



Extraction of illumination invariant facial features from a single image using nonsubsampling contourlet transform

Xiaohua Xie^{a,c}, Jianhuang Lai^{b,c,*}, Wei-Shi Zheng^d

^a School of Mathematics & Computational Science, Sun Yat-sen University, China

^b School of Information Science and Technology, Sun Yat-sen University, China

^c Guangdong Province Key Laboratory of Information Security, China

^d Department of Computer Science, Queen Mary University of London, UK

ARTICLE INFO

Article history:

Received 6 August 2009

Received in revised form

25 April 2010

Accepted 20 June 2010

Keywords:

Face recognition

Illumination normalization

Nonsubsampling contourlet transform

Multiplicative noise

ABSTRACT

Face recognition under varying lighting conditions is challenging, especially for single image based recognition system. Extracting illumination invariant features is an effective approach to solve this problem. However, existing methods are hard to extract both multi-scale and multi-directivity geometrical structures at the same time, which is important for capturing the intrinsic features of a face image. In this paper, we propose to utilize the logarithmic nonsubsampling contourlet transform (LNSCT) to estimate the reflectance component from a single face image and refer it as the illumination invariant feature for face recognition, where NSCT is a fully shift-invariant, multi-scale, and multi-direction transform. LNSCT can extract strong edges, weak edges, and noise from a face image using NSCT in the logarithm domain. We analyze that in the logarithm domain the low-pass subband of a face image and the low frequency part of strong edges can be regarded as the illumination effects, while the weak edges and the high frequency part of strong edges can be considered as the reflectance component. Moreover, even though a face image is polluted by noise (in particular the multiplicative noise), the reflectance component can still be well estimated and meanwhile the noise is removed. The LNSCT can be applied flexibly as neither assumption on lighting condition nor information about 3D shape is required. Experimental results show the promising performance of LNSCT for face recognition on Extended Yale B and CMU-PIE databases.

© 2010 Elsevier Ltd. All rights reserved.

1. Introduction

Face recognition technologies have been widely applied in the areas of intelligent surveillance, identity authentication, human-computer interaction, and digital amusement. However, one of the main limitations in deploying face recognition systems for practical use is their relatively low performance due to illumination variations. The face recognition vendor test (FRVT) 2002 [1] and the FRVT 2006 [2] have both revealed that large variation in illumination would seriously affect face recognition algorithms. Many well-known face descriptors such as local binary patterns (LBP) [4] and Gabor wavelet [3] have been proved to be effective for face recognition under good illumination condition, but their performances would degrade for large variations in illumination. So face illumination normalization is

a central task in face recognition, and many algorithms have been developed to tackle this issue.

1.1. Related work and analysis

The approaches of solving illumination problem in face recognition can be generally summarized into three categories¹ [32]: preprocessing and normalization technique [35–37], face modeling [33,38–40], and invariant feature extraction [4,5,20,21,41]. Methods of preprocessing and normalization process face image using image processing techniques, such as histogram equalization (HE), to normalize face image such that it appears to be stable under different lighting conditions. These approaches are always easy to implement, but it is still hard to obtain notable improvement for recognition. The model-based approach attempts to construct a generative 3-D face model that can be used to render face images of different poses and under

* Correspondence author at: School of Information Science and Technology, Sun Yat-sen University, China. Tel./fax: +86 20 84110175.

E-mail addresses: syxuixh@gmail.com (X. Xie), stsljh@mail.sysu.edu.cn (J. Lai), wszheng@ieee.org (W.-S. Zheng).

¹ We would like to point out that by using special hardware, one can obtain (near) infrared images invariant to visible light. Discussing this approach is beyond the scope of this paper.

varying lighting conditions. In these methods, a number of training samples are required and many assumptions are always made. Compared with the other two approaches, extracting illumination invariant features is a more effective approach for face recognition under various lighting conditions. Representative methods include local binary patterns (LBP) [4], Gabor feature [5], self quotient image (SQI) [41], logarithmic total variation (LTV) [20], and logarithmic wavelet transform (LWT) [21]. Recently, with the similar kernels of Gabor wavelet, the dual-tree complex wavelet transform (DT-CWT) [22] is used for face representation [43]. DT-CWT is good at capturing directional selective features in six different fixed orientations at dyadic scales and outperforms Gabor due to less redundancy and more efficient computation. Phase congruency [23] is also a kind of image feature insensitive to the variations in illumination and contrast. In order to end up with a reasonably small-size feature vector, a face recognition method based on the combination of the phase congruency and the Gabor wavelets has been proposed [45].

Based on the Lambertian theory, the intensity of a 2D surface I can be described as $I(x,y)=R(x,y)L(x,y)$ [6], where R and L are the reflectance component and illumination component, respectively. Reflectance component represents the intrinsic structures of the subject and can be regarded as a kind of illumination-invariance features. Because of this, the reflectance component is always extracted from a face image for recognition, and properly factorizing R and L from I is then a key research problem. Using multiple images of the same object, Weiss proposed a maximum-likelihood estimation method to obtain reflectance component [7]. However, for a single image, estimating R from I is an ill-posed problem [8]. To solve the problem, a common assumption is that L changes more slowly than R . Based on this assumption, various of low-pass filters such as homomorphic filtering [9] and Laplacian filtering [10] were developed to extract L , and R is obtained by $R=I/L$ or directly by high-pass filtering. Land and McCann [11] proposed Retinex model to estimate R as the ratio of the image I to the low-pass estimator L . However, these methods would all create halo artifacts under natural lighting condition. To reduce the halo artifacts, Jobson et al. [12] proposed to combine several low-pass copies as the estimation of L . Discontinuity preserving filtering is also used to estimate L , such as anisotropic diffusion [13,14,15], bilateral filtering [16,17], and mean shift filtering [18]. After initializing L by low-pass filtering, Chen et al. proposed the intrinsic illumination subspace based method to improve the result of estimation [19]. Chen et al. [20] proposed the logarithmic total variation (LTV) model to factorize face image. This method has several advantages compared to existing methods, particularly the capability of edge preserving and simple parameter selection.

The above filters lack the capabilities of multi-scale analysis. This limits their abilities in capturing multi-scale structures of an image. In order to address this problem, Zhang et al. [21] proposed using the wavelet analysis in the logarithmic domain for image decomposition. We call this method as LWT. As shown in [21], for face recognition the LWT outperform the LTV model. This may be due to the multi-scale capability of wavelet transform and therefore better edge-preserving ability is obtained in the low frequency illumination fields. However, since 2-D natural image is not the simple stack of 1-D piecewise smooth scan-lines, 2-D separable wavelet, as a simple tensor of 1-D wavelet, has weak directionality and is hard to effectually capture the geometrical structures of a face image due to its isotropy. In order to tackle this problem, a multi-scale and multi-directivity transform has to be considered. Nonsampled contourlet transform (NSCT) [27], which is a fully shift-invariant, multi-scale, and multi-direction transform, is therefore introduced and formulated for extracting more effective illumination invariant facial features.

We note that curvelet and contourlet have been used for face image recognition in [28–30]. These methods decompose face image into different frequency subbands and then directly use the coefficients of subbands as the facial features for recognition. In fact, they perform linear matrix decomposition in the original image domain rather than logarithmic domain. Direct linear decomposition in the image domain, however, cannot extract the intrinsic illumination invariant features very well, as a face image is nonlinearly represented by the reflectance and illumination components according to the Lambertian (reflectance) model, i.e. $I(x,y)=R(x,y)L(x,y)$, which is a multiplicative model. Furthermore, these methods may not extract the effectual discriminative features as the decomposed coefficients contain much redundant information.

1.2. Contributions

In a natural image, discontinuity points are typically located along contours owing to smooth boundaries of physical objects [24]. So an effective representation of image should have the following two abilities:

- (1) Isolating the discontinuities at edge points.
- (2) Exploring the smoothness along the contours.

Using square-shaped bases, 2-D wavelet will not achieve the second point. Unlike wavelet, the contourlet transform [26] represents image using the bases with different elongated shapes and in a variety of directions following the contour. These behaviors ensure that the contourlet transform has the above two abilities and captures more directional information than wavelet. As an advanced development of contourlet transform, the nonsampled contourlet transform (NSCT) [27] allows redundancy and can represent image more completely. Therefore, NSCT can efficiently capture the geometrical structures such as contours in a natural image and can accurately approximate the low frequency illumination fields. More importantly, strong edges, weak edges, and noises can be distinguished by NSCT. We will analysis that in the logarithm domain the low-pass subband of face image and the low frequency part of strong edges can be regarded as the illumination effect, while weak edges and the high frequency part of strong edges can be considered as the reflectance component. Accordingly, the logarithmic nonsampled contourlet transform (LNSCT), is proposed in this paper to estimate the reflectance component from a single face for face recognition. It is interesting to note that facial structures and noises (in particular multiplicative noises) can be distinguished by LNSCT. Therefore, the proposed methodology can estimate the reflectance component effectively even though the input image has been blurred by noise. In the rest of paper, we first briefly introduce NSCT (Section 2) before detailing LNSCT (Section 3). Our experimental results (Section 4) then show the significant improvement by using LNSCT for single image based face recognition.

2. Preliminary: NSCT

Contourlet transform [26] is a new extension of wavelet transform in the two-dimension case using non-separable and directional filter banks. With a rich set of basis images oriented at varying directions in multiple scales, contourlet transform can effectively capture the smooth contours merely using a small number of basis. Accordingly a better approximation of image can be obtained by contourlet transform as compared to wavelet transform [24].

