Illumination robust face representation based on intrinsic geometrical information

Soyel, H; Ozmen, B; McOwan, PW

Illumination Robust Face Representation Based on Intrinsic Geometrical Information

Hamit Soyel[†], Burcin Ozmen[‡] and Peter W. McOwan[†]

[†] Department of Computer Science, Queen Mary University of London, United Kingdom.
[‡] Department of Electrical and Electronic Engineering, Cyprus International University, via Mersin 10, Turkey.
{hsoyel, pmco}@eecs.qmul.ac.uk, bozmen@ciu.edu.tr

Abstract

The ability to recognize people is a key element for improving naturalistic human-robot and human-computer interaction systems. In this paper, we propose a binary nonsubsampled contourlet transform (B-NSCT) based illumination robust face representation. Faces are transformed into multi-scale and multi-directional contour information where the intrinsic geometrical structures are used for characterising facial texture. Experiments on the Yale B and CMU PIE databases illustrate that B-NSCT is highly insensitive to illumination variation.

1 Introduction

The ability to recognize and remember individuals is a fundamental requirement for complex interactions among people and represents a key element to improve naturalistic human-robot and human-computer interaction systems. Imprecise human identification will lead to the insufficient interactions. Therefore, it is necessary to recognize human faces with a high accuracy. Although various methods for face recognition have been proposed such as eigenfaces [14], fisherfaces [2] and Bayesian faces [11], the performance of most existing algorithms is highly sensitive to environmental illumination.

As the use of face recognition systems expands towards less restricted real-time approaches, illumination robust face representation becomes important. Previous studies proved that illumination variations are more significant than the inherent differences between individuals for face recognition [4, 3]. Extracting illumination invariant features or illumination insensitive measure is an effective approach to solve this problem. In general, classical and effective methods for varying lighting face recognition mainly include multi-scale retinex (MSR) [10], self quotient image (SQI) [15], logarithmic total variation (LTV) [5], multi-scale principal contour direction (MPCD) [6] and logarithmic nonsubsampled contourlet transform (LNSCT) [16].

In the MSR model, illumination variations are dealt with using the difference between an original image and its smoothed version in logarithm domain by combining several low-pass filters with different cut-off frequencies. But the halo effect of MSR is serious. The illumination effect is normalized by division over a smoothed version of image itself in the SQI model. However, the weighted Gaussian filter they used has trouble keeping sharp edges in low frequency illumination fields. LTV improves the SQI by adopting the total variation model which utilizes the edge preserving capability and simplify the parameter selection but has quite high computational expense.

The above methods lack the capabilities of multi-scale and multi-directional analysis. This limits their abilities in capturing multi-scale and multi-directional structures of face images. In order to address this problem MPCD is proposed for extracting illumination invariant features by applying multi-scale contour analysis. The method has demonstrated better recognition rates than the above mentioned ones by considering both spectrum and phase information. However, its parameter selection is usually empirical and complicated. Recently, LNSCT is proposed to estimate the reflectance component using nonsubsampled contourlet transform (NSCT) in logarithm domain. However, it was noted that the computational complexity limits LNSCT for real-time applications.

In this paper, we propose a novel representation based on a binary nonsubsampled contourlet transform (B-NSCT) for illumination robust face recognition. Grey scale images are transformed into multi-scale and multi-directional contour information where the intrinsic geometrical structures are used for characterising facial texture. Experimental results on Yale B and CMU PIE databases demonstrated that B-NSCT based face representation is effective, speedy and highly insensitive to illumination variation.

The rest of this paper is organized as follows. In Section 2, we explain the B-NSCT based methodology for illumination robust face representation. A variety of experimental results are presented in Section 3. Finally, the conclusion is give in Section 4.

