Abstract We propose a robust face recognition method under illumination variation. By using shadow compensation methods, we restore the image of a facial image taken under arbitrary illumination into an image that is similar to the image taken with frontal illumination. Then we apply a pixel selection method to these restored images in order to reduce the noise components, which can interfere with the extraction of discriminant features for face recognition. The experimental results for the CMU-PIE, Yale B and Multi-PIE databases show that the proposed method results in the improvement of recognition performance under illumination variation.

Keywords Shadow compensation; Fourier analysis; feature extraction; discriminant analysis; pixel selection; noise reduction; face recognition

1. Introduction

Face recognition is used to identify individuals from facial images by using face databases containing people’s identities. Compared to other types of biometric recognition, such as fingerprint, retinal or iris recognition, face recognition is less invasive and does not require a subject to be in proximity to or in contact with a sensor, which makes it widely applicable in user identification, e-commerce, access control, surveillance and human-computer interaction.

Numerous methods have been developed for face recognition over the last few decades. Most of them are capable of achieving high recognition rates in well-controlled environments. However, the accuracy of face recognition can deteriorate significantly when the images used for enrolment (gallery) and recognition (probe) are taken under different environmental conditions.

Illumination variation is one of the challenging problems to be solved for robust face recognition systems. The changes induced by illumination, such as cast shadows or attached shadows, can be larger than the innate differences between individuals. The methods used to solve the problems caused by illumination variation can be roughly classified into three categories [1], [2], [3]: face
modelling, invariant feature extraction and pre-processing and normalization. Face model-based methods construct a generative 3D face model that can be used to render facial images under various illumination conditions. Most of these methods require several images of the subject using light from different directions or 3D shape information for training [4], [5]. The methods belonging to the second category extract features, such as edge maps [6], gradient faces [7], Gabor features [8] or local binary patterns (LBP) [9], which are invariant to illumination variation. In pre-processing and normalization, various image processing techniques are used to compensate for the illumination variation [10], [12], [11], [13]. These methods [10], [12], [13] modify a facial image taken under arbitrary illumination conditions into an image that is similar to one taken under frontal illumination. The method in [13] restored a frontal illumination image from an image captured under arbitrary lighting conditions by using the ratio-image between a test image and a reference image, and iterated the process to obtain a visually better restored image. The quotient image (QI) [14], which is defined as the ratio between a test image and a linear combination of three noncoplanar illumination images, depends only on the albedo information and is free of illumination. The method in [11] used the ratio between a test image and its smoothed image to obtain a self-quotient image (SQI).

From human perception, facial images with frontal illumination have a good visual effect and are natural for the investigation of discriminant information for face recognition. In addition, the compensated images can be applied to well-known appearance-based face recognition methods such as Fisherface [15], DCV [16] or ERE [17]. In this paper, we focus on the methods belonging to the third category for handling illumination variation in face recognition.

We first present two state of the art methods [12], [19] for shadow compensation, which are simple and give good recognition performance compared to other methods. The method in [12] decomposes a facial image into magnitude components and phase components in the frequency domain using the Fourier transform. The facial image that was deteriorated by shadow was restored using the auxiliary magnitude, which compensates for the magnitude components distorted due to the illumination variation. In [19], the illumination variation was compensated for by adding an adequate average difference image depending on the light direction, under the assumption that the shadow on a facial image created by a similar light direction has common characteristics. However, the resultant compensated images of both the above methods still have many noisy pixels, which are likely to interfere in extracting discriminant features for face recognition. Thus, we can expect to obtain better discriminant features by eliminating noisy pixels in the compensated images, which will lead to improved face recognition performance. Even though various pixel selection methods can be used to remove noisy pixels [20], [21], [18], we modify the pixel selection method in [20] based on discriminant analysis and apply it to remove those noisy pixels. Experimental results show that the proposed method improves recognition performance in the presence of illumination variations.

The rest of this paper is organized as follows. Section 2 and 3 present how to compensate for the shadow in a facial image and how to remove noisy pixels in the compensated images, respectively. Section IV consists of the experimental results, followed by the conclusion in Section V.