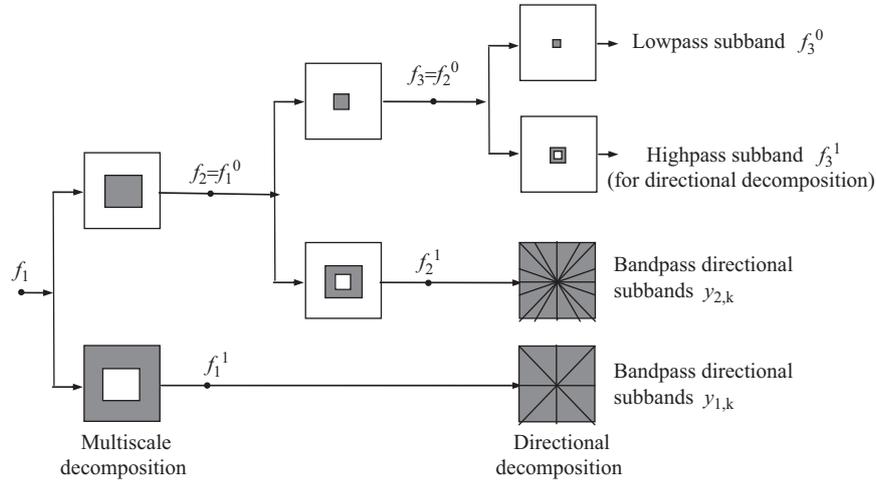


Fig. 1. Example of NSCT with three-stage pyramid decomposition.

The nonsampled contourlet transform (NSCT) [27], which allows redundancy, is a new development of contourlet transform. Allowing redundancy would make NSCT represent images more flexibly and more completely. Fig. 1 illustrates an overview of NSCT. NSCT is implemented by nonsampled filter bank structures. More specifically, it is constructed by combining the nonsampled pyramid (NSP) structure that ensures the multi-scale property and the nonsampled directional filter bank (NSDFB) structure that gives varying directions. Denote f_j as the input signal in the j -th level ($1 \leq j \leq J$). NSP first splits f_j into a low-pass subband f_j^0 and a high-pass subband f_j^1 using low-pass filter h_0 and high-pass filter h_1 :

$$f_j^i = h_i * f_j, \quad i = 0, 1 \quad (1)$$

where $*$ is convolution operator. Specially, the convolution formula is

$$f_j^i[n] = \sum_{k \in \text{supp}(h_i)} h_i[k] f_j[n - k \cdot S], \quad i = 0, 1, \quad n \in N \times N \quad (2)$$

where $S = 2^{j-1}I$ is the sampling matrix, I is the identity matrix, and $\text{supp}(h_i)$ is the compactly supported function of h_i . Then the high-pass subband f_j^1 is decomposed into several directional subbands by NSDFB and f_j^0 is for the next-stage decomposition. The NSDFB is constructed in cascade by combining two-channel fan filter banks and parallelogram filters without downsamplers and upsamplers. Consequently, the number of the directional subbands at a specific level is a power of two. We denote the equivalent filter for the k -th direction as u_k^{eq} , then the directional subbands can be obtained by

$$y_{j,k} = u_k^{eq} * f_j^1, \quad k = 1, \dots, 2^j \quad (3)$$

where 2^j is the number of directional subbands at the j -th level. This procedure would repeat on the low-pass subband by setting $f_{j+1} = f_j^0$ for the next level decomposition and the final low-pass subband is f_1^0 , so that directional subbands of different levels are generated. For the next level, all filters of pyramid are upsampled by 2 in both dimensions and this operation has been implied in the Eq. (2). It should be noted that filtering with the upsampled filters does not increase computational complexity. In our paper, the ‘maxflat’ filters and the ‘dmaxflat7’ filters are, respectively, selected for NSP and NSDFB. The details of design for filters are referred to [27].

The reconstruction of NSCT (invert NSCT) is also based on filtering operation according to the invert procedure of decomposition. Assume g_0 and g_1 are the corresponding synthesis filters of h_0 and h_1 , respectively, and v_k^{eq} is the synthesis filter of

u_k^{eq} . Then reconstruction of NSCT can be described as follows:

$$\begin{aligned} \hat{f}_j^0 &= \hat{f}_{j+1} \\ \hat{f}_j^1 &= \sum_{k=1}^{2^j} v_k^{eq} * y_{j,k}, \quad j = 1, \dots, J \\ \hat{f}_j &= g_0 * \hat{f}_j^0 + g_1 * \hat{f}_j^1 \end{aligned} \quad (4)$$

Given directional subbands $\{y_{j,k}\}_{j,k}$ and the low-pass subband f_j^0 , by setting $\hat{f}_{j+1} = f_j^0$ and iterating the procedure in Eq. (4) from the J -th level to the first level, the input signal can be reconstructed by $\hat{f} = \hat{f}_1$.

NSCT differs from other multi-scale analysis methods in that the contourlet transform allows for different and flexible number of directions at each scale. According to the direction information, directional filter bank can concatenate the neighboring singular points into (local) contours in the frequency domain, and therefore the detection of contours is obtained. By combination of NSP and NSDFB, NSCT is constructed as a fully shift-invariant, multi-scale, and multi-direction decomposition. It is worth to note that shift-invariant is very important. Lacking shift-invariance, pseudo-Gibbs phenomena may appear in image filtering [25]. Due to these merits, NSCT could highly benefit for image denoising and enhancement. Comprehensive comparison in Ref. [27] has demonstrated that NSCT performs better than wavelet and contourlet transform for image denoising. In this paper, we investigate the effect of NSCT for image decomposition and formulate the LNSCT based illumination invariant facial features extraction. In the following paragraphs, we will denote the decomposition procedure of NSCT as

$$\{(y_{j,k})_{j,k}, f_j^0\} = \text{NSCT}(f) \quad (5)$$

where f is the input signal, $\{y_{j,k}\}_{j,k}$ are directional subbands coefficients and f_j^0 is low-pass subband. On the other hand, the reconstruction of NSCT is denoted as

$$\hat{f} = \text{iNSCT}(\{(y_{j,k})_{j,k}, f_j^0\}) \quad (6)$$

3. Methodology

As a fully shift-invariant, multi-scale, and multi-direction transform, NSCT is suitable for preserving geometrical structures in natural scenes and is therefore an excellent technique for image

analysis. Specially, strong edges, weak edges, and noise can be distinguished in the NSCT domain. By extraction and recombination of these components in the logarithm domain, illumination component and reflectance component of face image can be effectually estimated. After that the reflectance component is considered as the illumination invariant features for face recognition.

3.1. LNSCT

According to the Lambertian reflectance function, the intensity of a face surface I can be represented as

$$I = R \odot L \quad (7)$$

where \odot indicates element-by-element multiplication, R stands for view independent reflectance component, which is a kind of intrinsic feature of object, determined only by the surface material, and L is the shading of a Lambertian surface which is the final light received at a certain location. In the reflectance model, neither assumption on the number of light sources nor 3D shape information is required. In the real life, the process of capturing images of objects and scenes, for instance vague lens, and inrush current of camera, usually introduces noises including additive noise and/or multiplicative noise into the images. When the input image has been polluted by noise, reflectance component may not be properly estimated by the filtering based technologies, since noise and reflectance component both produce low-magnitude coefficients in the frequency domain. However, if the image is blurred by multiplicative noise, the reflectance component can still be well estimated by our LNSCT model. Multiplicative noise is a type of signal-dependent noise, which may be generated by imaging device and is difficult to remove without impairing image details. Considering the multiplicative noise, a blurred face image I can be represented as

$$I = R \odot L \odot N \quad (8)$$

where N is the multiplicative noise. For illumination invariant face recognition, the problem here is how to factorize the reflectance component R from a given face surface I . We next investigate how

NSCT is used to estimate R in our LNSCT model. Generally, NSCT decomposes a signal into different frequency subbands in additive form. However the right hand side of Eq. (8) is multiplicative, we cannot apply NSCT directly on the original image to get the illumination invariant components. Fig. 2 illustrates the decomposition on the original face image by NSCT. It shows that the illumination has effects on both the high-frequency subband and low-frequency subband. In order to address this problem, we therefore consider the factorization in the logarithm domain. Taking logarithm transform on Eq. (8) yields

$$f = \log I = \log R + \log L + \log N \triangleq v + u + \eta \quad (9)$$

The logarithm transform turns the multiplicative model into the additive model. Its advantages are mainly two-fold: first, as an additive model, classical image estimation techniques can be applied; second, the logarithm transform of the luminance (similar to the logarithm of intensity) is a crude approximation to the perceived brightness, so logarithm transform can partially reduce the lighting effect. Furthermore, since the logarithm preserves structures, v , u , and η keep the similar characters as R , L , and N , respectively.

For an image, there are three categories of pixels: strong edges, weak edges, and noise, which can be classified by analyzing the distributions of their coefficients in the NSCT subbands [27]. The strong edges correspond to those pixels with large magnitude coefficients in all subbands. The weak edges correspond to those pixels with large magnitude coefficients in some directional subbands and small magnitude coefficients in the other directional subbands at the same scale. The noise corresponds to those pixels with small magnitude coefficients in all subbands. In the logarithm domain, these components still can be effectually separated. Fig. 3 illustrates the components analysis of face image in the logarithm domain by NSCT. It is interested to note that in the logarithm domain the illumination effects are just on the low-pass subband and the strong edges. Furthermore, noise can be distinguished from facial structures. This highly supports the feasibility of using LNSCT for the extraction of illumination invariant facial features.

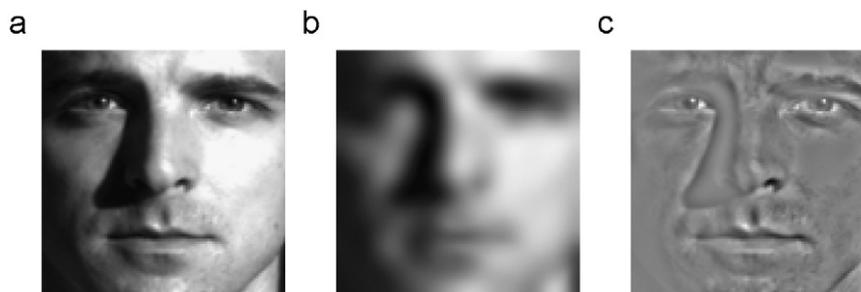


Fig. 2. Face image decomposition using direct NSCT. (a) Original face image, (b) low frequency part, and (c) high frequency part.

