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

Keywords: Face Recognition, Illumination variation, Nonsubsampled Contourlet Transform.

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 multiscale 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 by 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 demonstrate 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 given in Section 4.

2 Methodology

2.1 Nonsubsampled contourlet 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.

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 NSDFB 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].

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

Figure 1: Three-stage NSP decomposition

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:

\[ \{ C_{m,d} \}, m = 1, 2, \ldots k, d = 1, 2, \ldots l_m, k \in (1, 2, \ldots N), l_m = 2^N. \]
transform (B-NSCT) of an image can be defined as:

\[
B_{m,d}(x, y) = \begin{cases} 
1, & \text{if } \{C_{m,d}(x,y)\} > 0 \\
0, & \text{if } \{C_{m,d}(x,y)\} \leq 0
\end{cases}
\]

where \(\{C_{m,d}(x,y)\}\) is the coefficient point \((x, y)\) at the \(j\)th directional subband of \(i\)th 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 neighbor 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 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 \(\times\) 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 \((0^\circ - 12^\circ)\), subset 2 \((13^\circ - 25^\circ)\), subset 3 \((26^\circ - 50^\circ)\), subset 4 \((51^\circ - 77^\circ)\) and subset 5 \((above 77^\circ)\). 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-
Table 2: Recognition accuracy (%) of various methods when using images of subset 2 as training set.

<table>
<thead>
<tr>
<th>Method</th>
<th>Subset 1</th>
<th>Subset 2</th>
<th>Subset 3</th>
<th>Subset 4</th>
<th>Subset 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>MSR</td>
<td>99.17%</td>
<td>99.17%</td>
<td>98.57%</td>
<td>98.57%</td>
<td>99.47%</td>
</tr>
<tr>
<td>SQI</td>
<td>99.17%</td>
<td>99.17%</td>
<td>99.47%</td>
<td>99.47%</td>
<td>99.47%</td>
</tr>
<tr>
<td>LTV</td>
<td>100.00%</td>
<td>100.00%</td>
<td>100.00%</td>
<td>100.00%</td>
<td>100.00%</td>
</tr>
<tr>
<td>MPCD</td>
<td>100.00%</td>
<td>100.00%</td>
<td>100.00%</td>
<td>100.00%</td>
<td>100.00%</td>
</tr>
<tr>
<td>LNSCT</td>
<td>100.00%</td>
<td>100.00%</td>
<td>100.00%</td>
<td>100.00%</td>
<td>100.00%</td>
</tr>
<tr>
<td>B-NSCT</td>
<td>100.00%</td>
<td>100.00%</td>
<td>100.00%</td>
<td>100.00%</td>
<td>100.00%</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>Method</th>
<th>Subset 1</th>
<th>Subset 2</th>
<th>Subset 3</th>
<th>Subset 4</th>
<th>Subset 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>MSR</td>
<td>97.14%</td>
<td>98.33%</td>
<td>99.17%</td>
<td>98.42%</td>
<td>98.42%</td>
</tr>
<tr>
<td>SQI</td>
<td>99.17%</td>
<td>99.57%</td>
<td>100.00%</td>
<td>99.47%</td>
<td>99.47%</td>
</tr>
<tr>
<td>LTV</td>
<td>100.00%</td>
<td>100.00%</td>
<td>100.00%</td>
<td>100.00%</td>
<td>100.00%</td>
</tr>
<tr>
<td>MPCD</td>
<td>100.00%</td>
<td>100.00%</td>
<td>100.00%</td>
<td>100.00%</td>
<td>100.00%</td>
</tr>
<tr>
<td>LNSCT</td>
<td>100.00%</td>
<td>100.00%</td>
<td>100.00%</td>
<td>100.00%</td>
<td>100.00%</td>
</tr>
<tr>
<td>B-NSCT</td>
<td>100.00%</td>
<td>100.00%</td>
<td>100.00%</td>
<td>100.00%</td>
<td>100.00%</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>Method</th>
<th>Subset 1</th>
<th>Subset 2</th>
<th>Subset 3</th>
<th>Subset 4</th>
<th>Subset 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>MSR</td>
<td>87.14%</td>
<td>92.50%</td>
<td>96.67%</td>
<td>98.57%</td>
<td>98.57%</td>
</tr>
<tr>
<td>SQI</td>
<td>97.14%</td>
<td>94.17%</td>
<td>98.33%</td>
<td>97.14%</td>
<td>97.14%</td>
</tr>
<tr>
<td>LTV</td>
<td>100.00%</td>
<td>100.00%</td>
<td>100.00%</td>
<td>100.00%</td>
<td>100.00%</td>
</tr>
<tr>
<td>MPCD</td>
<td>100.00%</td>
<td>100.00%</td>
<td>100.00%</td>
<td>100.00%</td>
<td>100.00%</td>
</tr>
<tr>
<td>LNSCT</td>
<td>100.00%</td>
<td>100.00%</td>
<td>100.00%</td>
<td>100.00%</td>
<td>100.00%</td>
</tr>
<tr>
<td>B-NSCT</td>
<td>100.00%</td>
<td>100.00%</td>
<td>100.00%</td>
<td>100.00%</td>
<td>100.00%</td>
</tr>
</tbody>
</table>

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

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

<table>
<thead>
<tr>
<th>Method</th>
<th>Recognition rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>MSR</td>
<td>90.73%</td>
</tr>
<tr>
<td>SQI</td>
<td>91.11%</td>
</tr>
<tr>
<td>LTV</td>
<td>94.04%</td>
</tr>
<tr>
<td>MPCD</td>
<td>99.04%</td>
</tr>
<tr>
<td>LNSCT</td>
<td>99.51%</td>
</tr>
<tr>
<td>B-NSCT</td>
<td>99.82%</td>
</tr>
</tbody>
</table>

4 Conclusion

In this paper, we proposed a binary nonsampled 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.

Acknowledgements

This work was supported by the EU FP7 ICT-215554 project LIREC (LIving with Robots and intEractive Companions).

References


