Illumination Compensation and Normalization for Robust Face Recognition Using Discrete Cosine Transform in Logarithm Domain

Weilong Chen, Meng Joo Er, *Member, IEEE*, and Shiqian Wu, *Member, IEEE*

***Abstract*—This paper presents a novel illumination normalization approach for face recognition under varying lighting conditions. In the proposed approach, a discrete cosine transform (DCT) is employed to compensate for illumination variations in the logarithm domain. Since illumination variations mainly lie in the low-frequency band, an appropriate number of DCT coefficients are truncated to minimize variations under different lighting conditions. Experimental results on the Yale B database and CMU PIE database show that the proposed approach improves the performance significantly for the face images with large illumination variations. Moreover, the advantage of our approach is that it does not require any modeling steps and can be easily implemented in a real-time face recognition system.**

***Index Terms*—Discrete cosine transform, face recognition, illumination normalization, logarithm transform.**

I. Introduction

Face recognition has attracted significant attention because of its wide range of applications [1]. Recently, more researchers focus on robust face recognition such as face recognition systems invariant to pose, expression and illumination variations. Illumination variation is still a challenging problem in face recognition research area, especially for appearance-based approaches. The same person can appear greatly different under varying lighting conditions. A variety of approaches have been proposed to solve the problem [3]–[16]. These approaches can be generally classified into three main categories.

- **Preprocessing and Normalization**: In this approach, face images are preprocessed using some image processing techniques to normalize the images to appear stable under different lighting conditions. For instance, histogram equalization (HE), Gamma correction, logarithm transform, etc. are widely used for illumination normalization [3], [4]. However, nonuniform illumination variation is still difficult to deal with using these global processing techniques. Recently, adaptive histogram equalization (AHE) [2], region-based histogram equalization (RHE) [3], and block-based histogram equalization (BHE) [5] have also been proposed to cope with nonuniform illumination variations. Although recognition rates on face databases with nonuniform illumination variations can be improved compared with the HE, their performances are still not satisfactory. In [13], by combining symmetric shape-from-shading (SSFS) and a generic three-dimensional (3-D) model, the performance of face recognition under varying illuminations is enhanced. However, this method is only efficient for exact frontal face images and it is assumed that all faces share a similar common shape. In [3], the authors proposed a normalization method called quotient illumination relighting (QIR). This method is based on the assumption that the lighting modes of the images are known or can be estimated.
- **Invariant Feature Extraction**: This approach attempts to extract facial features which are invariant to illumination variations. Edge maps, derivatives of the gray-level and Gabor-like filters are investigated in [9]. However, empirical studies show that none of these representations are sufficient to overcome image variations due to changes in the direction of illumination. Another well-known feature extraction method is called Fisherface [also known as linear discriminant analysis (LDA)] which linearly projects the image space to a low-dimensional subspace to discount variations in lighting and facial expressions [11]. But, this method is a statistical linear projection method which largely relies on representativeness of the training samples. In [12], the quotient image is regarded as the illumination invariant signature image which can be used for face recognition under varying lighting conditions. Bootstrap database is required for this method and the performance degrades when dominant features between the bootstrap set and the test set are misaligned.
- **Face Modeling**: Illumination variations are mainly due to the 3-D shape of human faces under lighting in different directions. Recently, some researchers attempt to construct a generative 3-D face model that can be used to render face images with different poses and under varying lighting conditions [6], [7], [10] and [14]. A generative model called illumination cone was presented in [6], [7]. The main idea of this method is that the set of face images in fixed pose but under different illumination conditions can be represented using an illumination convex cone which can be constructed from a number of images acquired under variable lighting conditions and the illumination cone can be approximated in a low-dimensional linear subspace. In [10], the authors showed that the set of images of a convex Lambertian object obtained under a variety of lighting conditions can be well approximated by a 9D linear subspace. One of the drawbacks of the model-based approaches is that a number of images of the subject under varying lighting conditions or 3-D shape information are needed during the training phase. This drawback limits its applications in practical face recognition systems. In addition, existing model-based approaches assume that the human face is a convex object, i.e., the casting shadows are not considered. The specularity problem is also ignored even though the human face is not a perfect Lambertian surface.

To the best of our knowledge, one ideal way of solving the illumination variation problem is to normalize a face image to a standard form under uniform lighting conditions. In fact, the human visual system usually cares about the main features of a face, such as the shapes and relative positions of the main facial features, and ignores illumination changes on the face while recognizing a person. Accordingly, in this paper, we propose an illumination normalization approach to remove illumination variations while keeping the main facial features unimpaired. The key idea of the proposed approach is that illumination variations can be significantly reduced by truncating low-frequency discrete cosine transform (DCT) coefficients in the logarithm DCT domain. Our approach can be categorized into the first approach group although feature extraction can be carried out directly in the logarithm DCT domain.

Manuscript received October 16, 2004; revised March 9, 2005. This paper was recommended by Associate Editor Maja Pantic.

W. Chen is with the Computer Control Lab, Nanyang Technological University, Singapore 639798.

M. J. Er is with the Intelligent Systems Centre, Nanyang Technological University, Singapore 637533 (e-mail: EMJER@ntu.edu.sg).

S. Wu is with the Institute for Infocomm Research, Nanyang Technological University, Singapore 119613.

Digital Object Identifier 10.1109/TSMCB.2005.857353

The remainder of this paper is organized as follows. In Section II, we describe the illumination normalization approach in the logarithm DCT domain in detail. Experimental results and discussions are presented in Section III. Finally, conclusions are drawn in Section IV.

II. Illumination Normalization in the Logarithm DCT Domain

A. Logarithm Transform

Logarithm transform is often used in image enhancement to expand the values of dark pixels [9] and [21]. Here, we show why illumination compensation should be implemented in the logarithm domain. In a simple situation, the image gray level $f(x,y)$ can be assumed to be proportional to the product of the reflectance $r(x,y)$ and the illumination $e(x,y)$ [23], i.e.,

$$f(x,y) = r(x,y) \cdot e(x,y). \tag{1}$$

To our knowledge, the Retinex algorithm is related to the reflectance constancy [17]. The invariant property of reflectance ratio has been applied in object recognition [18]. Since the reflectance is a stable characteristic of facial features, our goal is to recover the reflectance of faces under varying illumination conditions. Taking logarithm transform on (1), we have

$$\log f(x,y) = \log r(x,y) + \log e(x,y). \tag{2}$$

It follows from (2) that in the logarithm domain, if the incident illumination $e(x,y)$ and the desired uniform illumination e' are given (e' is identical for every pixel of an image), we have

$$\begin{aligned}\log f'(x,y) &= \log r(x,y) + \log e' \\ &= \log r(x,y) + \log e(x,y) - \epsilon(x,y) \\ &= \log f(x,y) - \epsilon(x,y)\end{aligned} \tag{3}$$

where

$$\epsilon(x,y) = \log e(x,y) - \log e'$$

and $f'(x,y)$ is the pixel value under desired uniform illumination. From (3), we can conclude that the normalized face image can be obtained from the original image by using an additive term $\epsilon(x,y)$ called compensation term which is the difference between the normalized illumination and the estimated original illumination in the logarithm domain.