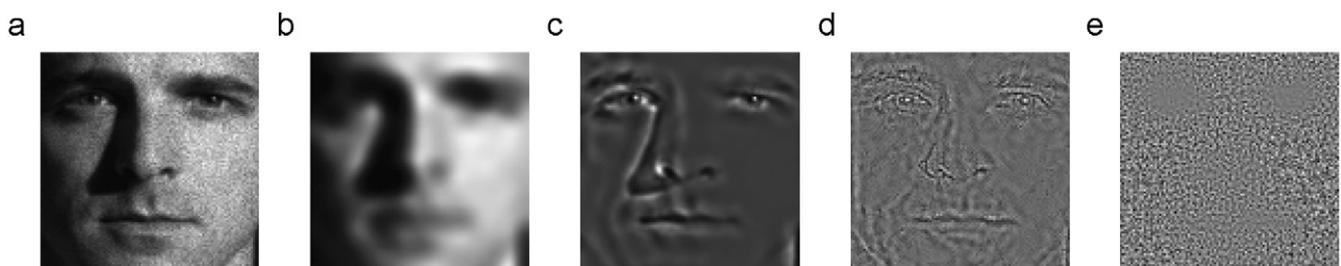


Fig. 3. Components analysis of face image using NSCT in the logarithm domain. (a) Noise-blurred image, (b) low-pass subband, (c) strong edges, (d) weak edges, and (e) noise.

On one hand, since facial small structures (e.g. eyes, nose, mouths, and eyebrow) are mostly composed of lines, edges, and small-scale objects [20], they contain weak edges. On the other hand, the illumination pattern is composed of direct light illumination and/or shadows cast by bigger objects, and is often of large scale. This gives sufficient reason to regard the low-pass subband belonging to illumination effect. The strong edges, however, cannot be simply treated as facial structures or illumination effect, since according to our observation they contain mostly facial structures as well as some illumination effect especially shadows. So we need to conduct a threshold operation to distinguish the facial structures and illumination effect in strong edges, where the low frequency part is treated as illumination effect and the high frequency part is regarded as facial structures. Fig. 4 illustrates the decomposition of strong edges. It shows that unlike the high frequency part, the low frequency part of strong edges hardly represents clear facial structures. It should be pointed out that by threshold operation, sometimes few facial large structures may be mistakenly classified as illumination effect, but this would scarcely affect face recognition process.

According to the above analysis, in the logarithm domain the low-pass subband of face image and the low frequency part of strong edges can be regarded as the illumination component, while weak edges and the high frequency part of strong edges can be considered as the reflectance component. These suggest the scheme of using LNSCT for extraction of facial features, which includes decomposition and recombination of image structures. Fig. 5 gives an example of the face decomposition, and the details are described below.

Denote the decomposition of f by NSCT as

$$(\{y_{j,k}\}_{j,k} f_j^0) = NSCT(f) \tag{10}$$

where f_j^0 is the low-pass subband and $y_{j,k}$ ($j=1, \dots, J, k=1, \dots, 2^l$) are bandpass directional subband coefficients at the k -th direction of the j -th scale. In order to estimate the reflectance component, a high-pass thresholding operation thr_T^ν is imposed on $y_{j,k}$ to obtain new directional subband coefficients $y'_{j,k}(n) = thr_T^\nu(y_{j,k}(n))$,

$n \in N \times N$. Note that using soft thresholding can avoid abrupt artifacts and obtain more pleasant visual results as compared to hard thresholding, so we use the soft high-pass thresholding operation thr_T^ν as follows:

$$thr_T^\nu(x) = \begin{cases} T, & \text{if } x \geq T \\ -T, & \text{if } x \leq -T \\ 0, & \text{if } MAX < c\sigma \\ x, & \text{otherwise} \end{cases} \tag{11}$$

where T is the threshold, σ is the noise standard deviation of the subbands at a specific pyramidal level, c is a parameter ranging from 0.01 to 0.1 subject to $c\sigma < T$, and MAX is the maximum magnitude of the coefficients for each pixel across directional subbands at the same scale, that is

$$MAX = \max_k \{|y_{j,k}(n)|\} \tag{12}$$

Specially $MAX < c\sigma$ means that the corresponding pixel produces small magnitude coefficients in all subbands of the same scale. In this case, the same spatial location of the original image can be thought to be polluted by noise and the coefficient should be set to zero. Since the low-pass subband belongs to the illumination component, it should be filtered out by setting $f_j^0(n) = 0$ for $\forall n$. Finally the logarithm of reflectance component ν is estimated by reconstructing the signal from the modified NSCT coefficients:

$$\hat{\nu} = iNSCT(\{y'_{j,k}\}_{j,k} f_j^0) \tag{13}$$

where $iNSCT$ is the inverse NSCT. Accordingly, the logarithm noise can be extracted by

$$\hat{\eta} = iNSCT(\{thr_T^\eta(y_{j,k})\}_{j,k} f_j^0) \tag{14}$$

where $f_j^0(n)$ is also assigned by $f_j^0(n) = 0$ for $\forall n$ and thr_T^η is defined as

$$thr_T^\eta(x) = \begin{cases} x, & \text{if } MAX < c\sigma \\ 0, & \text{otherwise} \end{cases} \tag{15}$$

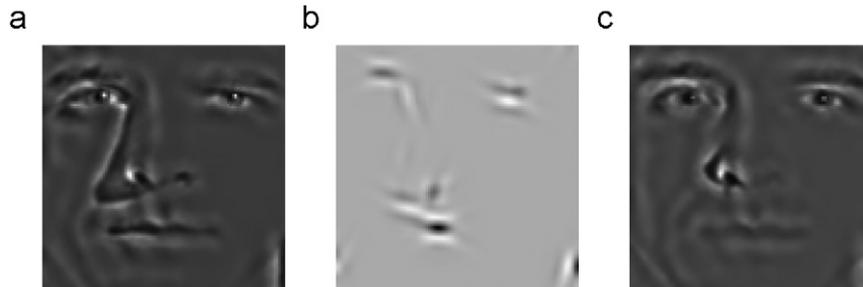


Fig. 4. Decomposition of strong edges. (a) Strong edges, (b) low frequency part, and (c) high frequency part.

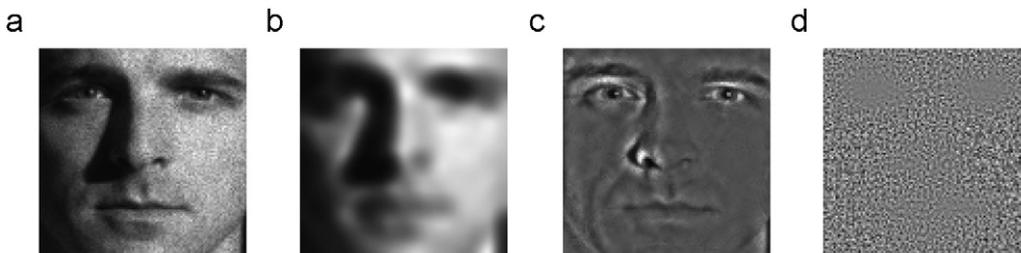


Fig. 5. Decomposition of face image using LNSCT. (a) Noise-blurred image, (b) illumination effect, (c) reflectance component, and (d) noise.

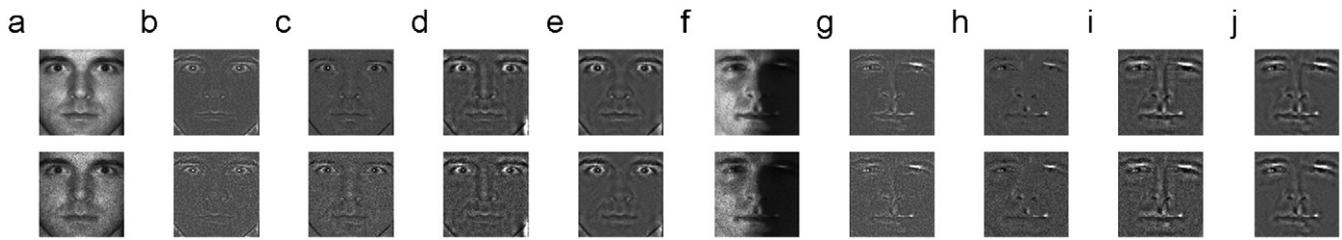


Fig. 6. Reflectance components estimated from noise-blurred images: (a) and (f) are noise-blurred images, (b) and (g) are the estimated results by SQL, (c) and (h) are the estimated results by LTV, (d) and (i) are the estimated results by LWT, (e) and (j) are the estimated results by proposed LNSCT. Different rows correspond to the noises of different standard deviations ($\sigma=0.1$ and 0.2 , respectively).

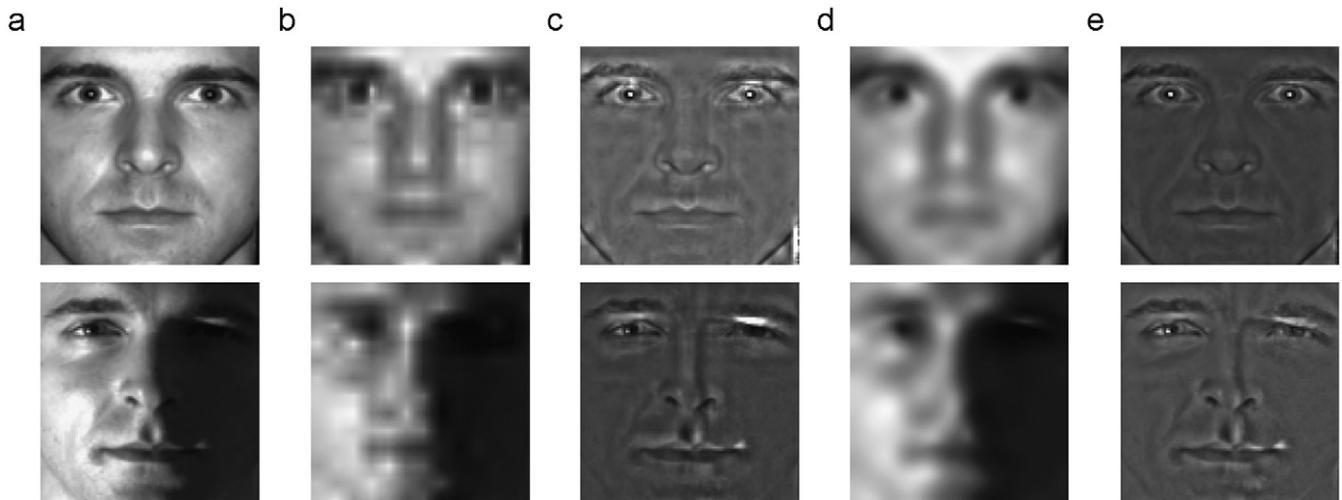


Fig. 7. Decomposition of noise-free face images using different methods. (a) are original images, (b) and (c) are the decomposition results using LWT, (d) and (e) are the decomposition results using LNSCT. (b) and (d) are the estimated illumination components, (c) and (e) are the estimated reflectance components.

