



Face recognition based on 2D images under illumination and pose variations

Sang-Il Choi^{a,*}, Chong-Ho Choi^a, Nojun Kwak^b

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

^bDivision of Electrical and Computer Engineering, Ajou University, San 5, Woncheon-dong, Yeongtong-gu, Suwon 443-749, Republic of Korea

ARTICLE INFO

Article history:

Received 27 March 2009

Available online 27 November 2010

Communicated by T. Tan

Keywords:

Face recognition

Illumination variation

Pose variation

Shadow compensation

Linear discriminant analysis

ABSTRACT

We propose a novel 2D image-based approach that can simultaneously handle illumination and pose variations to enhance face recognition rate. It is much simpler, requires much less computational effort than the methods based on 3D models, and provides a comparable or better recognition rate.

© 2010 Elsevier B.V. All rights reserved.

1. Introduction

Face recognition has received much attention due to its theoretical challenges as well as applications in user identification, surveillance and human–computer interaction. As a result, numerous methods have been developed for face recognition in the last few decades. Among them, appearance-based approaches such as Eigenface (Turk and Pentland, 1991) and Fisherface (Belhumeur et al., 1997), perform quite 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 and pose variations.

In order to overcome the problems due to illumination variation, many approaches based on 3D models have been proposed. Basri and Jacobs (2003) represented lighting by using spherical harmonics and described the effects of Lambertian reflectance as an analogy to convolution. Lee et al. (2007) represented a face image under arbitrary illumination using a linear combination of illuminated exemplars which were synthesized from photometric stereo images of training data. For pose variation, generic 3D shape models were used by considering a uniform face shape as a tool for transforming image pixels (Liu and Chen, 2005; Zhang et al., 2006). In (Jiang et al., 2005), 3D models were reconstructed from 2D images using feature-based or image-based techniques to estimate pose. In order to cope with both illumination as well as pose variations, Georghiades and Belhumeur (2001) presented the illumination cone model and applied it to various poses using

an affine warp. Zhou and Chellappa (2005) proposed an image-based method which unified the Lambertian reflectance model (Basri and Jacobs, 2003) and the eigen light-field (Gross et al., 2002). Recently, Zhang and Samaras (2006) presented a statistical method to deal with illumination and pose variations. The methods in (Romdhani et al., 2002, 2003; Romdhani and Vetter, 2003) presented 3D morphable models to characterize human faces. Since the shadows on a face are generated due to the complex 3D structure of a human head and a pose of a human head can be characterized by pitch, roll and yaw angles, the above approaches based on 3D information perform relatively well in dealing with pose and illumination variations. However, these methods require a large computational effort (Li et al., 2004) because they are based on either the knowledge of 3D structure such as albedos and surface normals or a special physical configuration. Moreover, the method in (Zhang and Samaras, 2006) requires to mark numerous image feature points manually to process one probe image, which is time-consuming. The fitting algorithm in (Romdhani et al., 2002, 2003; Romdhani and Vetter, 2003) is complicated and 3D face models have to be captured by a 3D laser scanner or special equipment, which are significant drawbacks in implementing an online real-time face recognition system.

Other simpler methods for face recognition based on 2D images have been proposed (Liu et al., 2005; Choi et al., 2007; Ruiz-delSolar and Quinteros, 2008). Shashua and Riklin-Raviv (2001) used the quotient image, which is an illumination invariant signature, to generate face images under arbitrary illumination conditions. Xie and Lam (2005) proposed a method to eliminate the influence due to illumination variation by using a 2D shape model, which separates an input image into a texture model and a shape model for retaining shape information. The methods in (Xie and Lam,

* Corresponding author. Tel.: +82 2 880 7313; fax: +82 2 885 4459.

E-mail addresses: kara@csl.snu.ac.kr (S.-I. Choi), chchoi@csl.snu.ac.kr (C.-H. Choi), nojunk@ajou.ac.kr (N. Kwak).

2006; Ahonen et al., 2006) tries to alleviate the effect of uneven illumination by using the techniques of local normalization or local binary pattern. In order to handle pose variation, Pentland et al. (1994) proposed a view-based eigenspace method and Huang et al. (2000) used a neural network with a view-specific eigenface for face recognition. Gross et al. (2002) presented the concept of light field to characterize the continuous pose space, and Liu (2004) proposed a Gabor-based kernel PCA using Gabor wavelets and a kernel. However, most of 2D image-based methods deal with either illumination or pose variation, and so it is difficult to apply them directly when both illumination and pose variations are present.

In this paper, we propose a new approach based on 2D images for handling illumination and pose variations simultaneously. We are motivated by the view-based face recognition methods (Liu, 2004; Pentland et al., 1994) that use different feature space for each different pose class. We first propose a simple pose estimation method based on 2D images, which uses a suitable classification rule and image representation to classify a pose of a face image. In order to represent the characteristic of each pose class, we transform a face image into an edge image, in which facial components such as eyes, nose and mouth in the image are enhanced. Then, the image can be assigned to a pose class by a classification rule in a low-dimensional subspace constructed by a feature extraction method. On the other hands, unlike general classification problems, pose classes can be placed sequentially from left profile to right profile in the pose space, and we can make use of the order relationship between classes. Therefore, in order to model the continuous variation in head pose, we investigate the performance of feature extraction methods for regression problems (Li, 1991, 1992; Kwak et al., 2008) and classification problems (Belhumeur et al., 1997; Fukunaga, 1990), where classes have an order relationship. Second, we propose a shadow compensation method that compensates for illumination variation in a face image so that the image can be recognized by a face recognition system designed for images under normal illumination condition. Generally, human faces are similar in shape in that they are comprised of two eyes, a nose and a mouth. Each of these components forms a shadow on a face, showing distinctive characteristics depending on the direction of light in a fixed pose. 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. Since the direction of light can change continuously, it is insufficient to represent the illumination variation with the shadow characteristic from only one discretized light category as in (Choi et al., 2007; Choi and Choi, 2007). Thus, we use more than one shadow characteristics to compensate for illumination variation by giving an appropriate weight to each estimated light category. Furthermore, we extend the compensation method that works not only for the frontal pose class but also for other pose classes as well. These shadow compensated images in each pose class are used for face recognition.

The proposed method consists of three parts which are pose estimation, shadow compensation and face identification (see Fig. 1). For a face image with multiple variations, the pose of the face image is estimated by using the proposed pose estimation method. After assigning a face image to an appropriate pose class, the face image is processed by the shadow compensation procedure customized for each pose class. These shadow compensated images are used for face identification by a classification rule.

The proposed method has the following advantages compared to other face recognition methods under illumination and pose variations. Unlike most of 2D image-based methods that deal with individual variation separately, the proposed method handles both illumination and pose variations. Moreover, the proposed method,

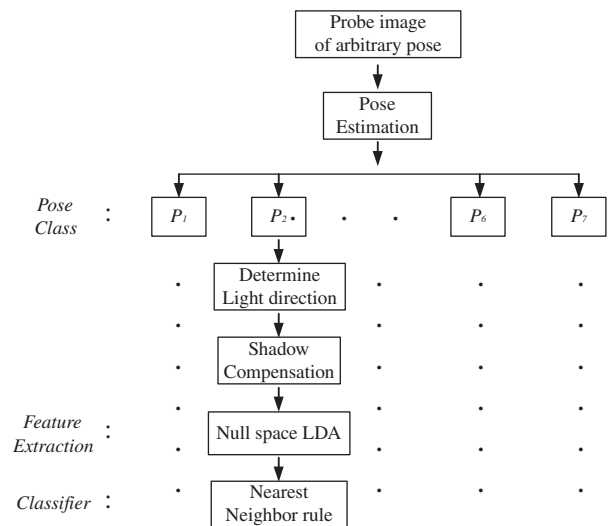


Fig. 1. The proposed architecture for face recognition.

which is based on 2D images, does not require to estimate the face surface normals or the albedos, and thus there is no need for any special equipment such as a 3D laser scanner (Romdhani et al., 2002, 2003; Romdhani and Vetter, 2003) or complicated computation. The proposed shadow compensation method also does not include image warping or iteration process. These make the proposed recognition system much simpler to implement, and this simplicity is an important factor for performing a face recognition system in real-time. Even for the simplicity of the proposed method, it works quite well and its recognition performance is better than or comparable to the algorithms based on 3D models which require 3D information.