B. Discrete Cosine Transform

There are four established types of Discrete Cosine Transforms (DCT's), i.e., DCT-I, DCT-II, DCT-III, and DCT-IV. The DCT-II is more widely applied in signal coding because it is asymptotically equivalent to the Karhunen–Loeve Transform (KLT) for Markov-1 signals with a correlation coefficient that is close to one [24]. For example, JPEG image compression is also based on the DCT-II [25]. The DCT-II is often simply referred to as "the DCT". The 2D $M \times N$ DCT is defined as follows:

$$C(u,v) = \alpha(u)\alpha(v) \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} f(x,y) \times \cos\left[\frac{\pi(2x+1)u}{2M}\right] \cos\left[\frac{\pi(2y+1)v}{2N}\right] \tag{4}$$

and the inverse transform is defined as

$$f(x,y) = \sum_{u=0}^{M-1} \sum_{v=0}^{N-1} \alpha(u)\alpha(v) C(u,v) \times \cos\left[\frac{\pi(2x+1)u}{2M}\right] \cos\left[\frac{\pi(2y+1)v}{2N}\right] \tag{5}$$

where

$$\alpha(u) = \begin{cases} \frac{1}{\sqrt{M}}, & u = 0 \\ \sqrt{\frac{2}{M}}, & u = 1, 2, \cdots, M-1 \end{cases}$$

$$\alpha(v) = \begin{cases} \frac{1}{\sqrt{N}}, & v = 0 \\ \sqrt{\frac{2}{N}}, & v = 1, 2, \cdots, N-1. \end{cases}$$

In the JPEG image compression standard, original images are initially partitioned into rectangular nonoverlapping blocks (8 × 8 blocks) and then the DCT is performed independently on the subimage blocks [25]. In the proposed approach, the DCT is performed on the entire face image to obtain all frequency components of the face image.

C. Illumination Compensation

Given a face image, illumination variations can be well compensated by adding or subtracting the compensation term $\epsilon(x,y)$ of (3) in the logarithm domain if we know where illumination variations and important facial features are. However, facial feature detection is a nontrivial task especially for face images with large illumination variations. Nevertheless, in a face image, illumination usually changes slowly compared with the reflectance except some casting shadows and specularities on the face. As a result, illumination variations mainly lie in the low-frequency band. Since we attempt to recognize faces using reflectance characteristic, illumination variations can be reduced by removing low-frequency components. It should be noted that only face images without hair are considered in our approach because the intensity of human's hair is a kind of low-frequency feature which will be impaired by discarding low-frequency components of face images. However, a human's hair is a kind of unstable feature which will change greatly with time. Therefore, in many face recognition systems, human's hair is not regarded as a kind of important facial feature.

The DCT can be used to transform an image from spatial domain to frequency domain. Besides, it can be implemented using a fast algorithm which significantly reduces the computational complexity. Low-frequency components of a face image can be removed simply by setting the low-frequency DCT coefficients to zero. Evidently, the resulting system works like a high-pass filter. Since illumination variations are mainly low-frequency components, we can estimate the incident illumination on a face by using low-frequency DCT coefficients. It follows from (4) that setting the DCT coefficients to zero is equivalent to subtracting the product of the DCT basis image and the corresponding coefficient from the original image. If n low-frequency DCT coefficients are set to zero, we have

$$\begin{aligned}F'(x,y) &= \sum_{u=0}^{M-1} \sum_{v=0}^{N-1} E(u,v) - \sum_{i=1}^{n} E(u_i,v_i) \\ &= F(x,y) - \sum_{i=1}^{n} E(u_i,v_i)\end{aligned} \tag{6}$$

where

$$E(u,v) = \alpha(u)\alpha(v)C(u,v)\cos\left[\frac{\pi(2x+1)u}{2M}\right]\cos\left[\frac{\pi(2y+1)v}{2N}\right].$$

Since illumination variations are expected to be in the low-frequency components, the term $\sum_{i=1}^{n} E(u_i,v_i)$ can be approximately regarded as the illumination compensation term. It follows from (3) that the term $F'(x,y)$ in (6) is just the desired normalized face image in the logarithm domain. Therefore, discarding low-frequency DCT coefficients in the logarithm domain is equivalent to compensating for illumination variations. This is the reason why DCT should be implemented in the logarithm domain.

The first DCT coefficient (i.e., the DC component) determines the overall illumination of a face image. Therefore, the desired uniform illumination can be obtained by setting the DC coefficient to the same value, i.e.,

$$C(0,0) = \log \mu \cdot \sqrt{MN} \tag{7}$$

where $C(0,0)$ is the DC coefficient of the logarithm image. For the convenience of understanding and visualization, we normally choose a value of μ near the middle level of the original image. In other words, the normal face has an average gray level of μ. It should be noted that we do not regard the skin color as a kind of facial feature because it is unstable when illumination changes. For example, the face of a black man normally has an average gray level below μ. It is actually regarded as a normal face under weak illumination conditions. It follows from (3) and (6) that the difference between the original DC component and the normalized DC component, together with the other discarded low-frequency AC components, approximately make up the compensation term $\epsilon(x, y)$.

D. Logarithm Domain Versus Original Domain

Since illumination variations mainly lie in the low-frequency band, we can approximately estimate them using the low-frequency DCT basis images and their corresponding coefficients. As a simple example to illustrate the idea, note that the half-lighted face image is highly correlated with the (0,1)th basis image. In other words, the illumination difference on a half-lighted face can be approximately estimated from the (0,1)th DCT coefficient. Therefore, the invariant reflectance can be obtained by discarding the (0,1)th DCT coefficient. As illustrated in Fig. 1, the facial features in the dark area of the original image are recovered much better by applying DCT on the logarithm image. In fact, only the brightness of the image is adjusted by discarding DCT coefficients of the original image, whereas discarding DCT coefficients of the logarithm image will adjust the illumination and recover the reflectance characteristic of the face.

E. Logarithm Image for Recognition

Human faces are not perfect Lambertian surfaces. In some cases, there are specularities on a face image which do not lie in the low-frequency band. Moreover, some shadows also lie in the same frequency band as the main facial feature. As a consequence, illumination variations on some small areas may not be correctly compensated by discarding the low-frequency coefficients. For example, some small areas under high illumination level may be incorrectly adjusted to even higher level. As we know, the addition operation in the logarithm domain is equivalent to multiplication in the original domain. If the logarithm image is restored to the original one, incorrect adjustment will make it even worse. Accordingly, in our approach, logarithm images are directly used for recognition, i.e., the inverse logarithm transform step is skipped. In fact, there are also physiological evidences that the response of the retina cells can be approximated as a log function of the intensity [9]. Fig. 1(c) and (d) shows the reconstructed nonlogarithm image and the logarithm image after discarding the (0,1)th DCT coefficient to correct half-lighted illumination, respectively.