2 Methodology

2.1 Nonsubsampled countourlet transform

In [8], Do and Vetterli proposed contourlet transform (CT) to represent two dimensional singularities, which is composed of Laplacian pyramid and directional filter bank. The transform can represent curve more sparsely due to its di-

rectionality and anisotropy. NSCT, based on the theory of CT, is a kind of multi-scale, multi-directional computation framework of discrete images. The whole course of NSCT is still composed of two stages, including multi-scale analysis and multi-directional analysis, which are similar to those of CT. The main difference lies in that, in the course of decomposition and reconstruction, traditional upsamplers and downsamplers in CT do not exist in NSCT any more, so that the NSCT is a fully multi-scale, multi-directional, good time-frequency property and shift invariant expansion. NSCT falls into two phases, including nonsubsampled pyramid (NSP) and nonsubsampled directional filter bank (NSDFB). The former phase ensures the multi-scale property by using two-channel nonsubsampled filter bank, and a low-pass image with a band-pass one can be produced at each NSP decomposition stage. The subsequent NSP decomposition stages are carried out to decompose the low-pass component available iteratively to capture the singularities in the image. As a result, NSP can result in $k+1$ subband images including one low-pass image and k band-pass images, whose sizes are all the same as that of the source image, where k denotes the number of decomposition stages. Figure 1 gives the NSP decomposition with $k=3$ stages with a 2-D low-pass filter is represented by its z-transform $H_0(z)$ where $z = [z_1, z_2]^T$ and $H_1(z) = 1 - H_0(z)$. The filters for subsequent stages are obtained by upsampling the filters of the first stage. This gives the multi-scale property without the need for additional filter design.
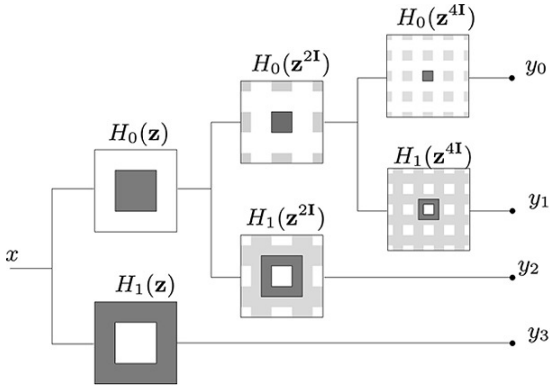

Figure 1: Three-stage NSP decomposition

The NSDFB, constructed by combining the directional fan filter banks devised by Bamberger and Smith [1], is two-channel nonsubsampled filter banks. NSDFB allows the direction decomposition with l stages in band-pass images from NSP at each scale and produces $2l$ directional subband images which have the same size as the source image. Thus, the NSDFB endows the NSCT with the multi-direction property and we can benefit a lot from the NS-DFB because it provides us with more precise directional detail information. A four-channel NSDFB, constructed with two-channel fan filter banks and parallelogram filters without downsamplers and upsamplers, is illustrated in Figure 2. Note that in the second level, the upsampled fan

filters $U_i(z^Q), i = 0, 1$ have checker-board frequency support, and when combined with the filters in the first level give the four directional frequency decomposition. In our paper, the 'maxflat' filters and 'dmaxflat7' filters are, respectively, selected for NSP and NSDFB. The concrete filter banks construction methods and more NSCT details can be found in [7].
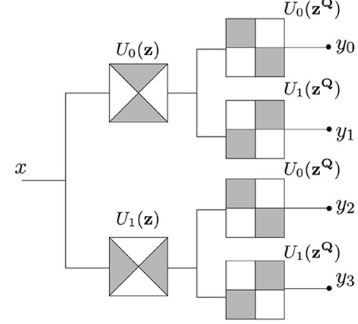

Figure 2: Four-channel nonsubsampled DFB

2.2 Binary nonsubsampled contourlet transform.

NSCT differs from other multi-scale analysis methods in that contourlet transform allows for different and flexible number of directions at each scale. According the direction information contours are obtained by directional filter bank concatenated the neighbouring singular points into local contours in the frequency domain. By combination of NSP and NSDFB, NSCT is constructed as a fully shift invariant, multi-scale, and multi-direction expansion that has better directional frequency localization and a fast implementation.