The rest of this paper is organized as follows. Section 2 explains how to assign a suitable pose class to an image. Section 3 explains how to compensate for the shadow in face images in each pose class. Section 4 presents the experimental results of the proposed method and its comparison with other methods. The conclusion follows in Section 5.

2. Pose estimation

In a view-based face recognition (Liu, 2004; Pentland et al., 1994), the pose estimation is to classify head orientation into one of several discrete orientation classes, e.g., frontal, left/right profiles, etc. Among the pose estimation methods based on 2D images such as the geometric methods, detector array methods, appearance template methods and subspace methods (Murphy-Chutorian and Trivedi, 2009), we use the subspace method that projects a probe image into a low-dimensional subspace to estimate its pose. We first divide the pose space into several pose classes from left profile to right profile. In a view-based face recognition, the pose estimation stage is important for face recognition performance because it is at the first stage in face recognition system to determine the pose class of an image.

To make pose estimation more reliable against the variations in subjects and environmental changes, it is necessary to find the characteristics that are mostly affected by pose variation. We use the geometrical distribution of facial components for pose estimation because the locations of facial components change depending on the pose. With this information, we can estimate the pose and determine the pose class by a classification rule.

In order to remove the redundant information for pose estimation, we transform a face image to an edge image. An edge image is

an effective representation of the distribution of facial components. Fig. 2 shows raw images and the corresponding edge images of different poses. As shown in Fig. 2(a), raw images contain not only the distribution of facial components but also other information such as texture, gray-level intensity, and appearance variation of subjects, and these can act as noise in estimating pose. On the contrary, as can be seen in Fig. 2(b) and (c), only the rough shapes of facial components are present in the edge images, while the other information disappears, i.e., the locations of facial components are enhanced in an edge image.

Shadow produces a large changes in a raw image, which makes it more difficult to properly classify the pose. There are two types of shadows that occur on a face image; one is an attached shadow and the other is a cast shadow. The attached shadow occurs in the regions of surface facing away from the light source. The cast shadow is caused by the blockage of light from a light source by some part of a subject, and projected onto another part of the subject itself. The edge images are effective especially when shadow is present on face images due to illumination variation. Since the attached shadow caused by illumination variation changes slowly in the spatial domain, an edge image, which is a high pass filtered image, reduces the effect of illumination variation. However, the edge images may be sensitive to cast shadow which contains high frequency components, and some traces of cast shadow may remain along with facial components in edge images. These traces in edge images can be alleviated in the process of constructing the subspace for pose classification. By adding the edge images under various illumination conditions in training set, the pose estimation can be reliably performed for the images under illumination variation.

Several edge detection algorithms have been proposed in image processing area. Among them, we adopt the Sobel edge detector (Gonzales and Woods, 2002) which uses two convolution kernels, one to detect changes in vertical contrast and another to detect changes in horizontal contrast. The Sobel edge detector is very simple and the edges produced by the Sobel edge detector enhance only the geometrical distribution of facial components eliminating unnecessary edge shapes compared to the Canny edge detector (Canny, 1986), which is another well known edge detector.

By applying a discriminant feature extraction method to these Sobel edge images from the images of training set, a subspace is constructed for each of K pose classes $\{P_k | k = 1, 2, \dots, K\}$ ($K = 7$ in the experiments in Section 4). The subspace for classification of the pose class is constructed by using a discriminant feature extraction method. The pose of each image projected into the subspace is classified by using the one nearest neighborhood rule with the l_2 norm as the distance metric, and a pose class P_k , $k = 1, 2, \dots, K$ is assigned to each image.

3. Weighted shadow compensation

3.1. Estimation of the direction of light

Unlike the methods in (Wang et al., 2007; Zhang and Samaras, 2006) which require 3D information to generate novel images, we obtain the shadow characteristic from 2D images without additional 3D information and then compensate for the shadow. For this, we first estimate the direction of light for each pose class. Since the shape of a human face is more convex in azimuth than in elevation, we divide the directions of light into L categories $\{C_l | l = 1, 2, \dots, L\}$ (here, $L = 7$) from the left side to the right side (see Fig. 6). C_4 implies that the light comes from the front. We denote the gray-level intensity of a face image (see Fig. 3) of $X(\text{height}) \times Y(\text{width})$ pixels as $I_{m,n}^{(k,l)}(x,y) \in \mathbb{R}^{X \times Y}$, where the subscripts $m (= 1, 2, \dots, M)$ and $n (= 1, 2, \dots, N_{(k,l)})$ denote the n th image of the m th individual when the direction of light belongs to category C_l in the pose class P_k . (We used $X = 120$ and $Y = 100$ in the experiments in Section 4.) Hence, the superscript (k,l) denotes that the pose class of the image is P_k and the direction of light belongs to C_l . To estimate the direction of light, we make a binary image with a threshold $(1/XY) \sum_{x=1}^X \sum_{y=1}^Y I_{m,n}^{(k,l)}(x,y)$ which is the average value of the gray-level intensities, for each face image (Choi et al., 2007). In order to reduce the influence of the background on a face image, we take a square mask of 80×80 (pixels) that only covers the center part of a face image of 120×100 (pixels).

Fig. 3 shows some examples of raw images and the corresponding binary images. As can be seen in Fig. 3(b), the black area moves from left to right depending on the light source, and so binary

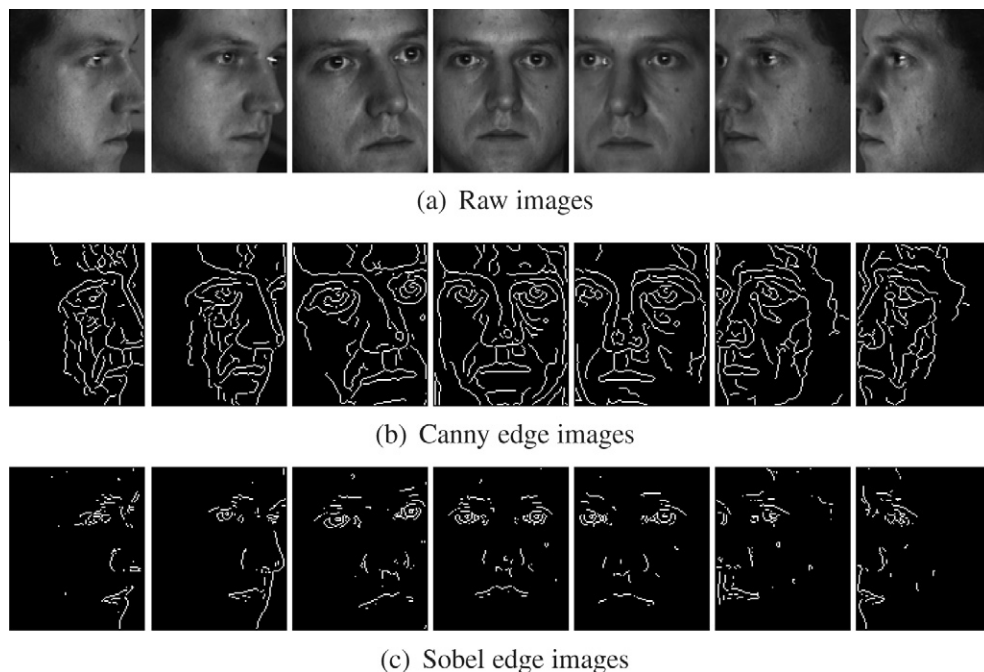


Fig. 2. Images under various pose classes (P_1, \dots, P_7) and corresponding edge images.