2. Shadow Compensation Methods

2.1 Shadow Compensation Based on Fourier Analysis (SCFA)

In [12], a shadow compensation method was performed using Fourier analysis to deal with illumination variation. The image signal in the spatial domain can be represented in the frequency domain using the Fourier transform [21]. Let us denote an image of $M \times N$ pixels as $I(x, y) \in \mathbb{R}^{M \times N}$. The Fourier transform of $I(x, y)$ is generally complex valued and thus it can be represented in terms of magnitude and phase as $F_I(u, v) = |F_I(u, v)| e^{i \theta_I(u, v)}$. The magnitude $|F_I(u, v)|$ describes the basic frequency content of a signal providing information about the relative magnitudes of the complex exponentials that make up $I(x, y)$. The phase angle $\theta_I(u, v)$, on the other hand, provides information concerning the relative phases of these exponentials. In a facial image, some of the most discriminant information is contained in the edges and regions of high contrast. Thus, the phase components, which have the structural information of the image, are very important in face recognition [12].

The shadows in a facial image due to illumination variations change the magnitude components of the image in the frequency domain. In contrast, the phase components are less prone to the effects of illumination variations [12]. In order to restore a facial image that had been deteriorated by shadows, the method in [12] compensated for the distorted magnitude components using the auxiliary magnitude $|F_{\text{aux}}(u, v)|$, which represents the characteristic of the magnitude for general human faces. If letting an arbitrary facial image with shadows and its Fourier representation be $I_A$ and $|F_{I_A}(u, v)| e^{i \theta_{I_A}(u, v)}$, respectively, the restored image $I_{\text{gen}}$ can be obtained from the compensated magnitude $|F_{C}(u, v)|$ and its own phase $\theta_{I_{\text{gen}}}(u, v)$ as follows:
\[ I_{R_u}(x, y) = \frac{1}{MN} \sum_{u=0}^{M-1} \sum_{v=0}^{N-1} F_c(u, v) e^{j2\pi (ux/M + vy/N)} \]  

(1)

where \[ |F_c(u, v)| = 1/2(|F_{l_a}(u, v)| + |F_{l_{ax}}(u, v)|) \] , \[ \Theta_c(u, v) = \Theta_{l_a}(u, v), x = 0, \ldots, M - 1 \text{ and } y = 0, \ldots, N - 1. \]

Figures 1(a) and 1(b) show the raw images and their restored images using (1), respectively.

2.2 Shadow Compensation Using Weighted Average Difference (SCWAD)

![Figure 1](image)  

(a)  
(b)  
(c)  

Figure 1. Examples of the images in which shadows are compensated. (a) Raw images. (b) Shadow compensated images using (1). (c) Shadow compensated images using (2).

Generally, human faces are similar in shape in that they are comprised of two eyes and two eyebrows, a nose and a mouth [10]. Each of these components makes a shadow on the face, thereby showing distinctive characteristics depending on the direction of light. If a facial image taken under frontal illumination is set as a reference image for a given facial image that is taken under an arbitrary illumination, the information about the shadow characteristics of the given facial image is contained in the difference image between the facial image and the reference image. Thus, if the light direction of a given facial image can be estimated, we can compensate for the shadow in the facial image using the average difference images \( D_{l_a} \), which is an average of the individual difference images for the same light direction.

First, \( L \) light direction categories \( \{ C_l \} \) \( l = 1, \ldots, L \) are defined from the left side to the right side. For a given facial image \( I_{l_a} \) , the binary image \( B(I_{l_a}) \) of \( I_{l_a} \) is produced in order to estimate the light direction. Then, the distances between \( B(I_{l_a}) \) and the binary images in each category \( C_l \) are calculated using a classification rule [19]. From the three nearest distances \( \text{dist}_{iN} \) , \( i = 1,2,3 \) and their corresponding categories \( C_l \), \( i = 1,2,3 \), the weight \( \omega_i \) is defined as

\[ \omega_i = \frac{\text{dist}_{iN}}{\sum_{i=1}^{3} \text{dist}_{iN}}. \]

The value \( \omega_i \) indicates how well the shadow of the given facial image can be characterized by \( D_{l_a} \). Finally, the shadow compensated image, \( I^C \) of \( I_{l_a} \), is obtained with the average differences as follows.

\[ I^C(x, y) = I_{l_a}(x, y) + \sum_{i=1}^{3} \omega_i D_{l_a}^{(i)}(x, y) \]  

(2)

Figure 1(c) shows the images compensated using (2).