It is worth to note that shift invariance is very important. Being shift invariant, each pixel of the transform subbands corresponds to that of the original image in the same spatial location. Therefore, we gather the geometrical information pixel by pixel from the NSCT coefficients. All directional contour subbands can be expressed as:

$$\{C_{m,d}\}, \ m = 1, 2, \ldots k, \ d = 1, 2, \ldots l_m, \\ k \in (1, 2, \ldots N), \ l_m = 2^N. \tag{1}$$

where m and d are the scale and direction of the decomposition respectively, k is the number of contour decomposition scale, l_m is the number of contour decomposition directions of mth scale and $\{C_{m,d}\}$ is the coefficient at the dth directional subband of the mth scale. As mentioned above directional contour subbands of NSCT only includes spectrum information and keeping the the most significant coefficients will directly lead to improvement in feature extraction [12]. We observe that in the NSCT domain, the illumination component corresponds to those pixels with positive coefficient points and the reflectance component mainly corresponds with negative coefficient points. Since the reflectance component is considered as the intrinsic facial features, it can be directly used for face recognition. Based on this observation binary nonsubsampled contourlet
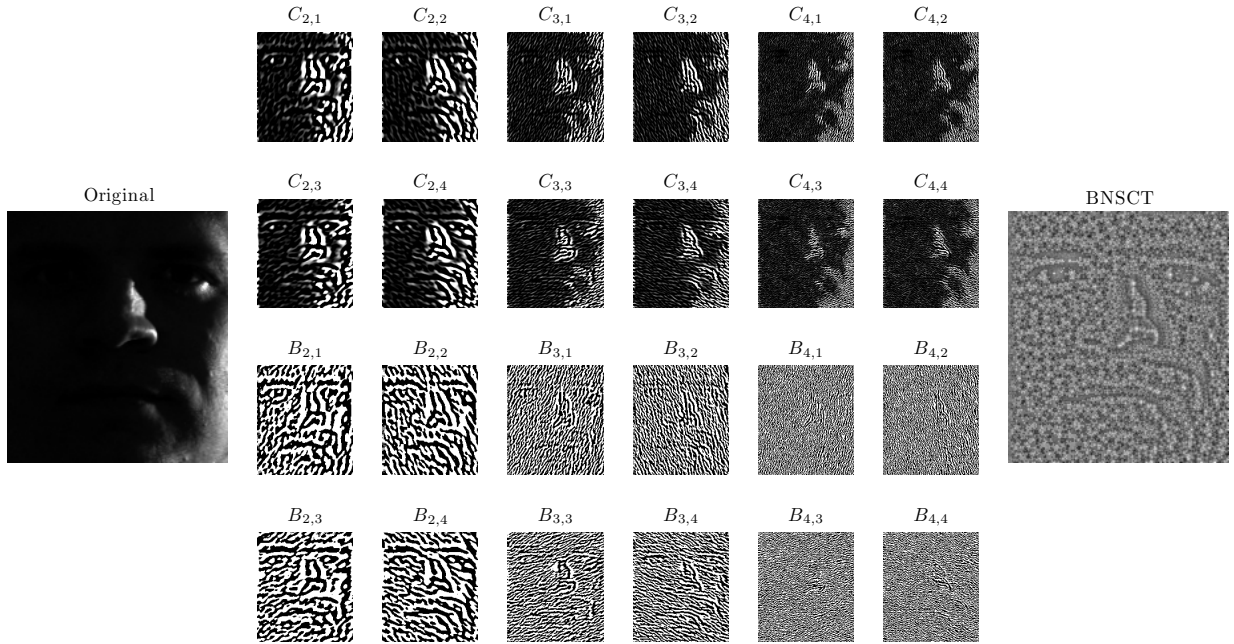
Figure 3: Example of constructing illumination invariant representation with three scales and four directional subbands in each scale.