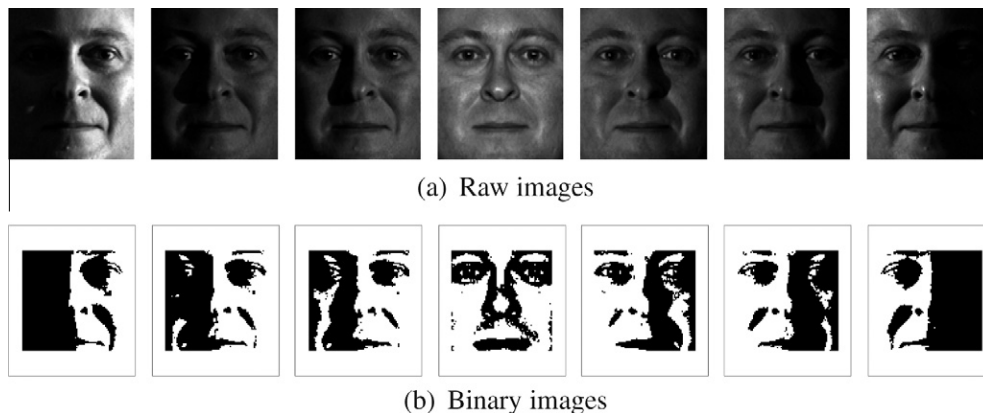


Fig. 3. Binary images for different light directions (P_4): (a) images under various illuminations; (b) corresponding binary images.

images can be effectively used to classify the direction of the light. With these binary images, we assign a category value $l (= 1, 2, \dots, L)$ to the light category C_l following the same procedure as in the pose estimation. We constructed the feature space by using PCA + LDA with the binary images transformed from the images in the CMU-PIE database (Sim et al., 2003), which was used for training at the pose estimation stage. We evaluated this light direction classification system with the Yale B database (Georghiadis and Belhumeur, 2001), which provides information about the location of the flashlight for each image. Fig. 4 shows the distribution of the horizontal angle between the direction of light and the frontal direction in each light category estimated by the proposed classification procedure. The vertical axis represents the angle between the light source and the frontal direction and the horizontal axis represents the light category. The positive value in Fig. 4 implies that the light source was from the right of the subject whereas the negative value implies otherwise. The category changes from C_1 to C_7 as the light source moves from left to right. In the figure, each vertical bar denotes one standard deviation of the estimated angle in both sides. As can be seen in Fig. 4, the mean of the angle that belongs to each category increases linearly as the index of the category increases, which implies that the direction of light was estimated very well by using the binary images. For a given angle in Fig. 4, it may belong to more than two categories, but it does not matter because the three nearest categories are used in the weighted shadow compensation that will be explained in the following. For images that belong to other pose classes, the direction of light was also estimated as the same procedure above.

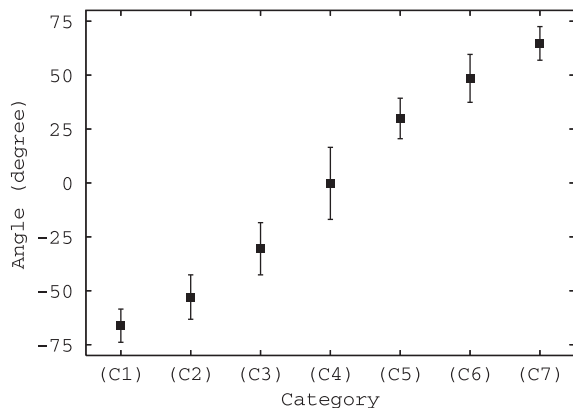


Fig. 4. The distribution of the angle between the direction of light and the frontal direction in each light category for P_4 .

3.2. Weighted shadow compensation

In order to alleviate the influence of the shadow as much as possible, all the images are pre-processed by the histogram equalization (Gonzales and Woods, 2002). Since most human faces are similar in shape, we can assume that the shadows on facial images in the same pose class and the same illumination category are also similar in shape, and the difference image between the images with and without the shadows contains the information on the illumination condition. We select one of the images under the frontal illumination in each pose P_k as a reference image ($I_m^{(k,ref)}$) (In the experiment, $ref = 4$). The gray-level intensity $I_m^{(k,l)}(x,y)$ at pixel (x,y) varies depending on the light category, and is different from that of $I_m^{(k,ref)}(x,y)$. We define the intensity difference between the images of $I_m^{(k,ref)}$ and $I_m^{(k,l)}$ at each pixel (x,y) as follows.

$$D_{m,n}^{(k,l)}(x,y) = I_m^{(k,ref)}(x,y) - I_m^{(k,l)}(x,y) \quad (1)$$

$$x = 1, 2, \dots, X, \quad y = 1, 2, \dots, Y$$

The intensity difference $D_{m,n}^{(k,l)}$ of one person is insufficient to compensate for the intensity differences of another person's images under various illumination conditions because $D_{m,n}^{(k,l)}$ contains information about the illumination condition as well as unique features of each individual. In order to compensate for the intensity difference due to illumination variation, we need to eliminate the influence of features that are innate to each individual. Therefore, we define the average intensity difference $D_A^{(k,l)}$ for the category C_l in the pose class P_k as follows:

$$D_A^{(k,l)}(x,y) = \frac{1}{MN^{(k,l)}} \sum_{m=1}^M \sum_{n=1}^{N^{(k,l)}} D_{m,n}^{(k,l)}$$

Note that there are no subscripts m or n in $D_A^{(k,l)}$. 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_l , it can be applied to any face image in compensating for the shadow formed by the light belonging to the category C_l in the pose class P_k . The average intensity difference, which is shown in Fig. 5, was made from the images for each pose class in the CMU-PIE database.

Since the direction of light can change continuously, it is too optimistic to expect that one average intensity difference contains sufficient information for shadow compensation in each pose class and light direction category. In the case of the example shown in Fig. 6, the direction of light for the test image is on the border between two categories C_2 and C_3 . Even though the direction of light belongs to category C_2 , the shadow on the image has shadow characteristics of both C_2 and C_3 . Thus, in order to handle such cases, we

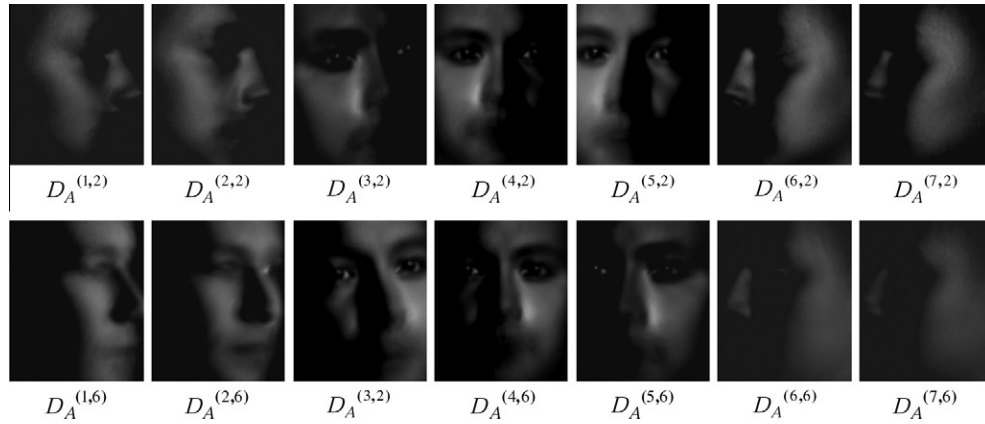


Fig. 5. Two average intensity differences for each pose (top, bottom).