Finally, the noise N , the reflectance component R , and the illumination component L can be respectively obtained by

$$\hat{R} = \exp(\hat{\nu}), \quad \hat{N} = \exp(\hat{\eta}), \quad \hat{L} = \exp(f - \hat{\nu} - \hat{\eta}) \quad (16)$$

Fig. 6 displays the reflectance components estimated by different algorithms from the blurred images with various noises. It is shown that the proposed LNSCT attains more robust results than other methods. This is because that the decompositions such as total variation (TV) model and Gaussian filtering used in other methods are just for separating the high- and low-frequency components of an image, and cannot distinguish the noise and edges.

Specially, in the noise-free case, we do not need to distinguish facial structures from noise but to separate high- and low frequency components. Accordingly, the Eq. (9) is reduced to

$$f = v + u \quad (17)$$

And the Eq. (11) can be simplified as

$$thr_T^v(x) = \begin{cases} T, & x \geq T \\ -T, & x \leq -T \\ x, & \text{otherwise} \end{cases} \quad (18)$$

Consequently, the illumination component L can be obtained by

$$\hat{L} = \exp(f - \hat{\nu}) \quad (19)$$

Fig. 7 illustrates the comparison of decomposition on noise-free images using LWT and LNSCT. As shown, due to the lack of multi-directivity, the blocking effects appear seriously in the reflectance component as well as the illumination component estimated by LWT. In contrast, the proposed LNSCT gets much properly smoother results.

Since the reflectance component is considered as the intrinsic facial features, it can be directly used for face recognition. Restoring the logarithmic image to the original one may enlarge the possible inaccuracy of estimation [32]. Accordingly, in our approach, for a grayscale image I , the logarithm of reflectance component, ν , is directly used as the facial features for face recognition. Summarily, the proposed algorithm is illustrated in Table 1.

3.2. Parameter selection

In order to implement the proposed method, several parameters, namely the number of scale, the number of directions at each scale for NSCT and the threshold T need to be set. For NSCT, the relevant parameters are set empirically. In general, the larger the size of an image is, the richer the scales are required. For example, it needs at least three scales for the image of size 100×100 , and no less than four scales for the image of size 256×256 . Furthermore, the number of directions in each level should be large than eight. Figs. 8 and 9 illustrate some estimated results using different NSCT parameters. In our experiments, 8, 16, 16 directions are selected in three scales from coarse to fine for the image of size 100×100 . Regarding the threshold T , it actually

indirectly depends on the face image itself. The analysis in Ref. [21] finds that the key facial features (reflectance component) in logarithm domain can be treated as “noise” in the denoising model. Accordingly we can use the methods for denoising to select the shrink T . Suggested by the authors of Ref. [27], we use

the BayesShrink [31] which is derived in a Bayesian framework and finds the soft-threshold that minimizes this Bayesian risk. It should be noted that we just use the similarity of noise and reflectance component to select model parameter, but the actual noise and reflectance component can still be distinguished in our method. Like [31], the threshold map $T_{j,k}$ is selected independently in each subband $y_{j,k}$. For the n -th coefficient, the corresponding threshold $T_{j,k,n}$ is calculated by

$$T_{j,k,n} = \frac{(\sigma_{j,k,n}^y)^2}{\hat{\sigma}_{j,k,n}^u}$$

$$\sigma_{j,k,n}^u = \sqrt{\max((\sigma_{j,k,n}^f)^2 - (\sigma_{j,k,n}^y)^2, 0)}$$

$$\sigma_{j,k,n}^f = \frac{1}{m^2} \sum_{y_{j,k}(n') \in N(y_{j,k}(n))} |y_{j,k}(n')|^2 \quad (20)$$

where $\sigma_{j,k,n}^u$ is the illumination component standard deviation, $(\sigma_{j,k,n}^y)^2$ is the reflectance component variance, and $\sigma_{j,k,n}^f$ is the observed signal (logarithmic face image) standard deviation of the n -th coefficient at the k -th directional subband of the j -th scale. The observed signal variances $\sigma_{j,k,n}^f$ are estimated locally using the neighboring coefficients contained in a $m \times m$ square window

Table 1
The LNSCT algorithm.

1. $I \leftarrow$ input image ($I=R \odot L \odot N$ or $I=R \odot L$)
2. Take the logarithm:
 $f = \log(I) = \log(R) + \log(L) + \log(N) \stackrel{\Delta}{=} v + u + \eta$ or $f = v + u$
3. Apply NSCT on f :
 $(\{y_{j,k}\}_{j,k} f_j^0) = NSCT(f)$
4. Impose a threshold operation on the NSCT coefficients: $y'_{j,k}(n) = thr_v^y(y_{j,k}(n))$
5. Set the low-pass subband $f_j^0(n) = 0$ for $\forall n$
6. Obtain the illumination invariant features by constructing the signal from the modified NSCT coefficients for face recognition:
 $\hat{v} = iNSCT(\{y'_{j,k}\}_{j,k} f_j^0)$

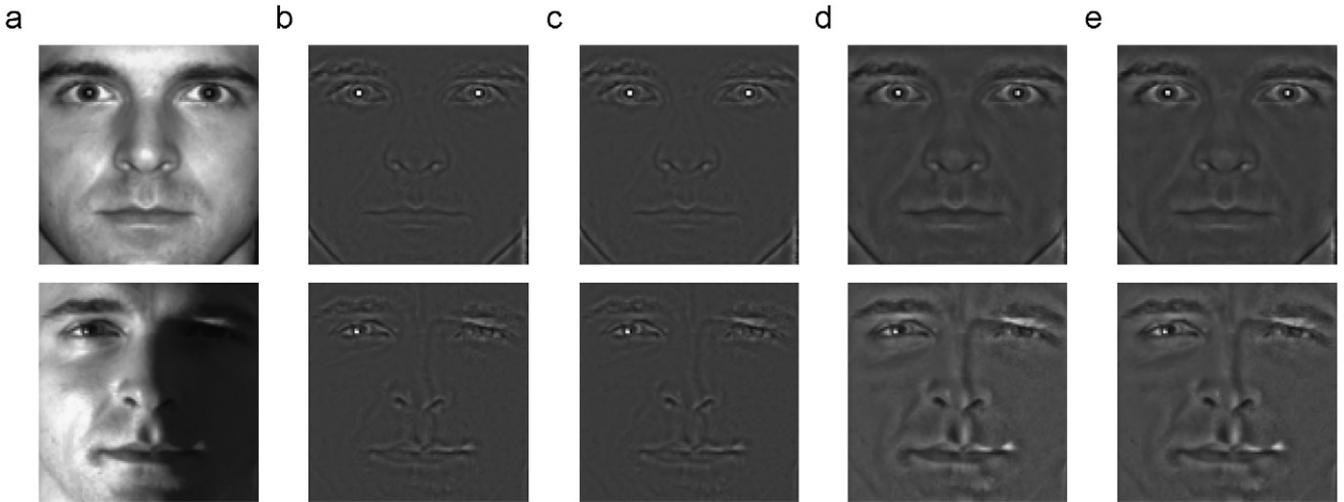


Fig. 8. Examples of reflectance component estimation from 100×100 images using different NSCT parameters. (a) are original images and (b)–(e) are the estimated results by using different numbers of directions in different scales. (b) 8 and 16 directions in two scales, (c) 16 and 8 directions in two scales; (d) 8, 16, and 16 directions in three scales, (e) 16, 8, and 16 directions in three scales from coarser to finer.

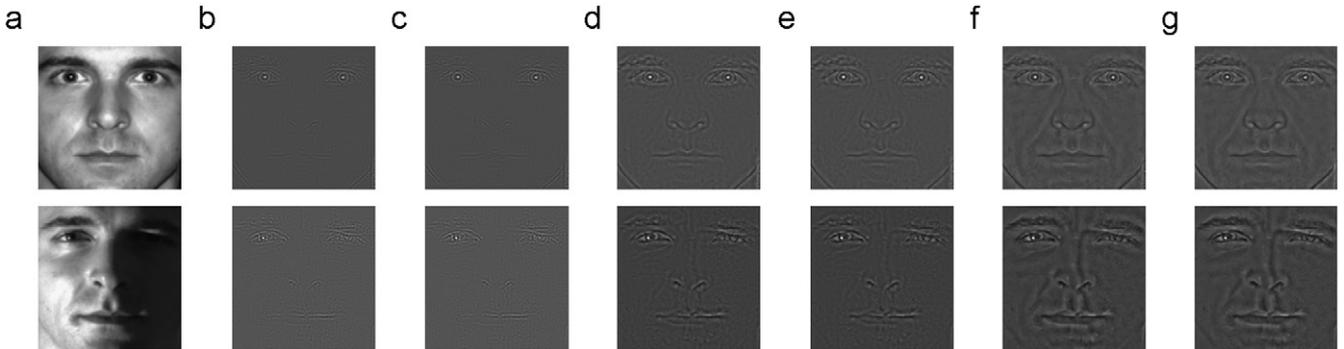


Fig. 9. Examples of reflectance component estimation from 256×256 images using different NSCT parameters. (a) are original images and (b)–(g) are the reflectance components estimated by using different numbers of directions in different scales. (b) 8 and 16 directions in two scales, (c) 16 and 8 directions in two scales, (d) 8, 16, and 16 directions in three scales, (e) 16, 8, and 16 directions in three scales, (f) 8, 16, 16, and 16 directions in four scales, (g) 4, 2, 8, and 16 directions in four scales from coarser to finer.

within each subband and m is generally set to be 5. For $\sigma_{j,k,n}^v$, it is inferred by using Monte Carlo technique [27], where the reflectance component variance in our model correspond to the noise variance in the denoising model. The variances in different subbands are computed for a few normalized noise images and the average value $(\sigma_{j,k,n}^M)^2$ is obtained. On the other hand, the reflectance variance is estimated using the median operator

$$\sigma = \frac{\text{median}(|y_{1,l_1}|)}{\lambda} \quad (21)$$

Finally the $(\sigma_{j,k,n}^v)^2$ is obtained by

$$(\sigma_{j,k,n}^v)^2 = \sigma^2 (\sigma_{j,k,n}^M)^2 \quad (22)$$

Different scale threshold λ would lead to different scale smoothness (see Fig. 10). Hence, the selection of λ is very important. According to our observation, λ should be selected from 0.001 to 0.01. In our experiments, λ is consistently set to be 0.003.