transform (B-NSCT) of an image can be defined as:

$$B_{m,d}(x,y) = \begin{cases} 1, & \text{if } \{C_{m,d}(x,y)\} > 0 \\ 0, & \text{if } \{C_{m,d}(x,y)\} \leq 0 \end{cases}$$
$$m = 1, 2, \ldots k, \ d = 1, 2, \ldots l_m. \qquad (2)$$

where $\{C_{m,d}(x,y)\}$ is the coefficient point (x,y) at the jth directional subband of ith scale. After extracting the intrinsic geometrical information by this simple binarization process, the illumination invariant representation can be reconstructed from the B-NSCT coefficients, $\{B_{m,d}(x,y)\}$, by inverse NSCT. Figure 3 shows an example of constructing illumination invariant by using contourlet transform with three scales and four directional subbands in each scale.

3   Experimental Results

In our experiments, the proposed B-NSCT system has been evaluated on Yale B [9] and CMU-PIE [13] databases. In the phase of recognition, dimensionality reduction is performed by PCA and the nearest neighbour classifier based on Euclidean distance is used for classification. The results yielded by our method were compared with state-of-the-art algorithms including MSR [10], SQI [15], LTV [5], MPCD [6] and LNSCT [16]. Quality of results is quantitatively evaluated by recognition rates. In the proposed method, we use three-scale NSCT decomposition and four directional subbands in each scale.

3.1   Experiments on the Yale B face database

The Yale B database contains 10 individuals under 64 different lighting conditions for 9 poses. Since we only focus on the illumination problem in this paper, only frontal

Method	Subset 2	Subset 3	Subset 4	Subset 5
MSR	99.17%	98.33%	92.86%	90.00%
SQI	99.17%	98.33%	93.57%	94.21%
LTV	100.00%	99.17%	97.86%	95.79%
MPCD	100.00%	100.00%	100.00%	99.47%
LNSCT	100.00%	100.00%	100.00%	99.47%
B-NSCT	100.00%	100.00%	100.00%	100.00%

Table 1: Recognition accuracy (%) of various methods when using images of subset 1 as training set.

face images under varying illumination conditions are chosen as samples. All the images are manually cropped by the positions of the eyes and the mouth and resized to 192 × 168, which include only the face with as little hair and background as possible. According to illumination angle, the cropped images fall into five subsets, *i.e.* subset 1 ($0°$ − $12°$), subset 2 ($13°$ − $25°$), subset 3 ($26°$ − $50°$), subset 4 ($51°$ − $77°$) and subset 5 (above $77°$). Figure 4 shows five images for each subset for one person and corresponding illumination invariants obtained by the proposed B-NSCT methodology. Performance evaluation is performed by choosing one of the five subsets as training and the remaining subsets for testing, respectively. The number of principal components retained for each method is 45. The recognition rates on different subsets tabulated in Table 1-5 illustrate the effectiveness of the proposed methodology.