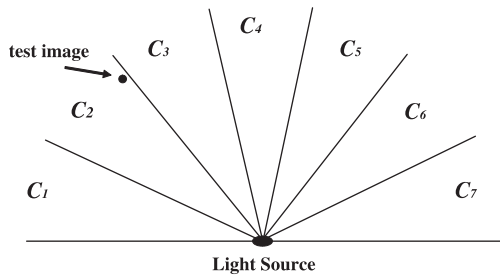


Fig. 6. Light direction categories.

assign weights to the average intensity differences depending on the category. After calculating the distances between the binary image of a test image and the binary images in each category C_i , the three nearest distances ($dist_i^{NN}$, $i = 1, 2, 3$) and their corresponding categories C_i , $i = 1, 2, 3$ are selected. The weight w_i , which implies the degree of contribution to the compensation, is determined based on these three nearest distances as the following.

$$w_i = \frac{dist_{(4-i)}^{NN}}{\sum_{i=1}^3 dist_i^{NN}} \quad (2)$$

Then, we obtain the shadow compensated image, $I_{m,n}^{C(k,l)}$ of $I_{m,n}^{(k,l)}$ with the average differences as follows:

$$I_{m,n}^{C(k,l)}(x,y) = I_{m,n}^{(k,l)}(x,y) + \sum_{i=1}^3 w_i D_A^{(k,l_i)}(x,y) \quad (3)$$

Fig. 7 shows some examples of the raw images from the CMU-PIE database and their shadow compensated images in various poses by using (3). As can be seen in Fig. 7(b), most of the shadow in raw images is removed in the shadow compensated images. Although the compensated images are slightly blurred in the compensation process, it does not compromise the final face recognition rates because most of the useful information for classification are located in the low frequency region (Jing et al., 2005). This shadow compensation method works well even for the images when there are horizontal and vertical variations in light source directions as in the Yale B database (see Table 8).

Fig. 8(a) shows three images, which are the raw image, the histogram equalized image and the compensated image. Fig. 8(b) shows probability mass functions (pmf), which are normalized histograms, of gray-level intensity values corresponding to these images. For the raw image, the components of the histogram are concentrated in the low side of the intensity scale. Although the

components with small intensity values spread over a wider range by the histogram equalization, a large portion still remains in the low side of the intensity scale. In the histogram of the shadow compensated image, pixels are quite uniformly distributed over the entire range of intensity values. It is known that an image, whose pixels tend not only to occupy the entire range of gray levels but also to be distributed uniformly, will have an appearance of high contrast and will exhibit a large variety of gray tones (Gonzales and Woods, 2002). Therefore, a face recognition system is expected to perform better with the shadow compensated images than the histogram equalized images. This will be confirmed in the experiments in Section 4.

4. Experimental results

We applied the proposed method to the CMU-PIE and Yale B databases to evaluate its performance. The facial components, i.e., the center of each eye and the bottom of the nose, were manually marked and their coordinates were recorded. Each face was cropped to include only the face and rotated based on these coordinates, and then rescaled to a size of 120×100 (pixels).

4.1. Pose estimation

In order to show the effectiveness of the proposed pose estimation method, we evaluated the performance of pose classification on the CMU-PIE database. The CMU-PIE database contains more than 40,000 facial images of 68 individuals, 21 illumination conditions, 13 poses and four different expressions. Among them, we selected the images of 65 individuals with seven pose indices (c22, c02, c05, c27, c29, c14, c34), so that the entire set consists of 21 images in 7 pose classes of 65 individuals ($21 \times 7 \times 65$ images in total). The training set was constructed by randomly choosing 3 images from each pose for each individual ($3 \times 7 \times 65$ images), while the test set consisted of all the other images ($18 \times 7 \times 65$ images). In order to estimate the pose of a face image, each of the seven pose classes was assigned a numerical target value from 1 (left profile) to 7 (right profile). We repeated this test three times by changing the composition of training and test sets, and computed the average classification rate (image indices: '02', '08', '17' for the first training set; '05', '08', '14' for the second training set; '03', '07', '18' for the third training set).

Unlike the general classification problem, since the pose classes can be placed sequentially from left profile to right profile in the pose space, there is an order relationship between classes and the distance between classes can be used as a measure of class similarity. For example, consider a pose estimation problem which

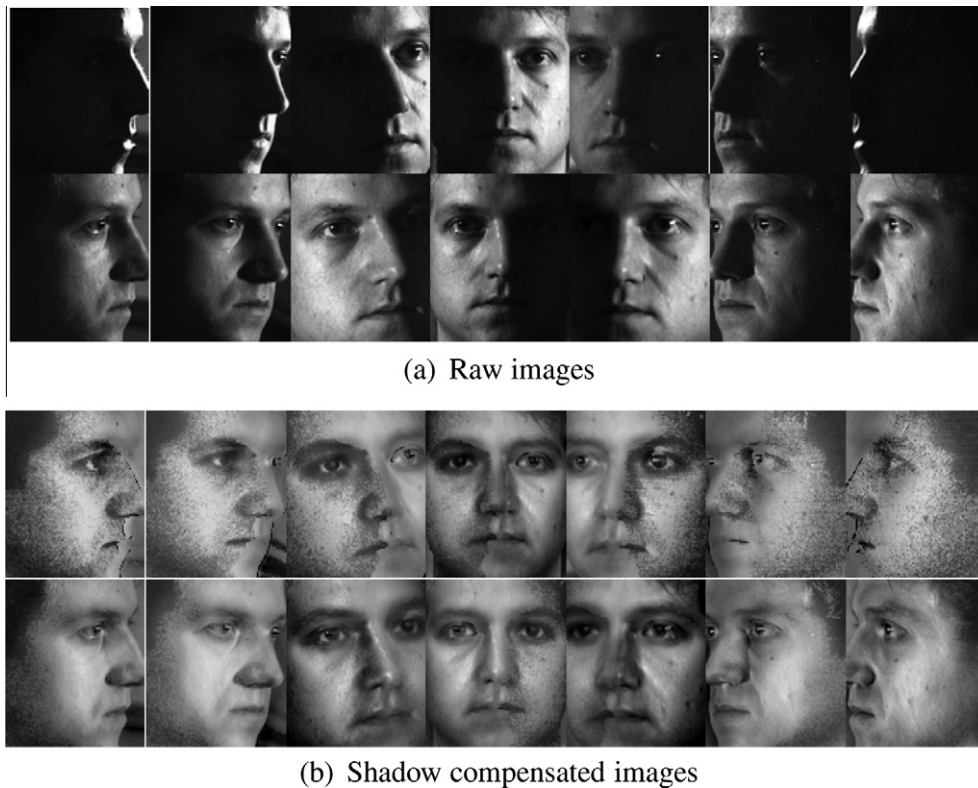


Fig. 7. Examples of the shadow compensated images in each pose.

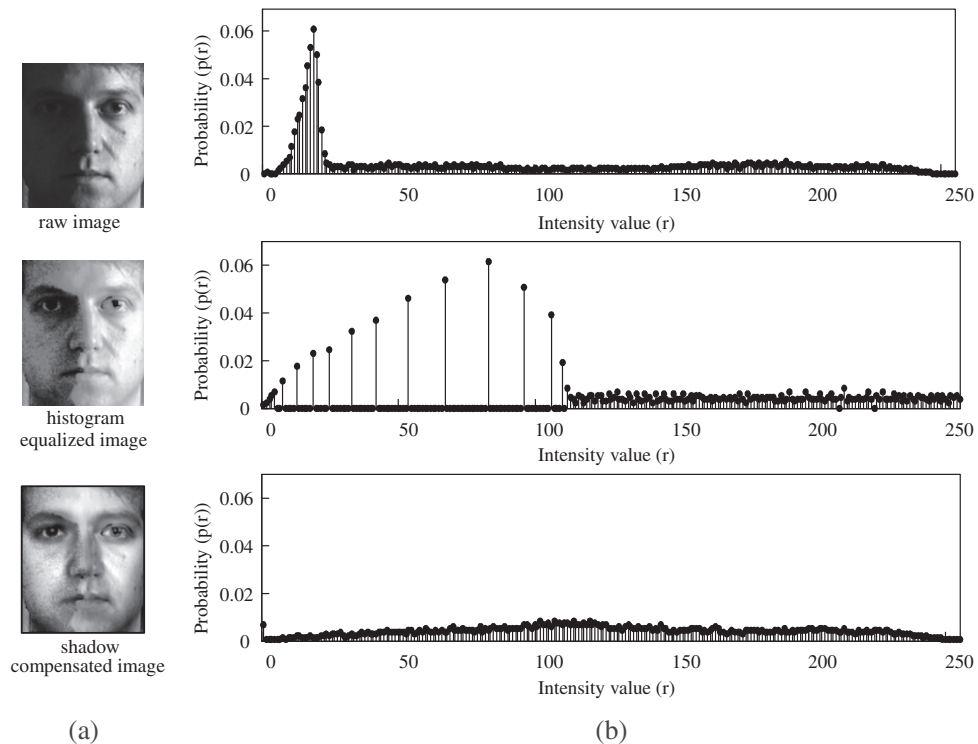


Fig. 8. Normalized histograms of the images: (a) three images; (b) normalized histograms corresponding to the images.