F. Discarding DCT Coefficients

As aforementioned, low-frequency DCT coefficients which are highly related to illumination variations should be discarded. There remains another issue: which and how many DCT coefficients should be discarded in order to obtain the well normalized face image?

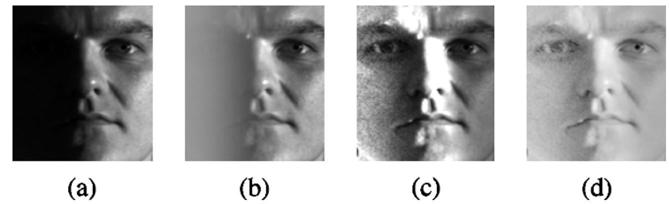


Fig. 1. (a) Original image. (b) Reconstructed image by applying DCT on the original image and discarding the (0,1) th DCT coefficient. (c) Reconstructed image by applying DCT on the logarithm image and discarding the (0,1) th DCT coefficient ($\mu = 100$). (d) Reconstructed logarithm image by applying DCT on the logarithm image and discarding the(0,1) DCT coefficient (i.e., (c) without the inverse logarithm transform).

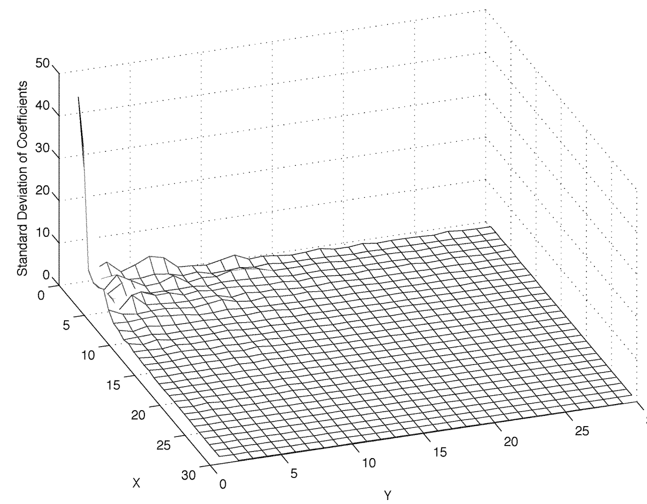


Fig. 2. Standard deviations of the logarithm DCT coefficients.

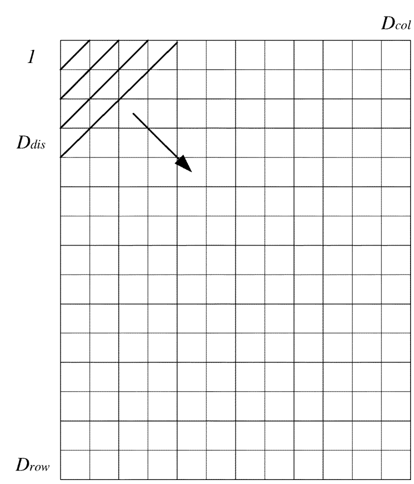


Fig. 3. Manner of discarding DCT coefficients.

Fig. 2 shows the standard deviations of the logarithm DCT coefficients which are calculated from 64 face images of the same subject (only the first 30 × 30 coefficients are shown). As we can see from Fig. 2, standard deviations of the coefficients with great magnitude are mainly located in the upper-left corner of the DCT coefficient matrix. Accordingly, illumination variations of face images can be reduced by discarding these low-frequency coefficients [the DC coefficient is set to a constant value according to (7)]. The manner of discarding DCT coefficients is shown in Fig. 3.

TABLE I
SUBSETS DIVIDED ACCORDING TO LIGHT SOURCE DIRECTION

Subsets	1	2	3	4	5
Lighting angle (°)	$0 \sim 12$	$13 \sim 25$	$26 \sim 50$	$51 \sim 77$	> 77
Number of images	70	120	120	140	190

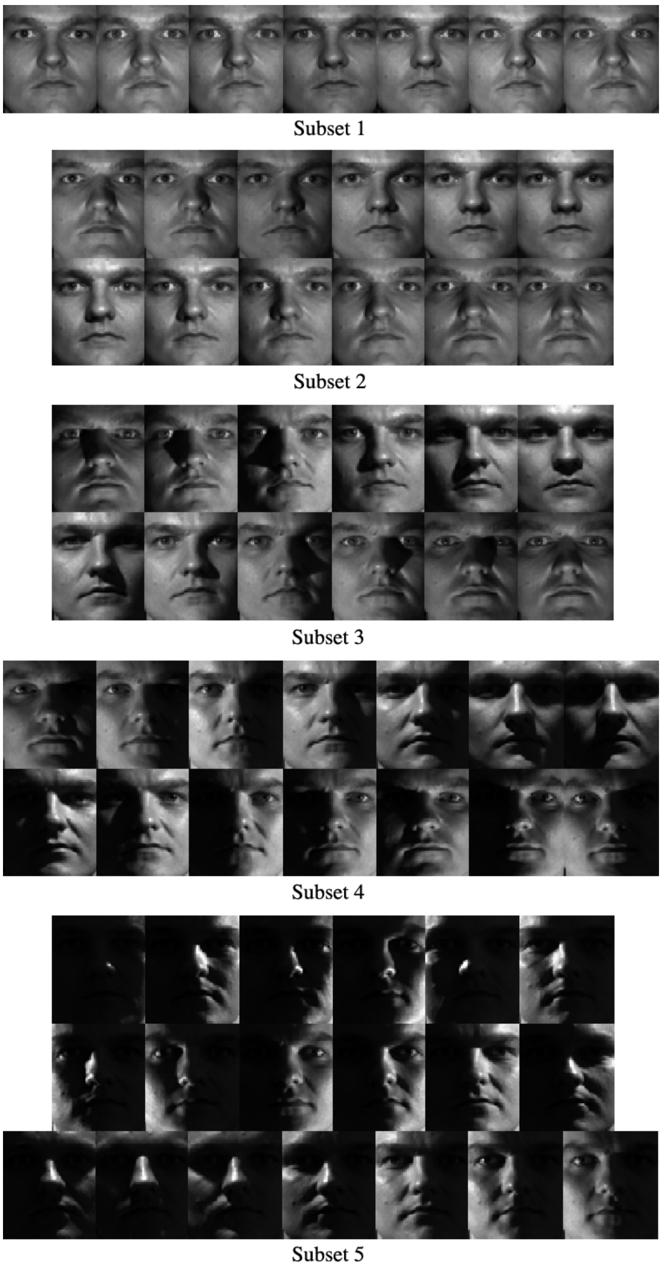


Fig. 4. Sample images of an individual divided into five subsets.