3. Pixel Selection Based on Discriminant Analysis

It is shown in the compensated images of Figs. 1(b) and 1(c) that noise components, which were generated in the process of compensating magnitude components or adding average difference images, are distributed over a whole facial image. This is because the average operations in the compensation procedures dilute the unique features of an individual in a facial image. In addition, the remaining cast shadows after shadow compensation can affect the extraction of discriminant features for face recognition. Therefore, in order to improve the performance of face recognition using these compensated images, we modify the pixel selection method based on discriminant analysis [20] and apply it to distinguish informative pixels from all pixels in the image. By eliminating the noisy pixels that have less discriminative information, we can expect to extract better discriminant features for face recognition, which will result in improved recognition performance.

The projection matrix \( W \) to be used for pixel selection is obtained using DCV [16]. The \( k \)-th image sample \( I_k \) can be represented as a vector \( x_k = O_k(I_k) \), where \( O_k(\cdot) \) is the lexicographic ordering operator that transforms a matrix into a vector [23]. When the training set contains \( N \) samples and \( c \) classes, each of which has \( N_l \) samples, between \( S_B \)- and within \( S_W \)-class scatter matrices are defined as

\[ S_B = \frac{1}{N} \sum_{i=1}^{c} N_l (\mu^l - \mu)(\mu^l - \mu)^T \]

\[ S_W = \frac{1}{N} \sum_{i=1}^{c} \sum_{x_k \in C_l} (x_k - \mu^l)(x_k - \mu^l)^T \]
where \( \mu \) is the total mean of the whole training set, \( \mu^i \) is the mean of the class \( c_i \), and \( x_k \) is a sample belonging to class \( c_i \). Then, the objective function of DCV is defined as follows:

\[
W = \arg \max_W \frac{|W^TS^BW|}{|W^TS^BW|}
\]

The projection matrix \( W \) consists of projection vectors \( w_i(= [\alpha_{i1}, \ldots, \alpha_{in}]^T) \) s \( (W = [w_1, w_2, \ldots, w_m]) \). The discriminate feature vector \( y_k = [y_{k1}, \ldots, y_{km}]^T \) is obtained from \( x_k \) as

\[
y_k = W^Tx_k \in \mathbb{R}^m, \quad k = 1, 2, \ldots, N. \tag{3}
\]

Each feature \( y_{ik}(k) \) in (3) is computed by projecting \( x_k \) onto \( w_i(y_{ik} = w_i^Tx_k) \). Since each feature is expressed by a linear combination of individual pixels \( x_{kl} \), \( l = 1, \ldots, n \), of which coefficients are \( \alpha_{il} \), the value \( |\alpha_{il}| \) indicates how much the \( l \)-th pixel \( x_{kl} \) contributes to constructing the \( i \)-th feature \( y_{ik} \). For example, if \( |\alpha_{il}| \) is larger than \( |\alpha_{ij}| \), \( x_{kl} \) can be regarded as a more important pixel for extracting a discriminant feature than \( x_{kj} \). Therefore, \( |\alpha_{il}| \) of each \( w_i \) can be used as a measurement of discriminative information in the \( i \)-th pixel \( x_{kl} \). Without defining an order vector separately [20], using \( |\alpha_{il}| \) as a measurement makes the pixel selection process more simple, which might offer advantages in terms of computation when dealing with high-dimensional data such as face images. From \( W \), which consists of \( m \) projection vectors \( w_i \), \( i = 1, \ldots, m \), different measurements can be produced. Since a projection vector with a larger eigenvalue has more discriminative power, these measurements are merged as a single measure vector \( M = [M_1, \ldots, M_n]^T \), based on the eigenvalue \( \lambda_i \) of each \( w_i \) as follows.

\[
M_i = \sum_{i} a_i |\alpha_{il}|, \quad a_i = \frac{\lambda_i}{\sum_{j} \lambda_j}
\]

Finally, \( n_i \) pixels corresponding to larger \( M_i \) s are selected to use in extracting the final features for face recognition.