3.2   Experiments on the CMU PIE face database

The CMU PIE has 68 subjects with different poses, illuminations and expressions. Only frontal face images un-
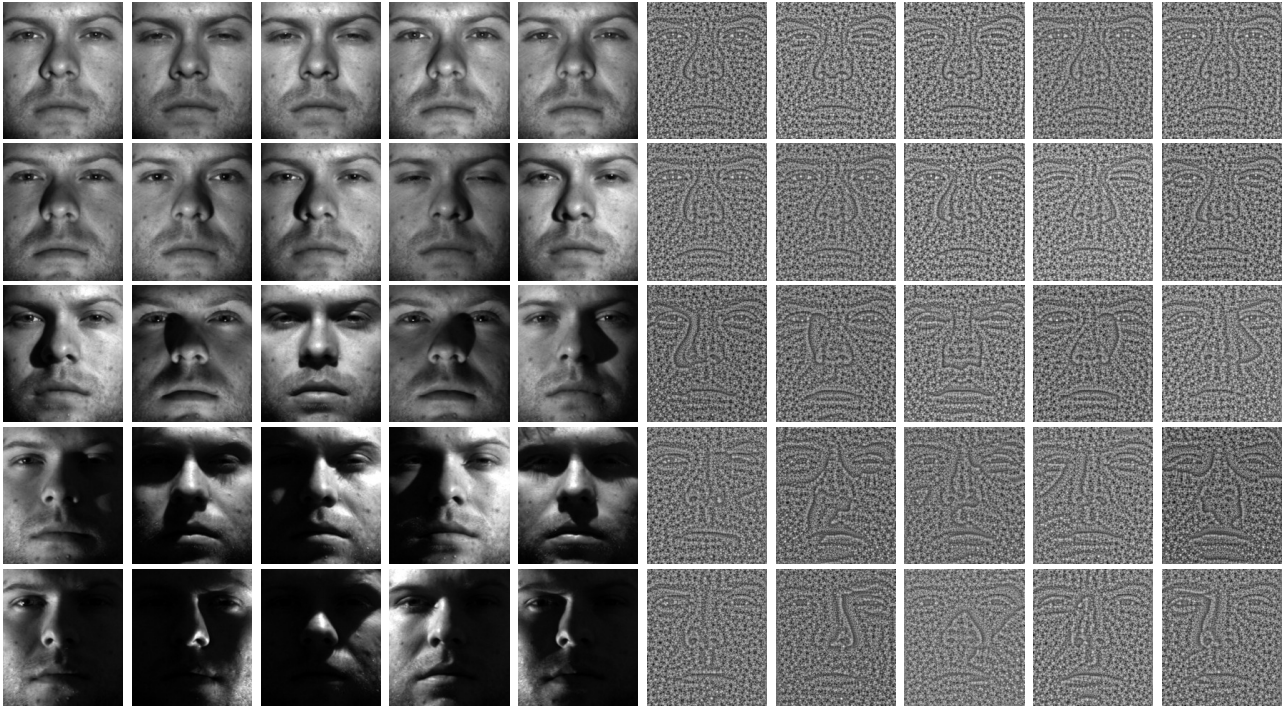
Figure 4: Original images and corresponding illumination invariant obtained by B-NSCT methodology.

Method	Subset 1	Subset 3	Subset 4	Subset 5
MSR	99.17%	96.67%	93.57%	94.21%
SQI	99.17%	98.33%	98.57%	99.47%
LTV	100.00%	98.33%	99.29%	99.47%
MPCD	100.00%	100.00%	100.00%	99.47%
LNSCT	100.00%	100.00%	100.00%	99.47%
B-NSCT	100.00%	100.00%	100.00%	100.00%

Table 2: Recognition accuracy (%) of various methods when using images of subset 2 as training set.

Method	Subset 1	Subset 2	Subset 4	Subset 5
MSR	100.00%	99.17%	98.57%	98.42%
SQI	100.00%	100.00%	98.57%	99.47%
LTV	100.00%	100.00%	99.29%	99.47%
MPCD	100.00%	100.00%	100.00%	99.47%
LNSCT	100.00%	100.00%	100.00%	100.00%
B-NSCT	100.00%	100.00%	100.00%	100.00%

Table 3: Recognition accuracy (%) of various methods when using images of subset 3 as training set.

Method	Subset 1	Subset 2	Subset 3	Subset 5
MSR	97.14%	98.33%	99.17%	98.42%
SQI	99.17%	98.57%	99.17%	97.90%
LTV	100.00%	100.00%	100.00%	99.47%
MPCD	100.00%	100.00%	100.00%	100.00%
LNSCT	100.00%	100.00%	100.00%	100.00%
B-NSCT	100.00%	100.00%	100.00%	100.00%

Table 4: Recognition accuracy (%) of various methods when using images of subset 4 as training set.