III. Experimental Results and Discussions

A. Face Database

In this paper, the Yale Face Database B and the CMU PIE Face Database are both used to evaluate the proposed approach. These two face databases contain face images with large illumination variations.

1) Yale Face Database B: There are ten individuals under 64 different lighting conditions for nine poses in the database. Since we are only concerned with the illumination problem in this paper, frontal face images under varying lighting conditions are used. As shown in Table I, the face images are divided into five subsets according to the angle between the light source direction and the camera axis. Interested readers may refer to [7] for more detailed information of the database. In the experiments, face images are all cropped and aligned in accordance with [7]. The distance between eyes is equal to four sevenths of the cropped window width and the face was centered along the vertical direction so that the two imaginary horizontal lines passing through the eyes and mouth are equidistant from the center of the cropped window. In this paper, all the face images are rescaled to the size of 120 × 105. Fig. 4 shows the images of one individual divided into five subsets based on different lighting conditions. We use Subset 1 as the training set and other subsets are used for testing.

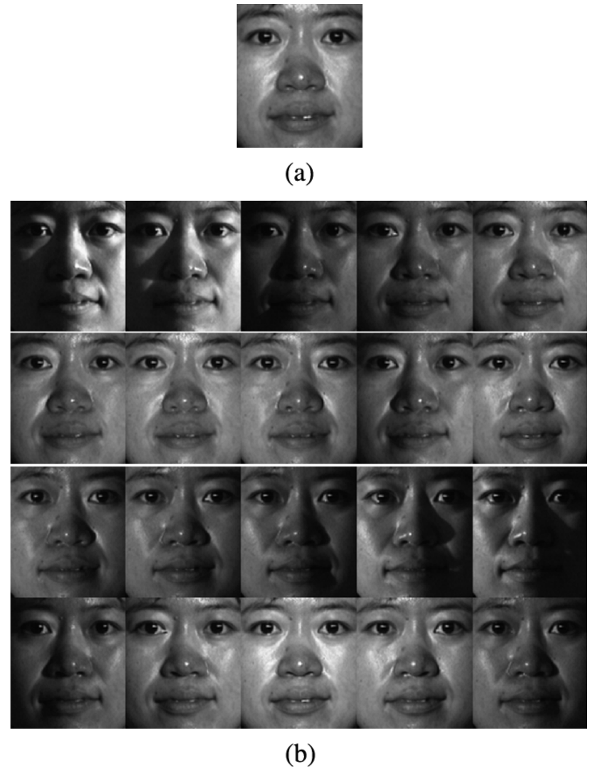


Fig. 5. Sample images of an individual in CMU PIE database. (a) Training. (b) Testing.

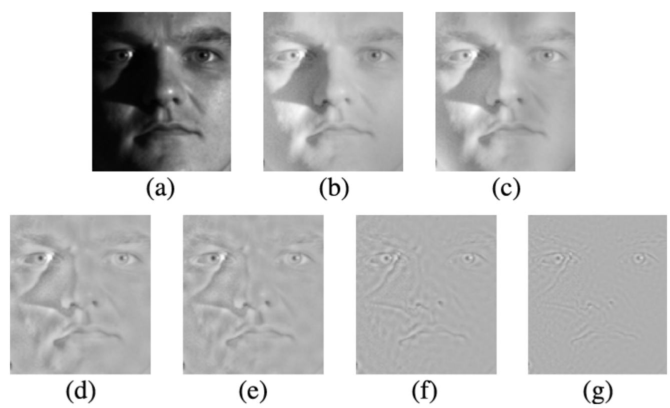


Fig. 6. Normalized logarithm images with different $D_{\rm dis}$: (a) original image; (b) $D_{\rm dis} = 3$; (c) $D_{\rm dis} = 6$; (d) $D_{\rm dis} = 15$; (e) $D_{\rm dis} = 20$; (f) $D_{\rm dis} = 35$; and (g) $D_{\rm dis} = 50$.

2) CMU PIE Face Database: In the CMU PIE database, there are 68 subjects with pose, illumination and expression (PIE) variations. In our experiments, only frontal face images under different lighting

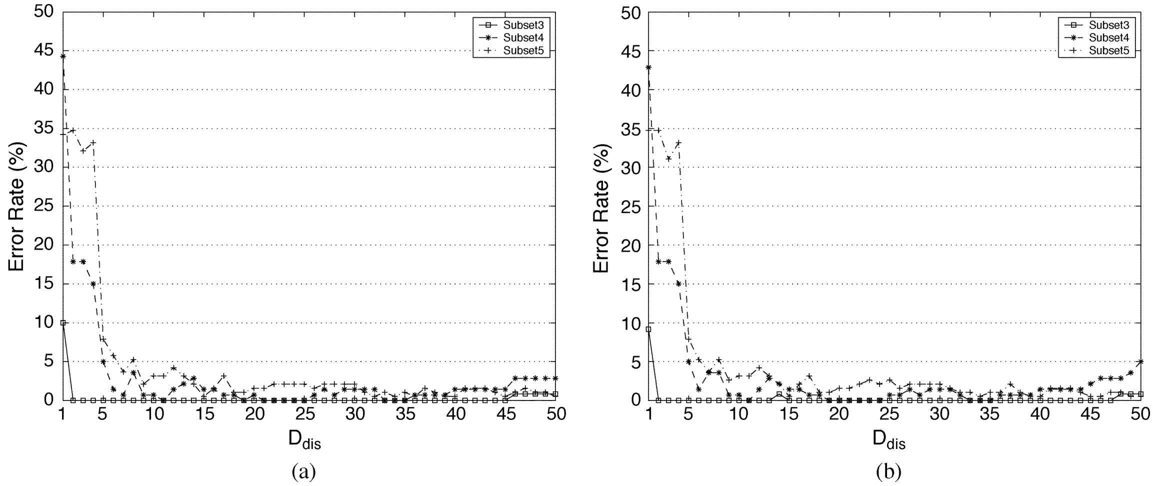


Fig. 7. Performance on the Yale B database with different D_{dis}. (a) Correlation. (b) Eigenfaces.

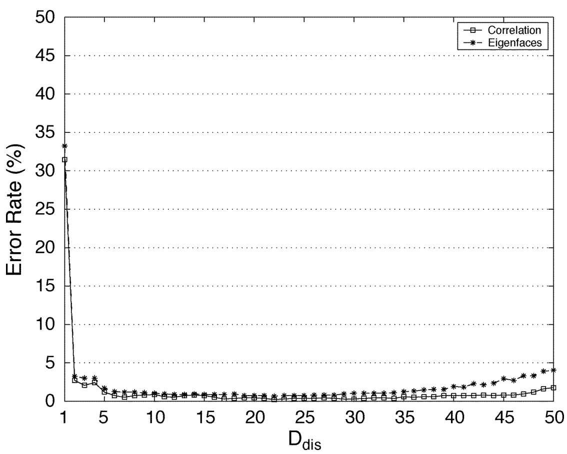


Fig. 8. Performance on the CMU PIE database with different D_{dis}.