The pixel selection method can be summarized as follows:

1) Obtain the projection vectors \( w_i \) s \( (i = 1, \ldots, m) \) using DCV.
2) Produce the measure vector \( M \) using \( w_i \) s, \( (i = 1, \ldots, m) \) and their eigenvalues \( \lambda_i \) s \( (i = 1, \ldots, m) \).
3) Select \( n_i \) pixels based on \( M_i \) s, \( (i = 1, \ldots, m) \).
4) Extract the final features with only the selected pixels for face recognition.

4. Experimental Results

In order to show the effectiveness of the proposed method, we have applied it to the CMU-PIE [24], Yale B [4] and Multi-PIE databases [26]. All the facial images were cropped in proportion to the distance between the eyes, which were manually detected, and were downsampled to a size of 120(pixels) x 100(pixels). The measure vector \( M \) for pixel selection was made with the images from the FERET database [25] that reflects general facial information well and has a large number of subjects. The features for recognition were extracted from these downsampled images using DCV [16] and one nearest neighbour rule was used with the \( l_2 \) norm as a classifier.

The CMU-PIE database contains images of 68 subjects with 21 illumination variations. Among them, we selected the images of only 65 subjects, because the images of the other subjects had some defects or did not include all types of illumination variations. Three images for each subject, which were taken with left side, right side and frontal illumination, were used for training, while the other 18 images not used for training were tested. Among the test images, one image under frontal illumination was used as a gallery image and the other 17 images were used as probe images.

Figure 2(a) shows the recognition rates for different number of features. Performing only histogram equalization (HIST) gave 27.9% better recognition rate than that of the raw images (RAW). The recognition rates for the shadow compensated images (SCFA [12] and SCWAD [19])
increased by 4.0% and 0.3%, respectively. As shown in Fig. 3(a), the SCFA-PS images gave the best recognition rates of 97.0%. Table 1 shows the comparisons of the proposed methods (SCFA-PS and SCWAD-PS) and various methods that deal with illumination variation for face recognition. The SCFA-PS and SCWAD-PS images are obtained by selecting 7000 and 8000 pixels, which were only 38.3% and 66.7% of the total pixels, from the SCFA and SCWAD images, respectively, using the pixel selection method. As can be seen in Table I, the proposed methods were better than other methods, showing that SCFA-PS and SCWAD-PS gave recognition rates of 97.0% and 93.6%, which were 1.0% and 1.3% better than SCFA and SCWAD respectively.

The Yale B database contains images of 10 individuals in nine poses and with 64 illuminations per pose. We used 45 facial 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 [4]. The index of the subset increases as the light source moves away from the front in taking the pictures. The images in Subset 1 were selected as a training set and the other images in Subsets 2, 3 and 4 were used as the test set.

Figure 2(b) shows the recognition rates for a different number of features which are based on the RAW, HIST, SCFA and SCWAD images. Similar to the cases of the CMU-PIE database, the histogram equalization alone gave 4.9% improvement in recognition rates compared to the RAW image, and the SCFA and SCWAD images additionally gave 3.8% and 9.3% increases, respectively. The SCFA-PS images gave the best recognition rates of 97.1%.

In Table 2, the recognition performances of the proposed methods are compared with those of various methods. The SCFA-PS and SCWAD-PS images are obtained by selecting 2000 and 9000 pixels, which were only 16.7% and 75.3% of the total pixels, from the SCFA and SCWAD images, respectively. As can be seen in Table 2, the proposed methods gave higher recognition rates compared to the other methods, showing that SCFA-PS and SCWAD-PS gave the recognition rates of 95.3% and 97.1%, which were 5.5% and 1.8% better than SCFA and SCWAD, respectively. In particular, the performance of SCWAD-PS images is most notable for Subset 4, where images are severely affected by shadows.