4. Experimental results and analysis

In our experiments, the proposed LNSCT has been evaluated on Extended YaleB [33] and CMU-PIE [34] databases. Face images from 38 individuals in Extended YaleB were captured under 64 different lighting conditions on 9 poses. In the experiment, only the 2432 ($=64 \times 38$) frontal face images from Extended YaleB database were used. All face images have been divided into five subsets according to the angle between the light source direction and the camera axis [33] as shown in Table 2. For CMU-PIE, the 1428 frontal face images from 68 individuals under 21 different illumination conditions with background lighting off were selected to use in the experiments. All images from the above face databases were simply aligned and resized to 100×100 .

In our method, for each face image I , the logarithm of reflectance component, v , is directly used as the facial features. The normalized correlation was selected as the similarity metric. The normalized correlation between two vectors, v_1 and v_2 , was calculated by

$$\delta(v_1, v_2) = \frac{v_1 \bullet v_2}{\|v_1\|_2 \|v_2\|_2} \quad (23)$$

where \bullet is the inner product and $\|\cdot\|_2$ is l_2 -norm. The higher the similarity $\delta(v_1, v_2)$ is, the closer the two samples v_1 and v_2 are.

In our experiment, we consider two kinds of evaluation, namely recognition and verification. In the recognition case, the nearest neighbor (NN) classifier was used. If there are in total M query images from all subjects, and M_1 images out of these M query images can be correctly recognized, then the recognition rate is M_1/M . For Extended YaleB, the recognition results on noise-free images will be reported on each subset. In the verification case, one frontal-illuminated image of each subject was registered as the reference image and the rest images were used as the probe ones. Note that for verification, we did not consider the partition of Extended YaleB anymore. Accordingly, for each subject from Extended YaleB database, there were 63 ($=64-1$) client samples and 2331 ($=63 \times 37$) impostor samples. So the number of genuine and impostor scores were 2394 ($=63 \times 38$) and 88578 ($=2331 \times 38$), respectively. For CMU, the number of genuine and impostor scores were 1360 ($=20 \times 68$) and 91120 ($=20 \times 67 \times 68$), respectively. We will report the ROC curve and equal error rate (EER) for each method to show the verification performance.

4.1. Face recognition on noise-free images

In this section, we report the performance of the LNSCT for face recognition on noise-free images. Accordingly the proposed model in Eq. (18) was used for evaluation here. It has been experimentally verified in Ref. [20] that the LTV model

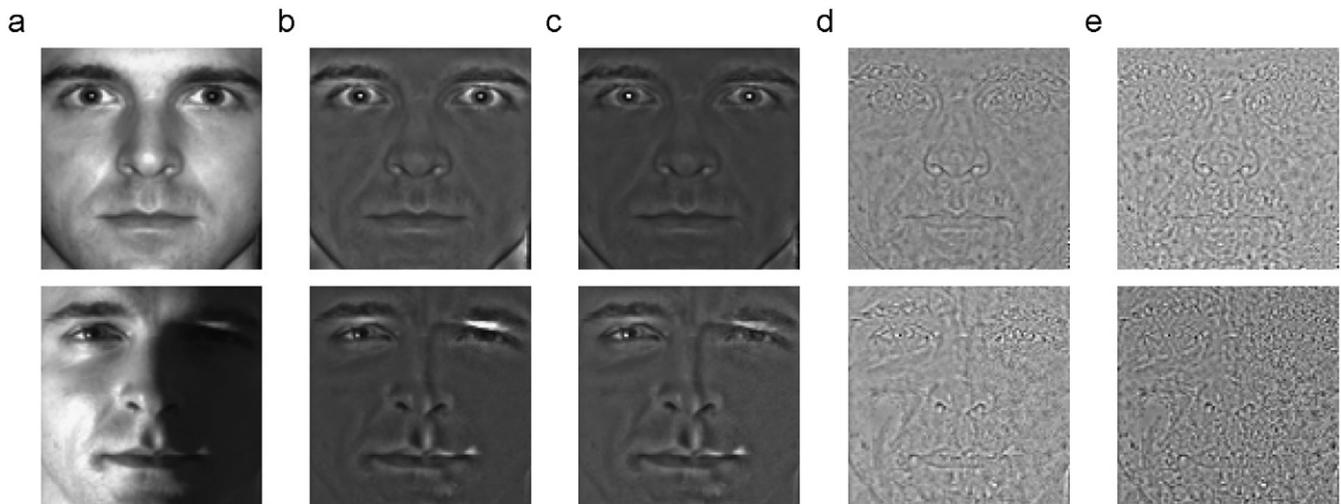


Fig. 10. Examples of reflectance component estimation with different λ under different illumination conditions. (a) Original images. (b) Estimated results with $\lambda=0.0005$. (c) Estimated results with $\lambda=0.003$. (d) Estimated results with $\lambda=0.05$. (e) Estimated results with $\lambda=0.1$.

Table 2
Subsets of Extended YaleB [33].

| Subsets | 1 | 2 | 3 | 4 | 5 |
|-----------------------|---------------------|----------------------|----------------------|----------------------|----------------------|
| Lighting angle (deg.) | 0–12 | 13–25 | 26–50 | 51–77 | > 77 |
| Number of images | $7 \times 38 = 266$ | $12 \times 38 = 456$ | $12 \times 38 = 456$ | $14 \times 38 = 532$ | $19 \times 38 = 722$ |

Table 3

Face recognition rates of different methods on noise-free images. The images under frontal illumination condition are used as the reference images.

| Database | Method | Recognition rate (%) | | | | |
|----------------|------------|----------------------|-----------------|----------------|----------------|----------------|
| CMU | LBP | 75.3676 | | | | |
| | Curvelet | 78.3824 | | | | |
| | Contourlet | 77.1324 | | | | |
| | DT-CWT | 89.5588 | | | | |
| | SQI | 99.0441 | | | | |
| | LTV | 99.7794 | | | | |
| | LWT | 99.5588 | | | | |
| | LNSCT | 99.9265 | | | | |
| | | Set 1 | Set 2 | Set 3 | Set 4 | Set 5 |
| Extended YaleB | LBP | 100.0000 | 100.0000 | 62.2807 | 10.3383 | 6.6482 |
| | Curvelet | 91.6667 | 100.0000 | 55.9211 | 14.8496 | 5.9557 |
| | Contourlet | 76.3158 | 98.6842 | 52.1930 | 22.3684 | 9.8338 |
| | DT-CWT | 98.6842 | 99.3421 | 76.7544 | 38.9098 | 13.8504 |
| | SQI | 100.0000 | 98.6842 | 71.2719 | 69.3609 | 63.9889 |
| | LTV | 100.0000 | 99.7807 | 78.5088 | 75.7519 | 82.4100 |
| | LWT | 100.0000 | 100.0000 | 82.0175 | 81.9549 | 70.7756 |
| | LNSCT | 100.0000 | 100.0000 | 83.3333 | 87.9699 | 84.3490 |

Table 4

Face recognition rates of different methods on noise-free images. The reference set (three images per individual) is randomly chosen.

| Database | Method | Recognition rate (%) | | | | |
|----------------|------------|----------------------|----------------|----------------|----------------|----------------|
| CMU | LBP | 81.8417 | | | | |
| | Curvelet | 92.8277 | | | | |
| | Contourlet | 89.1870 | | | | |
| | DT-CWT | 94.5770 | | | | |
| | SQI | 97.7760 | | | | |
| | LTV | 99.0100 | | | | |
| | LWT | 99.5020 | | | | |
| | LNSCT | 99.8760 | | | | |
| | | Set 1 | Set 2 | Set 3 | Set 4 | Set 5 |
| Extended YaleB | LBP | 81.5789 | 73.0789 | 64.4430 | 36.6880 | 31.7452 |
| | Curvelet | 79.7293 | 75.7763 | 63.0658 | 48.9699 | 29.7258 |
| | Contourlet | 76.8722 | 69.7018 | 68.9430 | 49.7331 | 35.8670 |
| | DT-CWT | 77.9474 | 76.7939 | 72.5570 | 51.8383 | 32.3407 |
| | SQI | 82.7440 | 85.1360 | 87.1060 | 81.1920 | 81.6140 |
| | LTV | 84.4820 | 80.8560 | 89.2940 | 85.4100 | 90.9700 |
| | LWT | 91.2560 | 86.5580 | 91.0840 | 93.2480 | 90.5920 |
| | LNSCT | 94.2940 | 88.9860 | 92.8120 | 94.4960 | 98.1960 |

outperform several popular algorithms for face recognition, such as quotient illumination relighting (QIR) [35], quotient image (QI) [40], and self quotient image (SQI) [41]. Thus, we mainly focus on the comparison between LNSCT and the state-of-the-art algorithms for face recognition under varying lighting conditions, including local binary patterns (LBP) [4], dual-tree complex wavelet transform (DT-CWT) [22], self quotient image (SQI) [41], logarithmic total variation (LTV) [20], and logarithmic wavelet transform (LWT) [21]. For LWT, three levels daubechies-3 wavelet was selected. For DT-CWT, the magnitudes of high-pass wavelet coefficients of four-level decomposition were used as facial features. We also report the results of similar techniques such as contourlet or curvelet approaches which directly use the high-pass coefficients of decomposition for face recognition. For contourlet, the numbers of scales and directions were set the same as LNSCT. As contourlet has special requirement on the resolution of image, we have to resize the face images into 128×128 for implementation. For curvelet, the 4-levels wrapping-based real-valued transform [44] was used. In the first experiment, for each subject only one image under normal (frontal) illumination condition was registered as the reference