Method	Subset 1	Subset 2	Subset 3	Subset 4
MSR	87.14%	92.50%	96.67%	98.57%
SQI	97.14%	94.17%	98.33%	97.14%
LTV	100.00%	94.17%	99.17%	98.57%
MPCD	100.00%	100.00%	100.00%	100.00%
LNSCT	100.00%	100.00%	100.00%	100.00%
B-NSCT	100.00%	100.00%	100.00%	100.00%

Table 5: Recognition accuracy (%) of various methods when using images of subset 5 as training set.

der different illumination conditions are used. Face images in the experiments are all cropped in the same way as the Yale B and resized to 128×128. Figure 5 gives 20 images for one subject under different illumination conditions, and corresponding illumination invariants obtained by the proposed B-NSCT methodology. In this experiment, 6 images per person are randomly chosen for training and the remaining 15 for testing. We run the simulation 60 times and average the results over them. The number of principal components is set to 165. The corresponding average recognition rates of various methods listed in Table 6 show that the proposed method is superior to others.

Experimental results verified that proposed method gets the highest recognition rate on all of the selected face databases, and significant improvements obtained in some challenging cases. Since the variations of lighting in CMU and the Set 1-2 of Yale B 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 Yale B, the performances
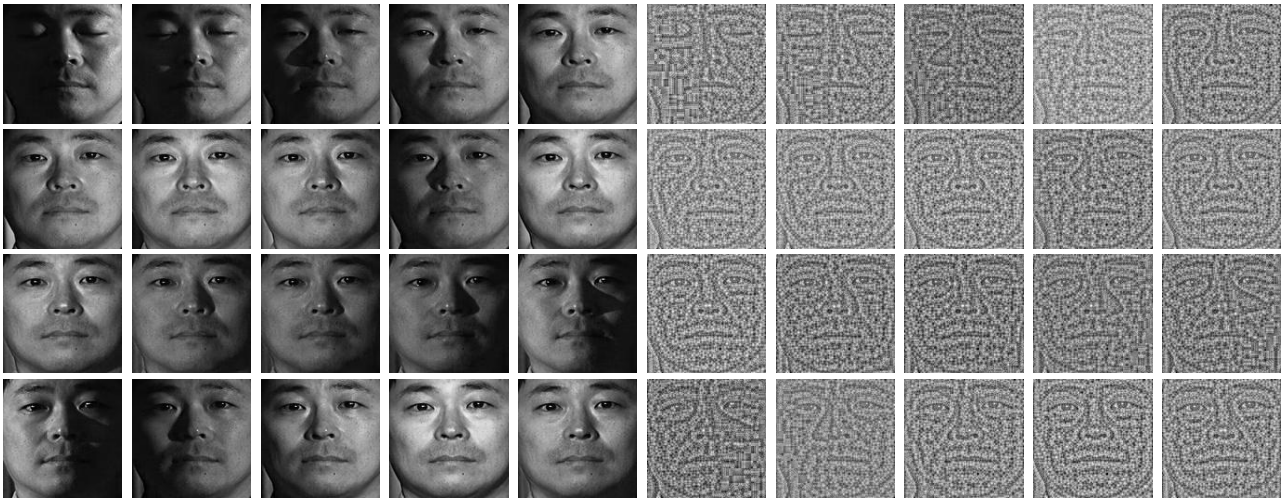
Figure 5: Original images and corresponding illumination invariant obtained by B-NSCT methodology.

Method	Recognition rate
MSR	90.73%
SQI	91.11%
LTV	94.04%
MPCD	99.04%
LNSCT	99.51%
B-NSCT	99.82%

Table 6: Average recognition rates (%) of various methods when randomly choosing six images per person as training set.

of the MSR, SQI and LTV drop dramatically, whereas B-NSCT can attain invariant recognition results. It is mainly because of the multi-scale and multi-directional analysis of NSCT, so that B-NSCT can still effectually extract intrinsic geometrical information from the face images with large variations in illumination.

Furthermore, proposed methodology does not require any parameter selection and on a PC with Intel Due Core 2.66 GHz CPU and 4 GB RAM, B-NSCT costs on average of 1.85s to decompose a 192×168 into directional subbands and to reconstruct illumination invariant representation using Matlab*. Hence, B-NSCT can be easily applied for a naturalistic real-time application.