<table>
<thead>
<tr>
<th>No. Feature (*)</th>
<th>16 (69)</th>
<th>24 (89)</th>
<th>32 (109)</th>
<th>40 (129)</th>
<th>48 (149)</th>
<th>56 (169)</th>
<th>64 (189)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Eigenfaces w/o 1st 2 [15]</td>
<td>71.0</td>
<td>71.9</td>
<td>72.3</td>
<td>72.0</td>
<td>72.0</td>
<td>71.9</td>
<td>71.9</td>
</tr>
<tr>
<td>FLD [15]</td>
<td>66.6</td>
<td>71.6</td>
<td>74.2</td>
<td>73.8</td>
<td>76.4</td>
<td>76.6</td>
<td>76.7</td>
</tr>
<tr>
<td>LBP [9]</td>
<td>82.6</td>
<td>84.7</td>
<td>85.3</td>
<td>86.2</td>
<td>86.2</td>
<td>85.9</td>
<td>85.0</td>
</tr>
<tr>
<td>SCAD [10]</td>
<td>60.5</td>
<td>66.8</td>
<td>70.9</td>
<td>73.9</td>
<td>74.8</td>
<td>74.7</td>
<td>76.1</td>
</tr>
<tr>
<td>SCFA [12]</td>
<td>80.9</td>
<td>85.3</td>
<td>86.2</td>
<td>87.1</td>
<td>88.2</td>
<td>89.0</td>
<td>89.7</td>
</tr>
<tr>
<td>SCWAD [19]</td>
<td>86.7</td>
<td>91.1</td>
<td>92.9</td>
<td>94.7</td>
<td>94.9</td>
<td>95.6</td>
<td>96.0</td>
</tr>
<tr>
<td>SCFA-PS</td>
<td>89.3</td>
<td>91.7</td>
<td>94.1</td>
<td>96.0</td>
<td>96.3</td>
<td>97.0</td>
<td>97.0</td>
</tr>
<tr>
<td>SCWAD-PS</td>
<td>89.3</td>
<td>88.1</td>
<td>89.3</td>
<td>91.0</td>
<td>91.6</td>
<td>92.9</td>
<td>93.6</td>
</tr>
</tbody>
</table>

*: No. Feature of Eigenfaces w/o 1st 2

Table 1. Recognition rates for different methods on the cmu-pie database (%).

<table>
<thead>
<tr>
<th>Data</th>
<th>Subset 1</th>
<th>Subset 2</th>
<th>Subset 3</th>
<th>Subset 4</th>
<th>All</th>
</tr>
</thead>
<tbody>
<tr>
<td>Eigenfaces w/o 1st 2 [15]</td>
<td>100</td>
<td>100</td>
<td>80.0</td>
<td>29.2</td>
<td>74.9</td>
</tr>
<tr>
<td>FLD [15]</td>
<td>100</td>
<td>100</td>
<td>91.4</td>
<td>67.5</td>
<td>88.7</td>
</tr>
<tr>
<td>SQI [11]</td>
<td>100</td>
<td>95.8</td>
<td>96.4</td>
<td>76.7</td>
<td>91.6</td>
</tr>
<tr>
<td>SCAD [10]</td>
<td>100</td>
<td>100</td>
<td>95.7</td>
<td>86.7</td>
<td>95.1</td>
</tr>
<tr>
<td>SCFA [12]</td>
<td>100</td>
<td>100</td>
<td>94.3</td>
<td>68.3</td>
<td>89.8</td>
</tr>
<tr>
<td>SCWAD [19]</td>
<td>100</td>
<td>100</td>
<td>95.7</td>
<td>87.5</td>
<td>95.3</td>
</tr>
<tr>
<td>SCFA-PS</td>
<td>100</td>
<td>100</td>
<td>97.9</td>
<td>85.0</td>
<td>95.3</td>
</tr>
<tr>
<td>SCWAD-PS</td>
<td>100</td>
<td>100</td>
<td>97.9</td>
<td>91.7</td>
<td>97.1</td>
</tr>
</tbody>
</table>

Table 2. Recognition rates for the yale b database (%).