consists of three pose classes 'front (0°)', 'half profile (45°)' and 'profile (90°)'. Then, a 'profile' image is closer to the 'half profile' image than the 'front' image. If a classifier happens to misclassify a 'profile' image, it would be better to classify it into a 'half profile'

image than a 'front' image. Thus, we can make use of the order relationship between classes in extracting features. These types of classification problems can be also considered to be similar to regression problems by assigning a numerical target value to each

of the pose classes. Therefore, in order to extract useful features for discriminating the pose of a face image, we investigated the performance of feature extraction methods for regression problem such as the Sliced Inverse Regression (SIR) (Li, 1991) and the Principal Hessian Directions (PHD) (Li, 1992), along with the conventional LDA which has been very successful for classification problems and LDA-r (Kwak et al., 2008), which is a variant of LDA to effectively handle classification problems with order relationship between classes.

When the pixels of an image are used as input variables of a 12,000-dimensional input space, and the Small Sample Size (SSS) problem occurs in extracting the features for pose estimation. To resolve this problem, in all the feature extraction methods, we pre-processed the dataset with PCA to reduce the dimension of input space. In SIR, the parameter S (number of slices) was set to 10, and in LDA-r, the weight function was used as $f(x) = \sqrt{||x| - \tau|}$ where the threshold τ was a multiple of the standard deviation σ_y of target variable y such that $\tau = \alpha\sigma_y$ and α was set to 0.1.

Table 1 shows the classification rates of pose estimation for the test images using several feature extraction methods. The numbers in the parentheses are the number of features. As can be seen in Table 1, the overall classification rates of PHD and SIR ($S = 10$), which makes use of order relationship between classes and are good for regression problems, are both below 50.0%. This is because the order relationship between pose classes is obscured in the subspace obtained with raw images which contain not only the information related to pose variation but also other information such as texture and individual characteristics. On the contrary, LDA and LDA-r, which are good for classification problems, give the overall classification rates of more than 84.0% and 88.0%, respectively.

Table 2 shows the classification rates for the edge image of the test images. As can be seen in Table 2, the overall classification rates for all the feature extraction methods are higher than their counterparts in Table 1. This is because the order relationship between classes becomes more prominent in edge images than in raw images, improving the classification performance for all the feature extraction methods. Note that the classification rates for SIR and PHD are 79.1% and 81.2%, respectively, which are better

Table 1
Classification rates in pose estimation for raw images (%).

Feature extraction	c22	c02	c05	c27	c29	c14	c34	Overall
LDA(6)	84.1	88.8	85.0	82.1	90.6	95.5	91.2	88.2
LDA-r(200)	79.2	86.7	81.9	77.8	85.0	92.9	88.8	84.6
SIR(1200)	62.0	37.9	39.4	30.9	37.0	48.9	51.9	44.0
PHD(1200)	61.9	42.5	41.0	31.5	38.6	49.9	53.3	45.4

Table 2
Classification rates in pose estimation for edge images by the Canny and Sobel edge detectors (%).

Feature extraction	c22	c02	c05	c27	c29	c14	c34	Overall
<i>(a) Canny detector</i>								
LDA(6)	97.1	87.4	89.8	87.2	89.1	87.0	98.5	90.9
LDA-r(200)	90.7	84.8	92.5	87.9	88.3	91.7	97.5	90.5
SIR(1200)	73.9	75.9	76.0	70.3	78.5	89.0	90.1	79.1
PHD(1200)	77.5	77.9	77.4	72.4	79.9	88.2	94.9	81.2
<i>(b) Sobel detector</i>								
LDA(6)	96.7	94.3	90.8	88.0	90.0	90.8	98.5	92.7
LDA-r(200)	95.8	95.3	92.9	88.3	88.1	92.5	98.5	93.1
SIR(1200)	87.1	88.2	84.2	79.9	84.3	86.1	96.8	86.6
PHD(1200)	87.6	88.9	84.8	81.0	85.3	86.0	96.1	87.1

Table 3
The number of misclassified poses for various distance d_p .

Method	$d_p = 1$	$d_p = 2$	$d_p = 3$	$d_p = 4$	$d_p = 5$
Raw image	120	12	7	50	2
Edge image (Canny)	136	8	4	1	1
Edge image (Sobel)	58	13	9	0	0

by more than 35% compared to those in Table 1. As expected, the Sobel edge detector performs better than the Canny edge detector for the pose classification problem. This is because, as shown in Fig. 2(b) and (c), the location change of the facial components depending on the pose class is more apparent by the Sobel edge detector than by the Canny edge detector. For the edge images produced by the Sobel edge detector, among all the feature extraction methods, LDA-r gives the best classification rate of 93.1%, which is 6.5%, 6.0% and 0.4% more than those of SIR, PHD and LDA, respectively.

In order to evaluate the effectiveness of using edge images instead of raw images in pose classification, we checked the distance d_p between the correct and misclassified pose class for each misclassified pose. An image in P_4 is closer to the images in P_5 compared to the images in P_6 , and if a classifier misclassifies an image in P_4 , it would be better to be classified into P_5 than P_6 . The distance d_p shows how bad the classification result is when a pose class is misclassified. In Table 3, the numbers are the number of misclassified poses for distance $d_p = s$, $s = 1, 2, 3, 4, 5$ when applying LDA-r to the raw images in the first training set ('02', '08', '17') and their edge images produced by the Canny and Sobel edge detectors. When $d_p = s$, it implies that there are $(s - 1)$ pose classes between the correct and assigned pose classes. As can be seen in Table 3, most classification errors for the edge images produced by both the Canny and Sobel edge detectors occur for $d_p = 1$, whereas the classification errors for the raw images occur in large numbers for $d_p = 2, 3, 4, 5$, which can severely degrade the performance of a view-based face recognition system.

4.2. Illumination variation

We selected the images of 65 individuals with seven pose indices (c22, c02, c05, c27, c29, c14, c34) as in Section 4.1 to evaluate the performance of the proposed shadow compensation. For each pose index, eighteen images of each individual, which were not used for training, were tested. Among the test images, one image under the frontal illumination for each pose index was used as a gallery image and the other seventeen images were used as probe images. The performance of a face recognition system is greatly affected by the selection of training images for face identification. Therefore, we selected three different groups of training sets for face identification, depending on the intensity of illumination variation (see Fig. 9).

The features were extracted from the shadow compensated images by the null space method (NLDA) (Cevikalp et al., 2005), which is widely used for face recognition. With the features, the one nearest neighborhood rule was used as a classifier with the Euclidean distance (l_2) as the distance metric.

Table 4 shows the recognition rate for the images under illumination variation when the pose class information was already given. As can be seen in Table 4, training with the images under intense illumination variations (the first group) gives the best results for all the image sets. This is because large illumination variation in training images can help to deal with various illumination conditions. Even when the correct pose class is given, the recognition rates for the raw images (I_{Raw}) in the top rows for each group drops drastically to 65.5% in average for the third group, in which

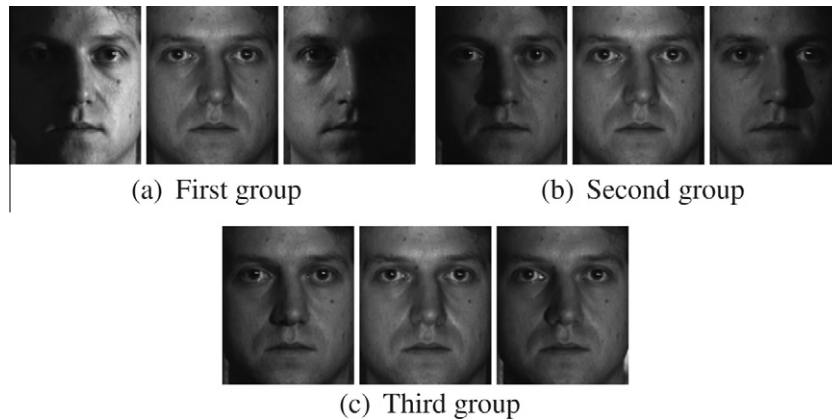


Fig. 9. Training images in each group: (a) first group; (b) second group; (c) third group.