Table 5

Equal error rates (EER) of different methods on noise-free images.

| Database | Method | EER |
|----------------|------------|---------------|
| CMU | LBP | 0.3265 |
| | Curvelet | 0.2706 |
| | Contourlet | 0.1809 |
| | DT-CWT | 0.2471 |
| | SQI | 0.0551 |
| | LTV | 0.0537 |
| | LWT | 0.0294 |
| | LNSCT | 0.0176 |
| Extended YaleB | LBP | 0.4745 |
| | Curvelet | 0.4645 |
| | Contourlet | 0.3538 |
| | DT-CWT | 0.4515 |
| | SQI | 0.2297 |
| | LTV | 0.2193 |
| | LWT | 0.1926 |
| | LNSCT | 0.1541 |

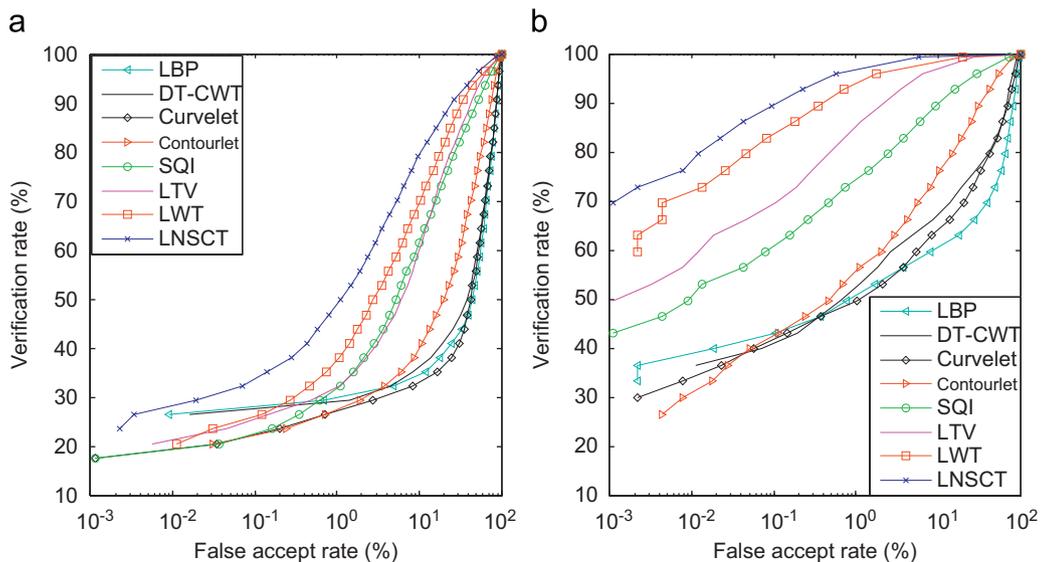


Fig. 11. ROC curves of different methods on noise-free images from (a) Extended YaleB and (b) CMU database.

Table 6
Face recognition rates of different methods on noise-blurred images. The images under frontal illumination condition are used as the reference images.

| Database | Method | Recognition rate (%) | | | |
|----------------|------------|----------------------|----------------|----------------|----------------|
| | | $\sigma=0.05$ | $\sigma=0.1$ | $\sigma=0.15$ | $\sigma=0.2$ |
| CMU | LBP | 7.0588 | 2.3529 | 1.5441 | 1.4706 |
| | Curvelet | 78.0147 | 77.7206 | 76.5441 | 75.9559 |
| | Contourlet | 77.0588 | 77.1324 | 76.7647 | 76.6912 |
| | DT-CWT | 89.1176 | 88.3824 | 87.2059 | 86.2500 |
| | SQI | 99.2647 | 1.4706 | 1.4706 | 1.4706 |
| | LTV | 99.6324 | 99.7059 | 99.7059 | 99.4118 |
| | LWT | 99.8529 | 97.5335 | 97.5335 | 97.4265 |
| | LNSCT | 99.8529 | 99.7794 | 99.8529 | 99.8529 |
| Extended YaleB | LBP | 3.0493 | 2.6316 | 2.6316 | 2.6316 |
| | Curvelet | 43.7761 | 44.0267 | 43.8179 | 43.4002 |
| | Contourlet | 43.7761 | 44.0267 | 43.8179 | 43.4002 |
| | DT-CWT | 55.556 | 55.0125 | 54.0936 | 53.6759 |
| | SQI | 2.8822 | 2.8822 | 2.8404 | 2.8404 |
| | LTV | 84.0434 | 83.9181 | 83.6257 | 83.8764 |
| | LWT | 82.5815 | 67.0844 | 67.0426 | 67.0008 |
| | LNSCT | 84.1270 | 84.2941 | 84.0852 | 84.2105 |

image, and the rest images were treated as the query ones. For Set 1–5 of Extended YaleB, there were 228(=266–1 × 38), 456, 456, 532, and 722 query images, respectively. For CMU, there were 1360(=20 × 68) query images. The face recognition rates on different face databases (subsets) are tabulated in Table 3. It is shown that our proposed method gets the highest recognition rates on all of the selected face databases, and significant improvements are obtained in some challenging cases. Since the variations of lighting in CMU and the Set 1–2 of Extended YaleB are relatively small, the performance discrepancy among the compared techniques is not too much notable. However, under the challenging lighting conditions, e.g. on the Set 3–5 of Extended YaleB, the performances of other methods drop, whereas LNSCT can attain robust recognition results. It is mainly because of the multi-scale and multi-directivity analysis of NSCT, so that LNSCT can still effectually extract facial features from the face images with large variations in illumination. Though, LWT was claimed to perform better than LTV in Ref. [21], our experimental results show that LWT does not always get higher recognition rates than LTV model especially on CMU database and Set 5 of Extended YaleB database. It may be

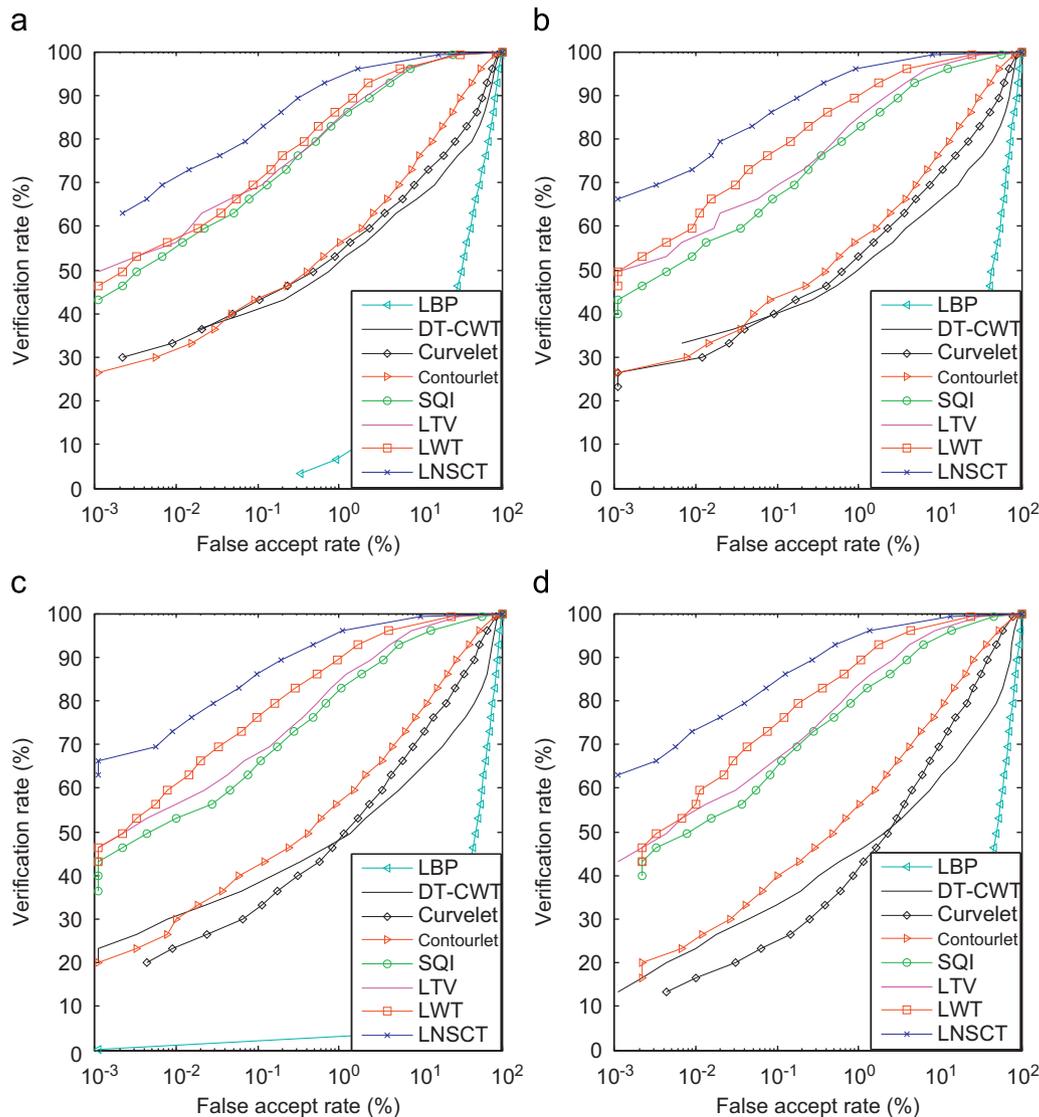


Fig. 12. ROC curves of different methods on blurred images with different level noises on CMU database. (a) $\sigma=0.05$, (b) $\sigma=0.1$, (c) $\sigma=0.15$, and (d) $\sigma=0.2$.

because in Ref. [21] multiple images of each subject were selected to form the training set. For LBP, curvelet, contourlet, and DT-CWT, they seem to do not work very well on the image with large variations in illumination. One main reason could be that unlike LNST, they are not based on a physical illumination model.