<table>
<thead>
<tr>
<th>No. Feature (*)</th>
<th>53 (35)</th>
<th>60 (95)</th>
<th>67 (135)</th>
<th>74 (175)</th>
<th>81 (215)</th>
<th>88 (255)</th>
<th>95 (295)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Eigenfaces w/o 1st 2 [15]</td>
<td>42.0</td>
<td>46.1</td>
<td>47.9</td>
<td>49.0</td>
<td>49.8</td>
<td>50.2</td>
<td>50.6</td>
</tr>
<tr>
<td>FLD [11]</td>
<td>81.3</td>
<td>82.2</td>
<td>82.7</td>
<td>83.0</td>
<td>83.4</td>
<td>83.5</td>
<td>83.6</td>
</tr>
<tr>
<td>SQI [11]</td>
<td>74.9</td>
<td>77.0</td>
<td>78.3</td>
<td>79.1</td>
<td>79.9</td>
<td>81.1</td>
<td>81.4</td>
</tr>
<tr>
<td>SCAD [10]</td>
<td>85.7</td>
<td>86.4</td>
<td>87.6</td>
<td>88.2</td>
<td>88.8</td>
<td>89.3</td>
<td>89.7</td>
</tr>
<tr>
<td>SCFA [12]</td>
<td>84.4</td>
<td>85.5</td>
<td>86.0</td>
<td>86.6</td>
<td>87.1</td>
<td>87.8</td>
<td>88.3</td>
</tr>
<tr>
<td>SCWAD [19]</td>
<td>83.1</td>
<td>84.5</td>
<td>86.0</td>
<td>86.7</td>
<td>87.7</td>
<td>88.7</td>
<td>89.5</td>
</tr>
<tr>
<td>SCFA-PS</td>
<td>87.8</td>
<td>88.5</td>
<td>89.5</td>
<td>89.9</td>
<td>90.6</td>
<td>91.1</td>
<td>91.6</td>
</tr>
<tr>
<td>SCWAD-PS</td>
<td>85.7</td>
<td>86.6</td>
<td>87.7</td>
<td>88.8</td>
<td>89.5</td>
<td>90.1</td>
<td>90.6</td>
</tr>
</tbody>
</table>

*: No. Feature of Eigenfaces w/o 1st 2

Table 3. Recognition rates for the multi-pie database (%).
The Multi-PIE database is recorded during four sessions over the course of six months. It contains a total 755,370 images from 337 different subjects. Individual session attendance varied between a minimum of 203 and a maximum 249 subjects. We used the 4,233 images of 249 subjects with 17 illumination variations from the first session. The recognition rates for the Multi-PIE database are shown in Table 3. The both SCFA-PS and SCWAD-PS images are obtained by selecting 7000 pixels, which were only 58.3% of the total pixels, from the SCFA and SCWAD images, respectively. As in the previous two databases, the proposed method also gave better recognition rates for the Multi-PIE database, which is the larger database, than the other methods. The recognition rates of SCFA-PS and SCWAD-PS were improved by 3.3% and 1.1% compared to SCFA and SCWAD, respectively.

5. Conclusions

Shadows caused by illumination variation can change the appearance of a face and degrade the face recognition performance. In order to alleviate the influence of shadow on face recognition, we compensated for the illumination variation by restoring the facial image deteriorated by shadow to the image similar to an image under frontal illumination using state of the art methods [12], [19]. In addition, we improved the face recognition performance by applying a pixel selection method based on discriminant analysis, which removed noisy pixels that were generated in the shadow compensation process. The experimental results for the CMU-PIE, Yale B and Multi-PIE databases showed that the proposed method improved the face recognition performance in the presence of illumination variation.

6. References