TABLE II
RECOGNITION PERFORMANCE COMPARISONS OF DIFFERENT METHODS

Methods	Error rate (%)			
	Yale B			CMU PIE
	Subset 3	Subset 4	Subset 5	
No normalization	10.8	51.4	77.4	43.0
Histogram equalization	9.2	54.2	41.1	47.8
Linear subspace [7]	0	15.0	n/a	n/a
Cones-attached [7]	0	8.6	n/a	n/a
Cones-cast [7]	0	0	n/a	n/a
Gradient angle [8]	0	1.4	n/a	n/a
Harmonic images [14]	0.3	3.1	n/a	n/a
Illumination ratio images [15]	3.3	18.6	n/a	n/a
Quotient illumination relighting [3]	0	9.4	17.5	n/a
9PL [16]	0	2.8	n/a	1.9
Our method	0	0.18	1.71	0.36

conditions (without expression variations) are selected. As shown in Fig. 5, the frontally lighted face images are chosen as the training set. The remaining 20 face images of each subject with different illumination variations are used for testing. Face images in the experiment are all cropped and aligned in the same way as images in the Yale B database.

B. Experimental Results

In the experiments, the nearest neighbor classifier based on the Euclidean distance is employed for classification. All the face images used in the experiments are normalized so that they have zero mean and unit variance.

An appropriate number of discarded DCT coefficients should be chosen in order to normalize the illumination well and not weaken important facial features. We employ the dimensionality of the discarded coefficients (D_{dis} shown in Fig. 3) to measure the extent of discarding coefficients. Fig. 6 shows examples of normalized logarithm images with different D_{dis}.

Results of employing correlation and Eigenface methods [22] on the normalized face images of both databases with different D_{dis} are shown in Figs. 7 and 8. For the Eigenface method, 50 principal components are used. It is evident that the error rate significantly decreases after a few DCT coefficients are discarded. As illustrated in Figs. 7 and 8, the best and stable performances approximately lie in the range of $18 \leq D_{\mathrm{dis}} \leq 25$. In other words, in this range, illumination variations are largely reduced while important facial features are preserved. For the Yale B database, the small number of subjects is one of the reasons that the performance does not drop even when D_{dis} is around 50 because the high-frequency features are enough to distinguish these few subjects. Another reason is that discarding low-frequency DCT coefficients keeps high-frequency features well. In fact, illumination variations and facial features are not perfectly separated with respect to frequency components. Some illumination variations, especially shadows and specularities, lie in the same frequency bands as some facial features do. As a consequence, in order to compensate for such illumination variations, some facial information, mainly low-frequency intensity variations of facial feature components, has to be sacrificed. Nevertheless, our experiments show that high performance can still be achieved without these features. The low-frequency features actually become less effective under large illumination variation conditions. From Figs. 7 and 8, we can see that the performance has significantly improved when $D_{\mathrm{dis}} > 5$. Therefore, for some applications without large illumination variations, especially shadowing, more low-frequency components can be preserved. It should be noted that our method is different from the DCT applied for dimensionality reduction in face recognition [20], in which only low-frequency coefficients are used as facial features. Our method should use higher frequency features in order to reduce the illumination variations. Nevertheless, if logarithm images are used for recognition based on their method,

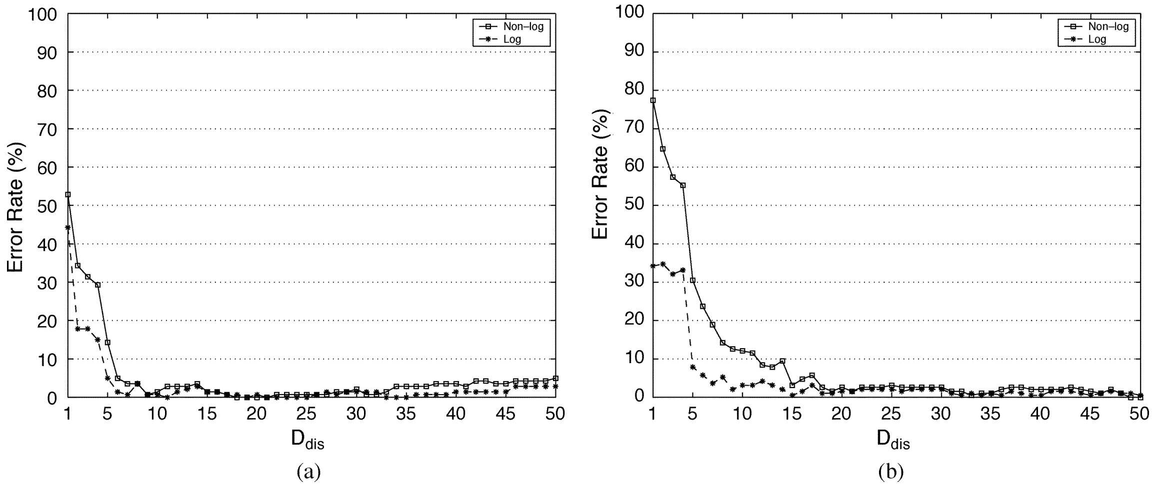


Fig. 9. Performance comparison between logarithm and nonlogarithm images (Yale B). (a) Subset 4. (b) Subset 5.

robustness against illumination variations can be easily improved by discarding several low-frequency DCT coefficients.

Comparison results with other methods dealing with illumination variations on both databases (mainly Yale B) are shown in Table II. The results of our normalization method are the average error rates of $18 \leq D_{\mathrm{dis}} \leq 25$. Some listed results of the existing methods are directly from other papers since they are based on the same database. We can see from Table II that the proposed method outperforms most of the existing methods except the cones-cast method. However, it should be pointed out that the illumination cone method needs much more complicated modeling steps, thus it cannot be applied in some practical applications. Moreover, in their paper, results on the most difficult subset (Subset 5) are not given.

C. Performance Comparison Between Logarithm and Non-Logarithm Images

As described in Section II-E, for face recognition, normalized logarithm face images should outperform nonlogarithm (i.e. with inverse logarithm transform) face images. Performance comparisons using the correlation method on the Yale B and CMU PIE databases are shown in Figs. 9 and 10, respectively. It is clear that better performance is achieved while using logarithm face images. It is more evident in the initial stage of discarding DCT coefficients for the reason that some higher frequency illumination variations are incorrectly estimated by using only a few low-frequency coefficients. As a consequence, logarithm face images should be used for recognition in the proposed approach.

D. Discarding DCT Coefficients Versus Discarding PCA Components

For the Eigenface method (PCA), it has been suggested that by discarding the three most significant principal components, variations due to lighting can be reduced. In [11], experimental results show that the Eigenface method performs better under variable lighting conditions after removing the first three principal components. However, the first several components not only correspond to illumination variations, but also some useful information for discrimination [11]. Besides, since the Eigenface method is highly dependent on the training samples, there is no guarantee that the first three principal components are mainly related to illumination variations. Fig. 11 shows the performance based on the Eigenface method by discarding different numbers of the first several principal components. It is evident that discarding first several principal components cannot improve the performance significantly.