In real-world applications, it is hard to guarantee that ideal images are used to establish the reference set. In some cases, the reference images are the ones under uncertain illumination conditions. In the second experiment, three images under uncertain illumination conditions for each person were randomly selected to form the reference set, and the rest images were for probing. The recognition results were averaged over 50 random trials. The final results are shown in Table 4. It shows that the proposed approach still obtains the highest recognition rates on all databases. The above experimental results show that the proposed LNST has consistent outstanding performance, no matter the reference images are under normal or uncertain illumination conditions.

In the verification mode, the ROC curves [42], which show false accept rate (FAR) versus face verification rate (FVT), are illustrated in Fig. 11, and the EER of each method is displayed in Table 5. It shows that LNST consistently gets the highest face verification rate and the lowest EER on Extended YaleB and CMU databases.

The above experimental results justify that in the noise-free case the LNST model could capture more robust intrinsic facial features which are more discriminative as compared to other state-of-the-art algorithms.

4.2. Face recognition on noise-blurred images

In the process of capturing or transmitting image, extra noise would be added into the image. Noise can seriously affect other facial image processing, such as features extraction and synthesis of image. So it is important to investigate the robustness of a feature-extraction method to noise. This section demonstrates the performance of LNST based on Eq. (11) for face recognition on noise-blurred images. For each face database, the query images have been polluted by the artificial multiplicative Gaussian noise with different standard deviations ($\sigma=0.05$, $\sigma=0.1$, $\sigma=0.15$, and $\sigma=0.2$, respectively) and the mean equals to 1. Then the facial features extracted by different methods from those blurred images were used for face recognition. For each subject, only one image under normal illumination condition was registered as the reference image and the rest images were the query ones.

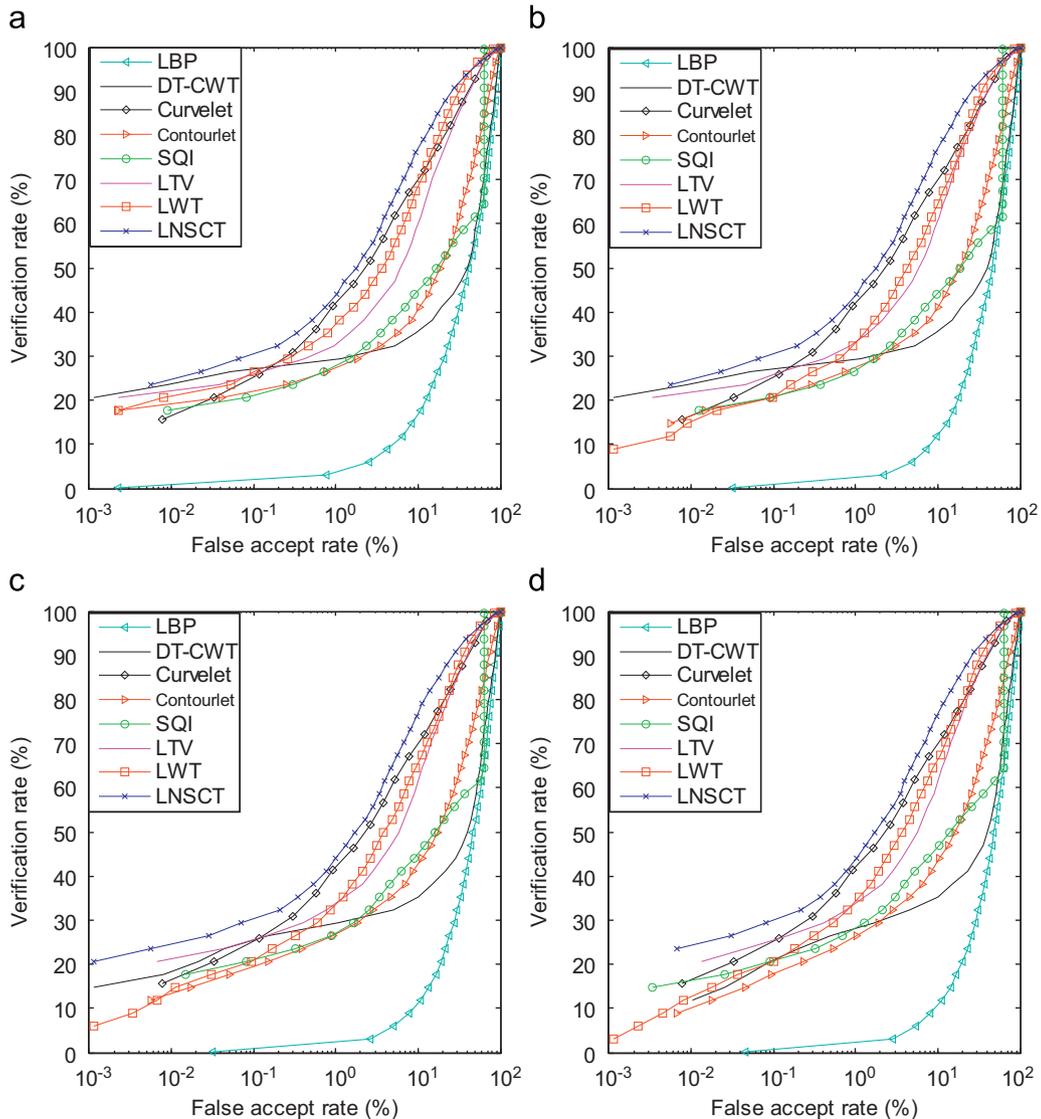


Fig. 13. ROC curves of different methods on blurred images with different level noises on YaleB database. (a) $\sigma=0.05$, (b) $\sigma=0.1$, (c) $\sigma=0.15$, and (d) $\sigma=0.2$.

Table 7
Equal error rates (EER) of different methods on noise-blurred images.

| Database | Method | EER | | | |
|----------------|------------|---------------|---------------|---------------|---------------|
| | | $\sigma=0.05$ | $\sigma=0.1$ | $\sigma=0.15$ | $\sigma=0.2$ |
| CMU | LBP | 0.4066 | 0.4721 | 0.4919 | 0.5015 |
| | Curvelet | 0.2250 | 0.2110 | 0.2059 | 0.2066 |
| | Contourlet | 0.1794 | 0.1735 | 0.1662 | 0.1640 |
| | DT-CWT | 0.2515 | 0.2588 | 0.2647 | 0.2757 |
| | SQI | 0.0551 | 0.0632 | 0.0632 | 0.0691 |
| | LTV | 0.0522 | 0.0551 | 0.0559 | 0.0581 |
| | LWT | 0.0478 | 0.0397 | 0.0397 | 0.0412 |
| | LSNCT | 0.0272 | 0.0228 | 0.0243 | 0.0243 |
| Extended YaleB | LBP | 0.4632 | 0.4766 | 0.4871 | 0.4925 |
| | Curvelet | 0.2066 | 0.2066 | 0.2066 | 0.2066 |
| | Contourlet | 0.3521 | 0.3469 | 0.3442 | 0.3409 |
| | DT-CWT | 0.4541 | 0.4570 | 0.4591 | 0.4591 |
| | SQI | 0.4018 | 0.4169 | 0.4077 | 0.3985 |
| | LTV | 0.2189 | 0.2168 | 0.2151 | 0.2147 |
| | LWT | 0.1909 | 0.2193 | 0.2059 | 0.2047 |
| | LSNCT | 0.1658 | 0.1612 | 0.1661 | 0.1642 |

Table 6 shows the recognition rates of different methods with respect to noise of different levels. On the noisy face images, the recognition performances of LTV and LNSCT keep more stable than other methods. Especially, the recognition rates of LBP and SQI are conspicuously low. In verification as shown in Figs. 12, 13, and Table 7, the LNSCT obtained higher ROC curve and notably lower EER as compared to the other methods. This is because LNSCT can in theory distinguish facial structures from multiplicative noises so that LNSCT can attain more robust extraction of facial features.

5. Conclusions

In this paper, the logarithm nonsubsampled contourlet transform (LSNCT) has been proposed to extract the illumination invariant features from a single face image for face recognition. Benefiting from the multi-scale and multi-directivity analysis, NSCT can efficiently capture the contours in natural image and also effectively approximate the low frequency illumination fields of a face image. Using NSCT as image decomposition in the logarithm domain, LNSCT extracts the strong edges, weak edges, and noises from a face image, and then effectively estimates the intrinsic features (reflectance component) for face recognition. We have also shown that by using LNSCT the illumination-invariant facial features can still be very well extracted from a noise-blurred image, as the noise, especially the multiplicative noise, can be effectively suppressed. Note that no lighting or 3D shape assumption and no training set are required for the proposed LNSCT. Encouraging experimental results on face databases under uneven lighting conditions have shown the effectiveness of the proposed method. It is noted that the computational complexity of our algorithm is mainly determined by the computational cost of NSCT. On a PC with Intel Due Core 2.66 GHz CPU and 3.25 GB RAM, NSCT (with codes provided by the authors of [27]) costs on average 16.45 s to decompose a 100×100 image using Matlab. This may still be a limitation of LNSCT for real-time application. Hence, a future research issue could be to develop fast computation or approximate decomposition methods for NSCT.

Acknowledgments

The authors would like to thank the authors of Refs. [20,22,27,44] who have offered the code of LTV, DT-CWT,

(Nonsubsampled) contourlet, and curvelet, respectively. The contourlet toolbox used in this paper can be downloaded at <http://www.mathworks.com/matlabcentral/fileexchange/8837>. This project was supported by the NSFC (60675016, 60633030), the 973 Program (2006CB303104) and NSF-Guangdong (U0835005).