Table 4
Recognition rates under illumination variation (%).

Group	Data set	c22	c02	c05	c27	c29	c14	c34	Overall
1	I_{Raw}	98.7	94.6	82.6	90.5	83.6	90.8	99.2	91.4
	I_{Hist}	99.7	99.7	99.7	100	99.1	99.9	99.9	99.7
	I^c	100	100	100	100	99.7	100	100	100
2	I_{Raw}	93.3	95.7	82.7	90.7	77.7	99.0	98.6	91.1
	I_{Hist}	99.3	99.4	98.7	98.7	99.1	99.2	99.0	99.1
	I^c	99.6	99.7	99.5	99.7	99.4	99.8	99.9	99.7
3	I_{Raw}	64.1	60.9	60.9	66.2	66.9	68.2	71.0	65.5
	I_{Hist}	91.2	91.0	80.3	88.3	95.8	96.6	94.8	91.1
	I^c	93.9	91.9	89.8	90.6	96.9	98.3	98.0	94.2

the illumination variation in the training images is the smallest. On the other hand, after the shadow compensation procedure, the overall recognition rates are much better (100%, 99.7% and 94.2%), i.e., the recognition rates for the proposed shadow compensated images decrease more slowly than those for the raw images or the histogram equalized images. This implies that the proposed shadow compensation procedure prevents the recognition rate from degrading rapidly even when the training images taken are under small illumination variation.

In addition, we computed the relative distance $d_{rel} = d_2/d_1$ for the images in the third group, where d_1 and d_2 are the distances between a probe image and its first and second nearest neighbors in the gallery images, respectively. d_{rel} shows the robustness of the face recognition system, and $\log_{10}(d_2/d_1)$ is called the confidence measure (Price and Gee, 2005). Thus, a higher d_{rel} indicates that the recognition result is more reliable. As d_{rel} decreases to a value near one, the probability of incorrect recognition increases. We can improve the confidence of a decision made on face recognition by accepting the results when d_{rel} is greater than a certain threshold value and rejecting the results otherwise. In other words, if d_{rel} is lower than a predefined threshold, the probe image is considered to be inadequate for face identification and is rejected by the classifier. Fig. 10 shows the correct recognition rate versus the rejection rate for various stages of compensation in the third group. As illustrated in the figure, the recognition rates improve as the rejection rate increases. For a given rejection rate in all the poses, the recognition rates for the shadow compensated images are always higher than those for the histogram equalized images. This means that the recognition system becomes more reliable through the proposed shadow compensation procedure in all pose classes.

We compared the proposed method with two other shadow compensation methods, which are mLBP (Modified Local Binary Pattern) (Froba and Ernst, 2004) and SQI (Self-Quotient Image)

(Wang et al., 2004). Table 5 shows that the recognition rates for training images of the second group. As can be seen from Table 5, I_{Raw} , I_{SQI} and I_{mLBP} give the average recognition rates of 91.1%, 96.0% and 98.4% for overall pose indices, respectively, whereas the average recognition rate for the images compensated by the proposed method (I^c) increases by 8.6%, 3.7% and 1.3% compared to I_{Raw} , I_{SQI} and I_{mLBP} , respectively.

4.3. Illumination and pose variations

In order to deal with both illumination and pose variations, a face recognition system was constructed with some of the images from the CMU-PIE database. For pose estimation, the features were extracted from the first training set in Section 4.1 by using LDA-r. And the appropriate features to determine the direction of light were extracted following to the procedures described in Section 3.1. For face identification, which is the third stage in Fig. 1, three images under different illumination conditions for seven pose indices (c22, c02, c05, c27, c29, c14, c34) were used for each individual as training images in constructing the feature space by using NLDA. Since the face identification stage comes after the pose classification stage as shown in Fig. 1, the error in pose classification can directly affect the result of face recognition. Since most of the misclassified poses were the immediate neighbor of the correct pose class as confirmed in Section 4.1, we used all the images of the correct pose class and its neighboring pose classes to construct the feature space in each pose class in the face identification stage. We expect that the resultant feature space makes the performance of face recognition more robust. For example, even if a probe image that actually belongs to pose class P_1 is classified to pose class P_2 , we can reduce the recognition error due to pose misclassification because the feature space for P_1 partially reflects the information of P_2 . We tested the rest of the images of the seven pose indices,

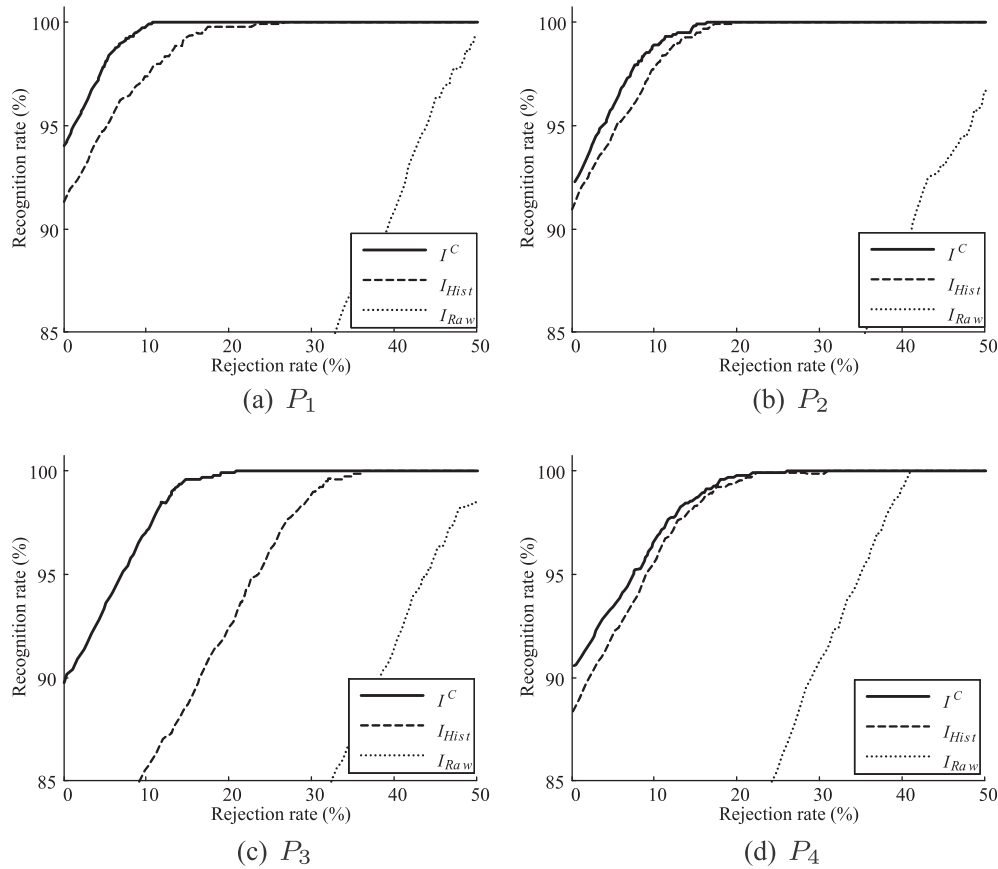


Fig. 10. Recognition rate versus rejection rate: (a) pose class P_1 ; (b) pose class P_2 ; (c) pose class P_3 ; (d) pose class P_4 .