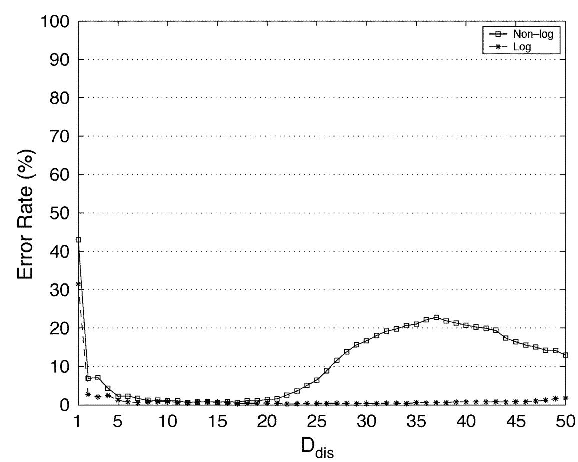


Fig. 10. Performance comparison between logarithm and nonlogarithm images (CMU PIE).

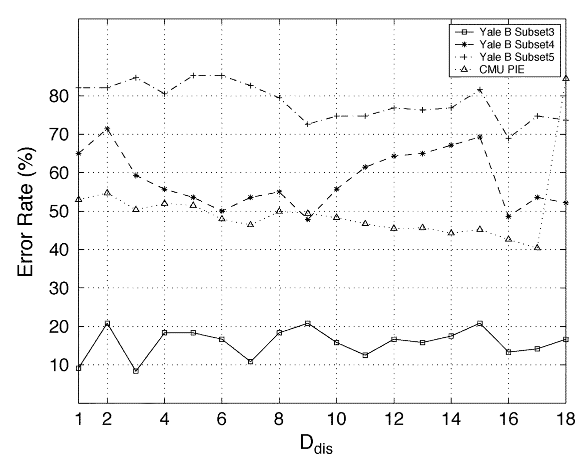
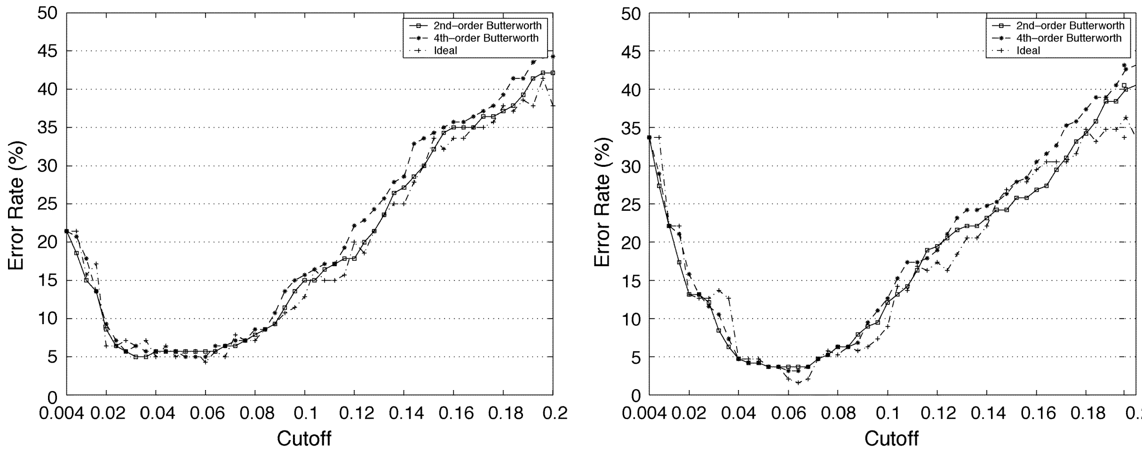


Fig. 11. Performance based on the Eigenface method by discarding different numbers of principal components (on original images, 50 principal components are used, i.e., the dimension of feature vectors is 50).

E. DCT Versus DFT

If the method of discarding DCT coefficients is regarded as a kind of filtering, DFT can also be employed since it is widely used as an image filtering method in the frequency domain. If DFT is employed instead



(a) (b)

Fig. 12. Performance on the Yale B database using high-pass filters with different transfer functions. (a) Subset 4. (b) Subset 5.

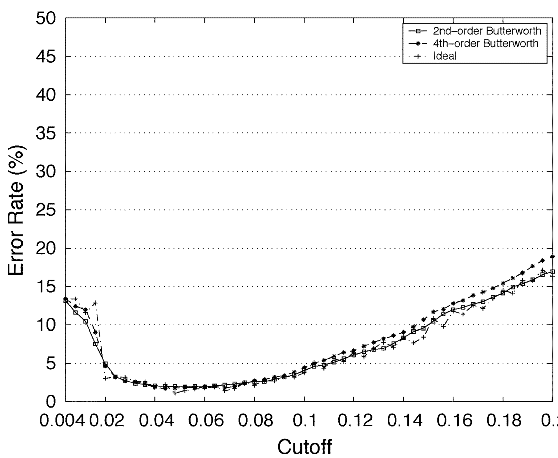


Fig. 13. Performance on the CMU PIE database using high-pass filters with different transfer functions.

of DCT, the proposed approach is similar to the so-called homomorphic filtering, an image enhancement approach which is used for contrast enhancement [21]. Homomorphic filtering can be accomplished by applying high-boost filtering or high-frequency emphasis filtering in the logarithm domain and subsequently transforming the filtered image to the original spatial domain using inverse logarithm transform. Different from the homomorphic filtering method for image enhancement, in illumination normalization applications, low-frequency illumination variations should be completely suppressed. Hence, high-pass filtering should be employed in the logarithm domain to reduce low-frequency illumination variations.

Normally, an ideal filtering method is not recommended for image filtering because unwanted ringing behavior will be generated [21]. However, in face recognition applications, we are more concerned with recognition performance rather than image quality. Fig. 12 shows the performances based on the correlation method of high-pass filters with different transfer functions, i.e., the second- and fourth-order Butterworth high-pass filters, and the ideal high-pass filter. As shown in Fig. 12 and Fig. 13, these transfer functions achieve similar performances. The only difference is that the Butterworth high-pass filter has a smoother performance curve with varying cutoff frequencies for the reason that, unlike the ideal filter, which has a sharp edge between passed and filtered frequencies, it employs a smooth transfer function. However, the best performance is achieved using the ideal high-pass filter. Moreover, the ideal filter is computationally less complex.

(a) (b) (c)

Fig. 14. (a) Second-order Butterworth filter. (b) Fourth-order Butterworth filter. (c) Ideal filter. $(\mathrm{cutoff} = 0.06)$.

Consequently, ideal filters can be used for illumination normalization for face recognition applications. Fig. 14 shows the filtered images using different transfer functions of high-pass filters.