References

- [1] P. Phillips, P. Grother, R. Micheals, D. Blackburn, E. Tabassi, J. Bone, FRVT 2002: Evaluation report, March 2003.
- [2] P. Phillips, W. Scruggs, A. O'Toole, P. Flynn, K. Bowyer, C. Schott, M. Sharpe, FRVT 2006 and ICE 2006 Large-Scale Results, National Institute of Standards and Technology, NISTIR, 2007.
- [3] W. Choi, S. Tse, K. Wong, K. Lam, Simplified Gabor wavelets for human face recognition, *Pattern Recognition* 41 (2008) 1186–1199.
- [4] T. Ahonen, A. Hadid, M. Pietikäinen, Face description with local binary patterns: application to face recognition, *IEEE Transactions on Pattern Analysis and Machine Intelligence* 28 (2006) 2037–2041.
- [5] W. Zhang, S. Shan, W. Gao, X. Chen, H. Zhang, Local Gabor binary pattern histogram sequence (LGBPHS): a novel non-statistical model for face representation and recognition, in: *Proceeding of IEEE International Conference on Computer Vision (ICCV)*, 2005.
- [6] B.K.P. Horn, in: *Robot Vision*, Academic, MA, 1990.
- [7] Y. Weiss, Deriving intrinsic images from image sequences, in: *Proceeding of IEEE International Conference on Computer Vision (ICCV)*, 2001.
- [8] R. Rammamorthi, P. Hanrahan, A signal-processing framework for inverse rendering, in: *Proceedings of ACM SIGGRAPH*, 2001.
- [9] T.G. Stockham Jr., Image processing in the context of a visual model, *Proceedings of IEEE* 60 (1972) 828–842.
- [10] B.K.P. Horn, Determining lightness from an image, *Computer Graphics and Image Processing* 3 (1974) 277–299.
- [11] E.H. Land, J.J. McCann, Lightness and Retinex theory, *Optical Society of America* 61 (1997) 1–11.
- [12] D.J. Jobson, Z. Rahman, G.A. Woodell, A multi-scale Retinex for bridging the gap between color images and the human observation of scenes, *IEEE Transactions on Image Processing* 6 (1997) 965–976.
- [13] P. Perona, J. Malik, Scale-space and edge detection using anisotropic diffusion, *IEEE Transactions on Pattern Analysis and Machine Intelligence* 12 (1990) 629–639.
- [14] J. Tumblin, G. Turk, LCIS: a boundary hierarchy for detail-preserving contrast reduction, in: *Proceedings of ACM SIGGRAPH*, 1999.
- [15] R. Gross, V. Brajovie, An image preprocessing algorithm for illumination invariant face recognition, in: *Proceeding of Fourth International Conference on Audio and Video Based Biometric Person Authentication*, 2003.
- [16] C. Tomasi, R. Manduchi, Bilateral filtering for gray and color images, in: *Proceedings of IEEE International Conference on Computer Vision*, 1998.
- [17] F. Durand, J. Dorsey, Fast bilateral filtering for the display of high-dynamic-range images, *ACM Transactions on Graphics* 21 (2002) 257–266.
- [18] D. Comaniciu, P. Meer, Mean shift: a robust approach toward feature space analysis, *IEEE Transactions on Pattern Analysis and Machine Intelligence* 24 (2002) 603–619.
- [19] C. Chen, C. Chen, Lighting normalization with generic intrinsic illumination subspace for face recognition, in: *Proceeding of IEEE International Conference on Computer Vision (ICCV)*, 2005.
- [20] T. Chen, W. Yin, X. Zhou, D. Comaniciu, T.S. Huang, Total variation models for variable lighting face recognition, *IEEE Transactions on Pattern Analysis and Machine Intelligence* 28 (2006) 1519–1524.
- [21] T. Zhang, B. Fang, Y. Yuan, Y. Tang, Z. Shang, D. Li, F. Lang, Multiscale facial structure representation for face recognition under varying illumination, *Pattern Recognition* 42 (2009) 251–258.
- [22] N.G. Kingsbury, Complex wavelets for shift invariant analysis and filtering of signals, *Journal of Applied and Computational Harmonic Analysis* 10 (2001) 234–253.
- [23] P.D. Kovess, Image features from phase congruency, *Videre: Journal of Computer Vision Research* 1 (1999) 1–26.
- [24] M.N. Do, M. Vetterli, The contourlet transform: an efficient directional multiresolution image representation, *IEEE Transactions on Image Processing* 14 (2005) 2091–2106.
- [25] R.R. Coifman, D.L. Donoho, Translation invariant de-noising, in: A. Antoniadis, G. Oppenheim (Eds.), *Wavelets and Statistics*, Springer-Verlag, New York, 1995, pp. 125–150.
- [26] M.N. Do, M. Vetterli, in: J. Stoeckler, G.V. Welland (Eds.), *Contourlets, BeyondWavelets*, Academic Press, New York, 2002.
- [27] L. Cunha, J. Zhou, M.N. Do, The nonsubsampled contourlet transform: theory, design, and applications, *IEEE Transactions on Image Processing* 15 (2006) 3089–3101.
- [28] W.R. Boukabou, A. Bouridane, Contourlet-based feature extraction with PCA for face recognition, in: *Proceedings of the 2008 NASA/ESA Conference on Adaptive Hardware and Systems Table of Contents*, Washington, DC, USA, 2008, 482–486.

- [29] T. Mandal, A. Majumdar, Q.M.J. Wu, Face recognition by Curvelet based feature extraction, *Lecture Notes in Computer Science* 4633 (2007) 806–817.
- [30] K. Delac, M. Grgic, M.S. Bartlett, Recent advances in face recognition, 978-953-7619-34-3, I-Tech, Vienna, Austria, 2008, 236.
- [31] S.G. Chang, B. Yu, M. Vetterli, Adaptive wavelet thresholding for image denoising and compression, *IEEE Transactions on Image Process* 9 (2000) 1532–1546.
- [32] W. Chen, M.J. Er, S. Wu, Illumination compensation and normalization for robust face recognition using discrete cosine transform in logarithm domain, *IEEE Transactions on Systems, Man and Cybernetics, Part B* 36 (2006) 458–466.
- [33] A. Georghiades, P. Belhumeur, D. Kriegman, From few to many: illumination cone models for face recognition under variable lighting and pose, *IEEE Transactions on Pattern Analysis and Machine Intelligence* 23 (2001) 643–660.
- [34] T. Sim, S. Baker, M. Bsat, The CMU pose, illumination, and expression (PIE) database, in: *Proceedings of IEEE Conference on Face and Gestures (FGR)*, May, 2002.
- [35] S. Shan, W. Gao, B. Cao, D. Zhao, Illumination normalization for robust face recognition against varying lighting conditions, in: *Proceedings of International Workshop Analysis and Modeling of Faces and Gestures*, 2003.
- [36] X. Xie, K. Lam, Face recognition under varying illumination based on a 2D face shape model, *Pattern Recognition* 38 (2005) 221–230.
- [37] S. Choi, C. Kim, C. Choi, Shadow compensation in 2D images for face recognition, *Pattern Recognition* 40 (2007) 2118–2125.
- [38] V. Blanz, T. Vetter, Face recognition based on fitting a 3D morphable model, *IEEE Transactions on Pattern Analysis and Machine Intelligence* 25 (2003) 063–1074.
- [39] H. Shim, J. Luo, T. Chen, A subspace model-based approach to face relighting under unknown lighting and poses, *IEEE Transactions on Image Processing* 17 (2008) 1331–1341.
- [40] A. Shashua, T. Riklin-Raviv, The quotient image: class-based re-rendering and recognition with varying illuminations, *IEEE Transactions on Pattern Analysis and Machine Intelligence* 23 (2001) 129–139.
- [41] H. Wang, S.Z. Li, Y. Wang, Generalized quotient image, in: *Proceeding of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2004.
- [42] P. Phillips, H. Moon, S. Rizvi, P. Rauss, The FERET evaluation methodology for face-recognition algorithms, *IEEE Transactions on Pattern Analysis and Machine Intelligence* 22 (2000) 1090–1104.
- [43] A. Eleyan, H. Ozkaramanli, H. Demirel, Complex wavelet transform-based face recognition, *EURASIP Journal on Advances in Signal Processing*, 2008, Article ID 185281.
- [44] E. Candes, L. Demanet, D. Donoho, et al., Fast discrete curvelet transforms, *Multiscale modeling and simulation* 5 (2007) 861–899.
- [45] E. Bezael, U. Efron, Efficient face recognition method using a combined phase congruency/Gabor wavelet technique, in: *Proceedings of SPIE, Optical Information Systems III*, August, 2005.

Xiaohua Xie is a Ph.D. candidate of Sun Yat-sen University in PR China. He is currently a visiting student, under the supervision of Prof. Ching Y. Suen, at Concordia University in Canada. His research interests include pattern recognition and computer vision, especially focusing on illumination and pose normalization of human face.

Jianhuang Lai was born in 1964. He received the M.Sc. degree in applied mathematics in 1989 and the Ph.D. degree in mathematics in 1999 from Sun Yat-sen University, Guangzhou, China. He joined Sun Yat-sen University in 1989, where currently, he is a Professor with the Department of Automation, School of Information Science and Technology. He has published over 60 papers in the international journals, book chapters, and conferences. His current research interests are in the areas of digital image processing, pattern recognition, multimedia communication, wavelets and their applications. He has taken charge of more than five research projects, including NSF-Guangdong (U0835005), NSFC (Numbers 60144001, 60 373 082, 60675016), the Key (Key grant) Project of Chinese Ministry of Education (Number 105 134), and NSF of Guangdong, China (Number 021 766, 06023194). He serves as a board member of the Image and Graphics Association of China and also serves as a board member and chairman of the Image and Graphics Association of Guangdong.

Wei-Shi Zheng received the B.S. degree in science with specialties in mathematics and computer science and the Ph.D. degree in applied mathematics from Sun Yat-Sen University, Guangzhou, China, in 2003 and 2008, respectively. He is a Postdoctoral Researcher at the Department of Computer Science, Queen Mary University of London, London, UK. He is now working on the European SAMURAI Research Project with Prof. S. Gong and Dr. T. Xiang. He has been a visiting student working with Prof. Z. Stan Li at the Institute of Automation, Chinese Academy of Sciences, Beijing, China, and an exchanged research student working with Prof. Yuen Pong C. at Hong Kong Baptist University, Hong Kong. His current research interests are in object association and categorization. He is also interested in discriminant/sparse feature extraction, dimension reduction, kernel methods in machine learning, and face image analysis. Dr. Zheng was awarded the HP Chinese Excellent Student Scholarship 2008.