Table 5

Recognition rates under illumination variation of different methods (%).

Method	c22	c02	c05	c27	c29	c14	c34	Overall
I_{Raw}	93.3	95.8	82.7	90.7	77.7	99.0	98.6	91.1
SQL (I_{SQL})	94.8	96.8	97.4	97.8	94.9	95.3	94.8	96.0
mLBP (I_{mLBP})	98.0	97.8	98.6	99.6	98.2	98.0	98.7	98.4
Proposed method (I^C)	99.6	99.7	99.5	99.7	99.4	99.8	99.9	99.7

Table 6

Recognition rates under illumination and pose variations (%).

Pose of probe	c22	c02	c05	c27	c29	c14	c34
Recognition rate	99.1	99.8	99.2	99.6	99.9	99.8	98.5
Pose of probe	c25	c37	c09	c07	c11	c31	Overall
Recognition rate	97.0	97.8	99.9	97.9	99.0	98.4	98.9

which were not used for training, and additionally tested the images of 62 individuals in the other six pose indices (c25, c37, c09, c07, c11, c31). Among 65 individuals in the six pose indices, three individuals did not have all the images of all pose and illumination variations, and so we excluded the images of those individuals in this experiment.

Table 6 shows the recognition rates for the images of the CMU-PIE database under both illumination and pose variations. As can be seen in Table 6, the average recognition rate over all the poses and illumination categories is 98.9% and the recognition rate does not change much depending on the pose class of the probe image. After investigating incorrectly recognized images, the cause of wrong recognition can be categorized into the following three

cases which are: (1) the images with the pose misclassification distance $d_p \geq 2$ were severely distorted in the shadow compensation procedure; (2) although most shadows in the face images were removed through the weighted shadow compensation, some cast shadows still remained around the nose; (3) in the process of face alignment by using the three coordinates of the facial components (both eyes and nose), inaccurate alignment caused distortion in the compensated image.

There are some noteworthy results in the face recognition under illumination and pose variations using the 3D models (Romdhani et al., 2006), and so we also compared the proposed method with the methods based on 3D models. Table 7, which gives the comparison results, shows that the proposed method is

Table 7
Recognition rates of different methods on the CMU-PIE database (%).

Method	Pose of a probe image			Time taken for identification
	Front (c_{27})	Half profile (c_{05})	Profile (c_{22})	
3DMM (Romdhani and Vetter, 2005)	99.9	99.3	89.4	2.5 min
Spherical-basis MM (Zhang and Samaras, 2006)	96.5	96.7	80.6	4 min
Zhou and Chellappa (2005)	97.0	88.0	52.0	1.5 s
The proposed method	99.5	99.4	99.0	1.5–1.7 s.

better than all the other methods. Although the experiments in (Romdhani and Vetter, 2005; Zhang and Samaras, 2006; Zhou and Chellappa, 2005) were performed with the CMU-PIE database, their results can not be directly compared to Table 7 due to the difference in the architecture of the recognition system. However, the method in (Romdhani and Vetter, 2005) requires much more computational effort in fitting 3D models to the gallery and probe images. Since it compares a probe image, which is frontal, to gallery images, which may not be frontal, the recognition rate varies from 89.4% to 99.9% depending on the pose matching between the gallery and probe images (Romdhani et al., 2006). The method in (Zhang and Samaras, 2006) requires a set of 60 manually marked image feature points on the inner part of a face in order to estimate the 3D shape, and its performance was also sensitive (80.6%–96.7%) to the pose matching between the gallery and probe images as in (Romdhani and Vetter, 2005). Even though the method in (Zhou and Chellappa, 2005) gave a recognition rate of 97% when the gallery and probe images were in frontal pose, the recognition rate degrades to 52% when the probe images were profile and the gallery images were frontal. On the other hand, the proposed method does not need the pose matching between the gallery and probe images because the pose for the probe image is estimated first, and then the probe image is projected to the feature space of its pose class. Table 6 shows that the variation of recognition rates in the proposed method is very small throughout all the pose classes. The running time is also an important factor in implementing a real face recognition system. The proposed method took 1.5–1.7 s to identify one input image by using MATLAB, whereas the methods in (Romdhani and Vetter, 2005) and (Zhang and Samaras, 2006) took 2.5 min and 4.5 min, respectively (Romdhani et al., 2006). Although the experiments were performed on different machines, the two orders of difference in processing time shows that the proposed method requires much less computational effort than those based on 3D models.

In order to see whether the face recognition system constructed with the CMU-PIE database can perform reliably for a different database, we tested the system with the images in the Yale B database. The Yale B database contains images of ten individuals in nine poses and 64 illuminations per pose. We used 45 face images for each subject in each pose which were further subdivided into four subsets (subset i , $i = 1, 2, 3, 4$) depending on the direction of light as in (Georghiades and Belhumeur, 2001). The direction of light source varied in both horizontally and vertically. The index of the subset increases as the light source moves away from the front during picture taking. Table 8 shows the comparison of recognition rates for the Yale B database. The proposed method gave recognition rates of 92.3%–95.8% for all of the poses. Only the fourth method in (Georghiades and Belhumeur, 2001) performed better, resulting in recognition rates of 94.5%–99.1%, but it was necessary to construct 3D models which required large computational effort. Also, it is important to note that the training set and the test set came from different databases in this experiment while these sets

Table 8
Recognition rates of different methods on the Yale B database (%).

Method	Pose of a probe image		
	front (Pose1)	12°(Pose2 – 6)	24°(Pose 7 – 9)
Correlation (Brunelli and Poggio, 1993)	70.9	24.2	12.8
Cone Approx. (Georghiades and Belhumeur, 2001)	100	34.7	18.0
Correlation with Planar Transformations (Georghiades and Belhumeur, 2001)	62.4	53.0	36.4
Cone Approx. with Planar Transformations (Georghiades and Belhumeur, 2001)	99.3	84.5	51.9
Cone Approx. with Full Pose (Georghiades and Belhumeur, 2001)	99.1	97.3	94.5
Proposed method	95.8	94.8	92.3

were from the same database in the others experiments in Table 8. This indicates that the performance of the proposed method is still expected to provide a good recognition rate for images from different database not used for training.

5. Conclusions

This paper proposes a novel approach to reduce the performance degradation of face recognition caused by illumination and pose variations. We constructed a feature space for each pose class by using a feature extraction method and compensated for illumination variation in each pose class. In order to estimate the pose and direction of light, we determined the pose class and light direction category based on the edge images and binary images, respectively. Since human faces are similar in shape, we can compensate for shadow variation in a face by adding a weighted average intensity difference depending on the direction of light. These compensated images can be used without making any modification to any other face recognition algorithms based on 2D images. By using appropriate feature spaces and the shadow compensation method, the recognition rate reached almost 99% on average for the CMU-PIE database under illumination and pose variations. Moreover, the compensated image makes the face recognition system reliable for all pose classes. Since the proposed method is based on 2D images and does not need to estimate 3D shape, it is computationally much more efficient than the other methods based on 3D models. Its recognition rate is also better than or comparable to other face recognition systems based on 3D models. This paper demonstrates that the face recognition system based on 2D images can be more efficient and effective under pose and illumination variations.

Acknowledgments

This work was supported by Mid-career Researcher Program through NRF grant funded by the MEST (400-20100014) and was partly supported by Korea Research Foundation Grant funded by Korean Government (KRF-2010-0004908).

References

- Ahonen, T., Hadid, A., Pietikäinen, M., 2006. Face description with local binary patterns: Application to face recognition. *IEEE Trans. Pattern Anal. Machine Intell.* 28 (12), 2037–2041.
- Basri, R., Jacobs, D.W., 2003. Lambertian reflectance and linear subspace. *IEEE Trans. Pattern Anal. Machine Intell.* 25 (2), 218–233.
- Belhumeur, P.N., Hespánha, J.P., Kriegman, D.J., 1997. Eigenfaces vs. fisherfaces: Recognition using class specific linear projection. *IEEE Trans. Pattern Anal. Machine Intell.* 19 (7), 711–720.