The DCT is employed in this paper because it has a few advantages over the DFT: 1) the DCT is real-valued instead of complex (i.e., it involves magnitude and phase) such that it is easier to be implemented than the DFT; 2) when the DFT transform coefficients are truncated, the Gibbs phenomenon causes the boundary points to take on erroneous values [21]; this can be observed from Fig. 14; 3) the DCT is more efficient for illumination variation estimation than the DFT. This can be experimentally shown in the following part.

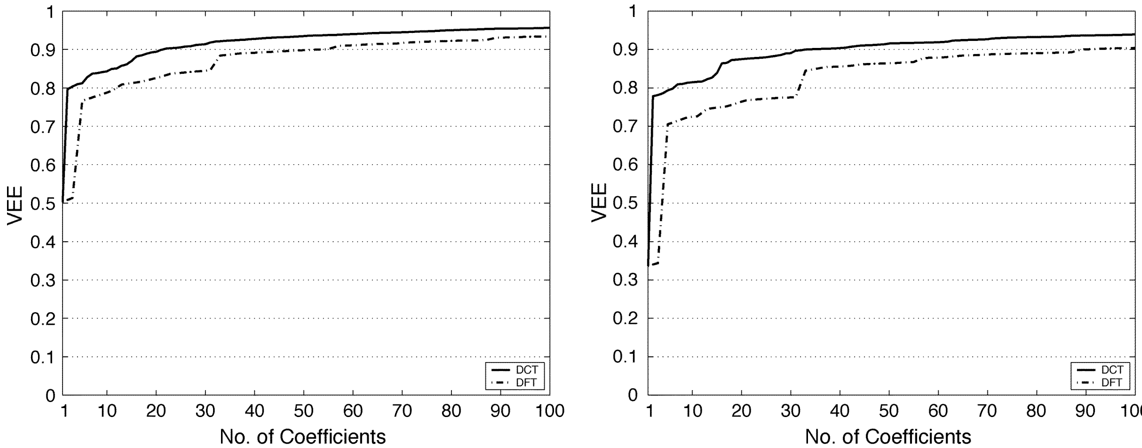
Similar to the definition of the energy packing efficiency (EPE) [24], we may define the variation estimation efficiency (VEE) as the performance criterion of variation estimation. For images of the same person, the VEE is the variance portion contained in the first M of N transform coefficients, i.e., for person P, and is given by

$$\mathrm{VEE}_P(M) = \frac{\sum_{i=0}^{M-1} E\left[(X_i - \bar{X}_i)^2\right]}{\sum_{i=0}^{N-1} E\left[(X_i - \bar{X}_i)^2\right]}. \tag{8}$$

The average VEE of the Yale B and CMU PIE database are respectively shown in Fig. 15(a) and (b). Obviously, the DCT has better VEE than the DFT, especially for the first few coefficients. In other words, DCT basis images are more correlated with illumination variations. We can also see from Table III that better recognition performance is achieved by using the DCT. (The results are the average error rates of the best performance range.)

F. Performance With Misaligned Face Images

The above experiments are all based on the well-aligned face images. Therefore, the higher frequency features are well utilized for recognition. In some practical applications, face images may not be aligned well. In this section, some experimental results on slightly misaligned



(a) (b)

Fig. 15. Average VEE of the Yale B and CMU PIE databases (Coefficients are sorted according to the DCT discarding and DFT filtering manner). (a) Yale B. (b) CMU PIE.

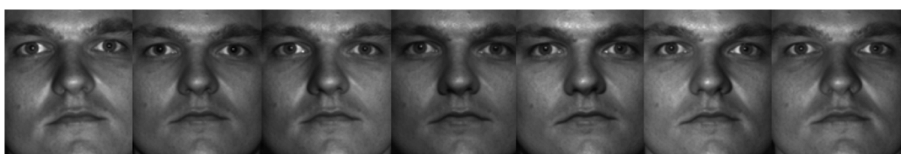
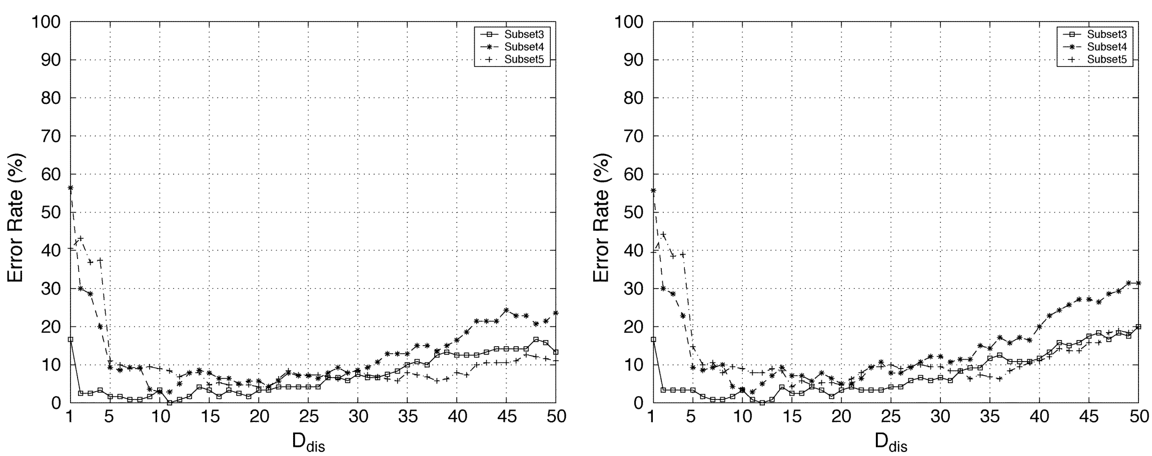


Fig. 16. Examples of misaligned face images from Yale B Subset 1.



(a) (b)

Fig. 17. Performance on the misaligned Yale B database with different $D_{\rm dis}$. (a) Correlation. (b) Eigenfaces.

face images are presented. Since the face alignment is based on eye coordinates, the misaligned face images are obtained by randomly adding offset errors to the eye coordinates such that there are small translation, scale and rotation variations in face images. Fig. 16 shows the slightly misaligned face images from Yale B Subset 1.

As shown in Fig. 17 and Fig. 18, the overall performance on misaligned images is worse than the well aligned images. This is the major drawback of appearance-based face recognition approaches. Besides, the performance degrades earlier in terms of $D_{\rm dis}$ because higher frequency features cannot be efficiently utilized. As a result, the value of $D_{\rm dis}$ should also be chosen taking into consideration accuracy of the alignment procedure. As aforementioned, discarding DCT coefficients is a tradeoff between low-frequency features and illumination variations. The proper $D_{\rm dis}$ should be chosen to minimize illumination variations as well as to keep low-frequency information as much as possible. Moreover, it also depends on how efficient the feature extraction method could utilize higher-frequency features that are essential for precise face recognition especially under large illumination variations. Nevertheless, the experimental results on these two databases show that recognition performance can be significantly improved when $D_{\rm dis} > 5$ even when the face images cannot be well aligned.

