as the rejection rate increases. The recognition rate is 100% with a 2.3% (\( T = 1.08 \)) rejection rate for the proposed method, while it is 95.3% for the histogram equalization (\( I_{hist} \)). This means that the recognition system is more reliable when the proposed method is used.

### 3.2. CMU-PIE illumination database

The CMU-PIE illumination database contains images of 65 individuals with 21 different illumination variations. Each of the three images for each individual, which are under the left side, right side and frontal illumination, were used for training, while the others were used for testing. The recognition rate for the histogram equalization was 92.9%. As can be seen in Fig. 7(b), the compensated images (\( I_{C} \)) give a recognition rate of 97.2%, which is approximately 30% and 4.3% better than \( I_{raw} \) images and \( I_{hist} \) images, respectively. The best recognition rate of 99.1% was obtained with 150 features when the local images were used in addition to the compensated image. Table 3 shows that the proposed method has better recognition rate than all the other methods. From Fig. 9(b), we can also see that the compensated images \( I_{C} \) and \( I_{CL} \) give much more reliable results than \( I_{hist} \). The mean of relative distance for correct recognition results is 1.67 for \( I_{hist} \) images and 1.88 for \( I_{C} \) images. Fig. 8(b) shows that if \( T \) is set to 1.12, the recognition rate for \( I_{CL} \) images is 100% at the rejection rate of 5.5% (\( T = 1.12 \)), whereas it is 95.6% for \( I_{hist} \) images according to Fig. 9(b).

### Table 2

<table>
<thead>
<tr>
<th>Method</th>
<th>Direction of light</th>
<th>Subsets 1 and 2</th>
<th>Subset 3</th>
<th>Subset 4</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Correlation [16]</td>
<td></td>
<td>100</td>
<td>76.7</td>
<td>26.4</td>
<td>70.9</td>
</tr>
<tr>
<td>Eigenfaces [1]</td>
<td></td>
<td>100</td>
<td>74.2</td>
<td>24.3</td>
<td>69.6</td>
</tr>
<tr>
<td>Eigenfaces w/o first 3 [2]</td>
<td></td>
<td>100</td>
<td>80.8</td>
<td>33.6</td>
<td>74.2</td>
</tr>
<tr>
<td>Fisherface [2]</td>
<td></td>
<td>100</td>
<td>95.8</td>
<td>53.6</td>
<td>84.1</td>
</tr>
<tr>
<td>Linear subspace [23]</td>
<td></td>
<td>100</td>
<td>100</td>
<td>85.0</td>
<td>95.4</td>
</tr>
<tr>
<td>Cone-attached [6]</td>
<td></td>
<td>100</td>
<td>100</td>
<td>91.4</td>
<td>97.3</td>
</tr>
<tr>
<td>Harmonic image [9]</td>
<td></td>
<td>100</td>
<td>99.7</td>
<td>96.9</td>
<td>98.9</td>
</tr>
<tr>
<td>Gradient angle [24]</td>
<td></td>
<td>100</td>
<td>100</td>
<td>98.6</td>
<td>99.6</td>
</tr>
<tr>
<td>Cone-cast [6]</td>
<td></td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>Shen’s image [11]</td>
<td></td>
<td>100</td>
<td>98.3</td>
<td>96.4</td>
<td>98.4</td>
</tr>
<tr>
<td>Local normalization [14]</td>
<td></td>
<td>100</td>
<td>100</td>
<td>96.4</td>
<td>98.9</td>
</tr>
<tr>
<td>The proposed method</td>
<td></td>
<td>100</td>
<td>100</td>
<td>98.6</td>
<td>99.6</td>
</tr>
</tbody>
</table>

*Fig. 8. Probability distribution of the relative distance for the cases of correct and incorrect recognition results for \( I_{CL} \) images: (a) the Yale B database; (b) the CMU-PIE illumination database.*
equalization on local images of the eyes, nose, and mouth, we can obtain additional features for robust face recognition. By using both the compensated images and the local images, the recognition rate exceeds 99% for both the Yale B database and the CMU-PIE illumination database, exceeding the performance of other methods in most cases. Moreover, the compensated image makes the face recognition system more reliable. The proposed method has several advantages. The category to which direction of light belongs can be easily found regardless of skin color and the illumination condition because the binary image is constructed based on the average value of the gray-level intensity of pixels in an image. The shadow compensation is simple to use because it requires only one average intensity difference for the shadow compensation depending on the category to which direction of light belongs. Since the proposed method is based on 2D images, it is computationally much simpler than the other methods based on 3D models. We expect that the proposed method can also be applied to compensate for illumination variation in images of other objects.

Acknowledgment

This work was supported by the Korea Research Foundation Grant (KRF-2005-041-D00491) funded by the Korean Government (MOEHRD).

References


4. Conclusions

This paper proposes a novel approach to reducing the degradation of face recognition rate by illumination variation. Since human faces are similar in shape in general, we can compensate the shadow variation in faces relatively easily by adding the average intensity difference regardless of the individual, depending on the direction of light. By applying the histogram

Fig. 9. Recognition rate versus rejection rate: (a) the Yale B database; (b) the CMU-PIE illumination database.

Table 3

<table>
<thead>
<tr>
<th>Method</th>
<th>Recognition rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Correlation [16]</td>
<td>85.4</td>
</tr>
<tr>
<td>Eigenfaces [1]</td>
<td>79.7</td>
</tr>
<tr>
<td>Eigenfaces w/o first 3 [2]</td>
<td>89.5</td>
</tr>
<tr>
<td>Fisherface [2]</td>
<td>92.9</td>
</tr>
<tr>
<td>Local Normalization [14]</td>
<td>98.9</td>
</tr>
<tr>
<td>The proposed method</td>
<td>99.1</td>
</tr>
</tbody>
</table>

The Yale B database

Rejection rate (%)  0  5  10  15  20
Recognition rate (%) 75  80  85  90  95

The CMU-PIE database

Rejection rate (%)  0  5  10  15  20
Recognition rate (%) 70  75  80  85  90

Table 3

Recognition rates for different methods on the CMU-PIE database (%)


About the Author—S.-I. Choi received a B.S. degree in the Division of Electronic Engineering from Sogang University, Korea in February 2005. He is currently pursuing a M.D. degree in the School of Electrical Engineering and Computer Science from Seoul National University. His research interests include face recognition, feature extraction, neural networks, and their application.

About the Author—C. Kim received a B.S. degree in the School of Electrical Engineering from Seoul National University, Korea in 1998. He was a research engineer at venture companies in Korea from 1998 to 2002. He is currently pursuing a Ph.D. degree in the School of Electrical Engineering and Computer Science from Seoul National University. His research interests include face recognition, pattern classification, neural networks, and their applications.

About the Author—C.-H. Choi received a B.S. degree from Seoul National University, Korea in 1970 and M.S. and Ph.D. degrees from the University of Florida, Gainesville, in 1975 and 1978, respectively. He was a senior researcher with the Korea Institute of Technology from 1978 to 1980. He is currently a professor in the School of Electrical Engineering and Computer Science, Seoul National University. He is also affiliated with the Automation and Systems Research Institute, Seoul National University. His research interests include control theory and network control, neural networks, system identification, and their applications.