- Brunelli, R., Poggio, T., 1993. Face recognition: Features versus templates. *IEEE Trans. Pattern Anal. Machine Intell.* 15 (10), 1042–1052.
- Canny, J., 1986. A computational approach to edge detection. *IEEE Trans. Pattern Anal. Machine Intell.* 8 (6), 679–698.
- Cevikalp, H., Neamtu, M., Wilkes, M., Barkana, A., 2005. Discriminative common vectors for face recognition. *IEEE Trans. Pattern Anal. Machine Intell.* 27 (1), 914–919.
- Choi, S.I., Choi, C.H., 2007. An effective face recognition under illumination and pose variations. In: *Proc. IEEE Internat. Joint Conf. on Neural Networks – IJCNN 2007*. pp. 914–919.
- Choi, S.I., Kim, C., Choi, C.H., 2007. Shadow compensation in 2D images for face recognition. *Pattern Recognition* 40 (7), 2118–2125.
- Froba, B., Ernst, A., 2004. Face detection with the modified census transform. In: *Proc. sixth Internat. Conf. on Face and Gesture Recognition – FG 2004*. pp. 91–96.
- Fukunaga, K., 1990. *Introduction to Statistical Pattern Recognition*, second ed. Academic Press.
- Georgiades, A.S., Belhumeur, P.N., 2001. From few to many: Illumination cone models for face recognition under variable lighting and pose. *IEEE Trans. Pattern Anal. Machine Intell.* 23 (23), 643–660.
- Gonzales, R.C., Woods, R.E., 2002. *Digital Image Processing*, second ed. Prentice Hall.
- Gross, R., Matthews, I., Baker, S., 2002. Eigen light-fields and face recognition across pose. In: *Proc. fifth Internat. Conf. on Face and Gesture Recognition – FG 2002*. pp. 3–9.
- Huang, F.J., Zhou, Z., Zhang, H.J., Chen, T., 2000. Pose invariant face recognition. In: *Proc. fourth Internat. Conf. on Face and Gesture Recognition – FG 2000*. pp. 245–250.
- Ruiz-delSolar, J., Quinteros, J., 2008. Illumination compensation and normalization in eigenspace-based face recognition: A comparative study of different pre-processing approaches. *Pattern Recognition Lett.* 29 (14), 1966–1979.
- Jiang, D., Hu, Y., Yan, S., Zhang, L., Gao, W., 2005. Efficient 3D reconstruction for face recognition. *Pattern Recognition* 38 (6), 787–798.
- Jing, X.Y., Tang, Y.Y., Zhang, D., 2005. A fourier-LDA approach for image recognition. *Pattern Recognition* 38 (3), 453–457.
- Kwak, N., Choi, S.I., Choi, C.H., 2008. Feature extraction for regression problems and an example application for pose estimation of a face. In: *Proc. fifth Internat. Conf. on Image Analysis and Recognition – ICIAR 2008*. pp. 435–444.
- Lee, S.W., Lee, S.H., Moon, S.H., Lee, S.W., 2007. Face recognition under arbitrary illumination using illuminated exemplars. *Pattern Recognition* 40 (5), 1605–1620.
- Li, K., 1991. Sliced inverse regression for dimension reduction. *J. Amer. Statist. Assoc.* 86, 316–342.
- Li, K., 1992. On principal hessian directions for data visualization and dimension reduction: Another application of stein's lemma. *J. Amer. Statist. Assoc.* 87, 1025–1039.
- Li, Q., Ye, J., Kambhmettu, C., 2004. Linear projection methods in face recognition under unconstrained illuminations: A comparative study. In: *Proc. Internat. Conf. on Computer Vision and Pattern Recognition – CVPR 2004*, vol. 2, pp. 474–481.
- Liu, C., 2004. Gabor-based kernel PCA with fractional power polynomial models for face recognition. *IEEE Trans. Pattern Anal. Machine Intell.* 26 (5), 572–581.
- Liu, D.H., Lam, K.M., Shen, L.S., 2005. Illumination invariant face recognition. *Pattern Recognition* 38 (7), 1705–1716.
- Liu, X., Chen, T., 2005. Pose-robust face recognition using geometry assisted probabilistic modeling. In: *Proc. Internat. Conf. on Computer Vision and Pattern Recognition*. pp. 502–509.
- Murphy-Chutorian, E., Trivedi, M., 2009. Head pose estimation in computer vision: A survey. *IEEE Trans. Pattern Anal. Machine Intell.* 31 (4), 609–626.
- Pentland, A., Moghaddam, B., Starner, T., 1994. View-based and modular eigenspaces for face recognition. In: *Proc. Internat. Conf. on Computer Vision and Pattern Recognition – CVPR 1994*. pp. 84–91.
- Price, J.R., Gee, T.F., 2005. Face recognition using direct, weighted linear discriminant analysis and modular subspaces. *Pattern Recognition* 38 (2), 209–219.
- Romdhani, S., Blanz, V., Vetter, T., 2002. Face identification by fitting a 3D morphable model using linear shape and texture error functions. In: *Proc. European Conf. on Computer Vision – ECCV 2002*. pp. 3–19.
- Romdhani, S., Blanz, V., Vetter, T., 2003. Face recognition based on fitting a 3D morphable model. *IEEE Trans. Pattern Anal. Machine Intell.* 25 (9), 1–14.
- Romdhani, S., Ho, J., Vetter, T., Kriegman, D.J., 2006. Face recognition using 3-D models: Pose and illumination. *Proceedings of the IEEE* 94 (11), 1977–1999.
- Romdhani, S., Vetter, T., 2003. Efficient, robust and accurate fitting of a 3D morphable model. In: *Proc. Internat. Conf. on Computer Vision – ICCV 2003*. pp. 59–66.
- Romdhani, S., Vetter, T., 2005. Estimating 3D shape and texture using pixel intensity, edges, specular highlights, texture constraints and a prior. In: *Proc. Internat. Conf. on Computer Vision and Pattern Recognition – CVPR 2005*, vol. 2, pp. 986–993.
- Shashua, A., Riklin-Raviv, T., 2001. The quotient image: Class based re-rendering and recognition with varying illuminations. *IEEE Trans. Pattern Anal. Machine Intell.* 23 (2), 129–139.
- Sim, T., Baker, S., Bsat, M., 2003. The CMU pose, illumination, and expression database. *IEEE Trans. Pattern Anal. Machine Intell.* 25 (12), 1615–1618.
- Turk, M., Pentland, A., 1991. Eigenfaces for recognition. *J. Cognitive Neurosci.* 3 (1), 71–86.
- Wang, H., Li, S., Wang, Y., 2004. Face recognition under varying lighting conditions using self quotient image. In: *Proc. sixth Internat. Conf. on Face and Gesture Recognition – FG 2004*. pp. 819–824.
- Wang, Y., Liu, Z., Hua, G., Wen, Z., Zhang, Z., Samaras, D., 2007. Face re-lighting from a single image under harsh lighting conditions. In: *Proc. Internat. Conf. on Computer Vision and Pattern Recognition – CVPR 2007*. pp. 1–8.
- Xie, X., Lam, K.M., 2005. Face recognition under varying illumination based on 2D face shape model. *Pattern Recognition* 38 (2), 221–230.
- Xie, X., Lam, K.M., 2006. An efficient illumination normalization method for face recognition. *Pattern Recognition Lett.* 27 (6), 609–617.
- Zhang, L., Samaras, D., 2006. Face recognition from a single training image under arbitrary unknown lighting using spherical harmonics. *IEEE Trans. Pattern Anal. Machine Intell.* 28 (3), 351–363.
- Zhang, X., Gao, Y., Leung, M., 2006. Automatic texture synthesis for face recognition from single views. In: *Proc. Internat. Conf. on Pattern Recognition*. pp. 1151–1154.
- Zhou, S.K., Chellappa, R., 2005. Image-based face recognition under illumination and pose variations. *J. Opt. Soc. Amer.* 22 (2), 217–229.