4 Conclusion

In this paper, we proposed a binary nonsubsampled contourlet transform (B-NSCT) based illumination robust face representation. The proposed method can extract illumination invariant from multi-scale and multi-directional space. Experimental results on the Yale B and CMU-PIE databases verified that binary representation of the intrinsic geometrical information extracted from directional contour coefficients is quite efficient, effective and speedy. In addition, compared with other methods B-NSCT can be directly applied in a naturalistic human-robot or a human-computer interaction system without any preprocessing and conditions.

References

[1] R. H. Bamberger and M. J. T. Smith. A filter bank for the directional decomposition of images: theory and design. *IEEE Trans. Signal Process.*, 40(4):882–893, 1992.

[2] P.N. Belhumeur, J.P. Hespanha, and D.J. Kriegman. Eigenfaces vs. fisherfaces: recognition using class specific linear projection. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 19(7):711–720, jul 1997.

[3] Kevin W. Bowyer, Kyong Chang, and Patrick Flynn. A survey of approaches and challenges in 3d and multi-modal 3d+2d face recognition. *Computer Vision and Image Understanding*, 101(1):1–15, 2006.

[4] R. Chellappa, C.L. Wilson, and S. Sirohey. Human and machine recognition of faces: a survey. *Proceedings of the IEEE*, 83(5):705–741, may 1995.

[5] T. Chen, Wotao Yin, Xiang Sean Zhou, D. Comaniciu, and T.S. Huang. Total variation models for variable lighting face recognition. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 28(9):1519–1524, sept. 2006.

[6] Y. Cheng, C.L. Wang, Z.Y. Li, Y.K. Hou, and C.X. Zhao. Multiscale principal contour direction for varying lighting face recognition. *Electronics Letters*, 46(10):680–682, 13 2010.

*The contourlet toolbox used in this paper can be downloaded at http://www.mathworks.com/matlabcentral/fileexchange/10049-nonsubsampled-contourlet-toolbox.

[7] A.L. da Cunha, Jianping Zhou, and M.N. Do. The nonsubsampled contourlet transform: Theory, design, and applications. *Image Processing, IEEE Transactions on*, 15(10):3089–3101, oct. 2006.

[8] M.N. Do and M. Vetterli. The contourlet transform: an efficient directional multiresolution image representation. *Image Processing, IEEE Transactions on*, 14(12):2091–2106, dec. 2005.

[9] A.S. Georghiades, P.N. Belhumeur, and D.J. Kriegman. From few to many: illumination cone models for face recognition under variable lighting and pose. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 23(6):643–660, jun 2001.

[10] D.J. Jobson, Z. Rahman, and G.A. Woodell. A multiscale retinex for bridging the gap between color images and the human observation of scenes. *Image Processing, IEEE Transactions on*, 6(7):965–976, jul 1997.

[11] Baback Moghaddam, Tony Jebara, and Alex Pentland. Bayesian face recognition. *Pattern Recognition*, 33(11):1771–1782, 2000.

[12] D.D.-Y. Po and M.N. Do. Directional multiscale modeling of images using the contourlet transform. *Image Processing, IEEE Transactions on*, 15(6):1610–1620, june 2006.

[13] T. Sim, S. Baker, and M. Bsat. The cmu pose, illumination, and expression database. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 25(12):1615–1618, dec. 2003.

[14] M. Turk and A. Pentland. Eigenfaces for Recognition. *Journal of Cognitive Neuroscience*, 3(1):71–86, January 1991.

[15] Haitao Wang, S.Z. Li, and Yangsheng Wang. Face recognition under varying lighting conditions using self quotient image. In *Automatic Face and Gesture Recognition, 2004. Proceedings. Sixth IEEE International Conference on*, pages 819–824, may 2004.

[16] Xiaohua Xie, Jianhuang Lai, and Wei-Shi Zheng. Extraction of illumination invariant facial features from a single image using nonsubsampled contourlet transform. *Pattern Recogn.*, 43:4177–4189, December 2010.