TABLE III
RECOGNITION PERFORMANCE COMPARISON BETWEEN DCT AND DFT (CORRELATION)

Methods	Error rate (%)			
	Yale B			CMU PIE
	Subset 3	Subset 4	Subset 5	
DCT ($18 \le D_{dis} \le 25$)	0	0.18	1.71	0.36
DFT ($0.052 \le Cutoff \le 0.068$)	0.83	5.29	2.63	1.62

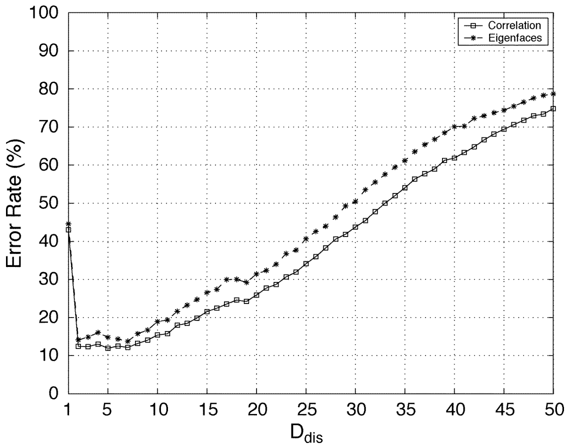


Fig. 18. Performance on the misaligned CMU PIE database with different D_{dis}.

IV. Conclusions

A novel illumination normalization approach is proposed in this paper. Illumination variations under different lighting conditions can be significantly reduced by discarding low-frequency DCT coefficients in the logarithm domain. Our approach has several advantages: 1) no modeling step and bootstrap sets are required; 2) our approach is very fast and it can be easily implemented in a real-time face recognition system; and 3) the proposed approach outperforms most of existing approaches. Nevertheless, the shadowing and specularity problems are not perfectly solved because they lie in the same frequency band as some facial features. Our future work will focus on reducing illumination variations caused by shadows and specularities. Furthermore, higher frequency facial features are more difficult to extract while poses and expressions change. Currently, we are exploring an efficient feature extraction method to make good use of higher frequency facial features.

Acknowledgment

The authors would like to thank Yale University for the use of the Yale Face Database B and Dr. Athinodoros S. Georghiades for providing useful information on this database. The authors would also like to thank Dr. S. Baker for providing the CMU PIE database.

References

[1] R. Chellappa, C. L. Wilson, and S. Sirohey, "Human and machine recognition of faces: a survey," *Proc. IEEE*, vol. 83, no. 5, pp. 705–740, May 1995.

[2] S. M. Pizer and E. P. Amburn, "Adaptive histogram equalization and its variations," *Comput. Vis. Graph., Image Process.*, vol. 39, no. 3, pp. 355–368, 1987.

[3] S. Shan, W. Gao, B. Cao, and D. Zhao, "Illumination normalization for robust face recognition against varying lighting conditions," in *Proc. IEEE Workshop on AMFG*, 2003, pp. 157–164.

[4] M. Savvides and V. Kumar, "Illumination normalization using logarithm transforms for face authentication," in *Proc. IAPR AVBPA*, 2003, pp. 549–556.

[5] X. Xie and K.-M. L, "Face recognition under varying illumination based on a 2D face shape model," *Pattern Recognit.*, to be published.

[6] P. N. Belhumeur and D. J. Kriegman, "What is the set of images of an object under all possible illumination conditions," *Int. J. Comput. Vis.*, vol. 28, no. 3, pp. 245–260, Jul. 1998.

[7] A. S. Georghiades, P. N. Belhumeur, and D. W. Jacobs, "From few to many: illumination cone models for face recognition under variable lighting and pose," *IEEE Trans. Pattern Anal. Mach. Intel.*, vol. 23, no. 6, pp. 630–660, Jun. 2001.

[8] H. F. Chen, P. N. Belhumeur, and D. J. Kriegman, "In search of illumination invariants," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, vol. 1, 2000, pp. 13–15.

[9] Y. Adini, Y. Moses, and S. Ullman, "Face recognition: the problem of compensating for changes in illumination direction," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 19, no. 7, pp. 721–732, Jul. 1997.

[10] R. Basri and D. W. Jacobs, "Lambertian reflectance and linear subspaces," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 25, no. 2, pp. 218–233, Feb. 2003.

[11] P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman, "Eigenfaces versus Fisherfaces: recognition using class specific linear projection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 19, no. 7, pp. 711–720, Jul. 1997.

[12] A. Shashua and T. Riklin-Raviv, "The quotient image: class-based re-rendering and recognition with varing illuminations," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 23, no. 2, pp. 129–139, Feb. 2001.

[13] W. Zhao and R. Chellappa, "Illumination-insensitive face recognition using symmetric shape-from-shading," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, 2000, pp. 286–293.

[14] L. Zhang and D. Samaras, "Face recognition under variable lighting using harmonic image exemplars," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, vol. 1, 2003, pp. 19–25.

[15] J. Zhao, Y. Su, D. Wang, and S. Luo, "Illumination ratio image: synthesizing and recognition with varying illuminations," *Pattern Recognit. Lett.*, vol. 24, pp. 2703–2710, 2003.

[16] K.-C. Lee, J. Ho, and D. J. Kriegman, "Acquiring linear subspaces for face recognition under variable lighting," *IEEE Trans. Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 5, pp. 684–698, May 2005.

[17] E. H. Land and J. J. McCann, "Lightness and retinex theory," *J. Opt. Soc. Amer.*, vol. 61, pp. 1–11, 1971.

[18] S. K. Nayar and R. M. Bolle, "Reflectance based object recognition," *Int. J. Comput. Vis.*, vol. 17, no. 3, pp. 219–240, 1996.

[19] T. Sim, S. Baker, and M. Bsat, "The CMU pose, illumination, and expression database," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 25, no. 12, pp. 1615–1618, Dec. 2003.

[20] Z. M. Hafed and M. D. Levine, "Face recognition using the discrete cosine transform," *Int. J. Comput. Vis.*, vol. 43, no. 3, pp. 167–188, 2001.

[21] R. C. Gonzalez and R. E. Woods, *Digital Image Processing*. Reading, MA: Addison-Wesley, 1992.

[22] M. A. Turk and A. P. Pentland, "Eigenfaces for recognition," *J. Cog. Neurosci.*, vol. 3, pp. 71–86, 1991.

[23] B. K. P. Horn, *Robot Vision*. Cambridge, MA: MIT Press, 1986.

[24] K. R. Rao and P. Yip, *Discrete Cosine Transform: Algorithms, Advantages, Applications*. Boston, MA: Academic, 1990.

[25] W. Pennebaker and J. Mitchell, *JPEG Still Image Data Compression Standard*. New York: Van Nostrand Reinhold, 1993.