

FACE RECOGNITION BASED ON NONSUBSAMPLED CONTOURLET TRANSFORM AND BLOCK-BASED KERNEL FISHER LINEAR DISCRIMINANT

Biao Wang, Weifeng Li and Qingmin Liao

Visual Information Processing Laboratory
Department of Electronic Engineering/Graduate School at Shenzhen, Tsinghua University, China

ABSTRACT

Face representation, including both feature extraction and feature selection, is the key issue for a successful face recognition system. In this paper, we propose a novel face representation scheme based on nonsubsampling contourlet transform (NSCT) and block-based kernel Fisher linear discriminant (BKFLD). NSCT is a newly developed multiresolution analysis tool and has the ability to extract both intrinsic geometrical structure and directional information in images, which implies its discriminative potential for effective feature extraction of face images. By encoding the NSCT coefficient images with the local binary pattern (LBP) operator, we could obtain a robust feature set. Furthermore, kernel Fisher linear discriminant is introduced to select the most discriminative feature sets, and the block-based scheme is incorporated to address the small sample size problem. Face recognition experiments on FERET database demonstrate the effectiveness of our proposed approach.

Index Terms— Face representation, nonsubsampling contourlet transform, kernel Fisher linear discriminant, local binary pattern.

1. INTRODUCTION

As one of the most typical applications of image analysis and computer vision, automatic face recognition has remained an extensively studied topic during the last several decades. Numerous effective approaches have been proposed to address this problem to date [1]. However, machine based algorithms are still far from surpassing the remarkable face recognition ability of human vision system, especially under uncontrolled circumstances. The main challenges arise from the fact that the interpersonal variations are relatively small, but the intrapersonal variations could be large due to variant factors such as expression, illumination, pose, aging and so on.

Face representation, which includes both feature extraction and feature selection, is acknowledged to be a key issue to face recognition technology. Generally speaking, extracting suitable feature sets could diminish the intrapersonal variations, while at the same time provide enough discriminative power for different person. Feature selection aims to reduce the dimensionality and retain the most discriminative feature sets.

Basically, feature extraction methods can be divided into three categories. The first category takes advantage of the geometric features, namely, relative positions and sizes of face components (e.g. eyes, noses and mouths). These methods have difficulty when variations of facial appearances exist, for example, closed eyes, eyes with glasses and open mouth. The second category considers a face image as a specific texture pattern and then extract the features by using

This work was supported by the Shenzhen-Hongkong Innovation Circle Project No. ZYB200907070030A.

local descriptors, such as local binary pattern (LBP) [2], local XOR pattern (LXP) [3]. However, in these methods, features are extracted directly from the gray images which are sensitive to variations due to uneven illumination. The third category tries to take advantage of the transformation features in the frequency domain or wavelet domain, such as discrete fourier transform (DFT) [4], discrete wavelet transform (DWT) [5], and Gabor wavelet [3][6]. Compared to LBP, transformation based features prove to be much more robust to variant variations. Among them, Gabor wavelet is well-known as one of the most successful feature extraction methods for face images because it can provide a multiscale and multi-orientation representation.

In this paper, we propose a novel face representation method by utilizing nonsubsampling contourlet transform (NSCT) and block-based kernel Fisher linear discriminant (BKFLD). Thanks to its multiscale, multidirection, anisotropy and shift-invariance, NSCT could not only provide multiresolution analysis, but also well capture both the geometric structure and directional information in images, which makes it suitable for feature extraction of facial images. By further encoding the NSCT coefficient images through LBP operator, we obtain a robust feature descriptor, named as local NSCT binary patterns (LNSCTBP). Furthermore, BKFLD is proposed to select the most discriminative feature sets, while at the same time avoiding the small sample size problem encountered in the traditional kernel Fisher linear discriminant (KFLD) methods. Experimental results on FERET database demonstrate the effectiveness of our proposed method.

The remainder of the paper is organized as follows. The overview of NSCT and its merits are given in Section 2. Then in Section 3, we introduce the proposed LNSCTBP descriptor, and present our feature selection scheme based on BKFLD. Experiments are given to illustrate the effectiveness of our method in Section 4. Section 5 concludes the paper.

2. NONSUBSAMPLED CONTOURLET TRANSFORM

Traditional wavelets for image processing are actually the tensor product of 1D wavelet, and they have only three directions, namely, horizontal, vertical and diagonal. Wavelets are optimal to capture point singularities. However, for high dimensional signals like images, which consist of higher order singularities, wavelets can only reveal image features across edges, but not the features along edges.

Contourlet transform (CT) was first proposed by Do and Vetterli in [7]. It's designed to be multiscale and multidirectional by utilizing laplacian pyramid (LP) and directional filter bank (DFB), which makes it suitable to discover the 2D geometry of any digital image. Compared to wavelets, CT holds improved directional elements and better ability to represent two dimensional singularities. However,

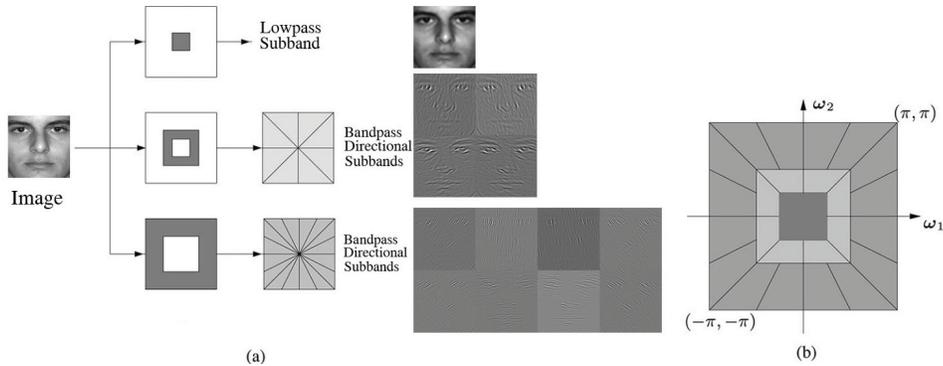


Fig. 1. Nonsampled contourlet transform: (a) NSFB structure and (b) idealized frequency partitioning obtained by NSFB.

one main disadvantage is that the contourlet transform is not shift-invariant.

To overcome the aforementioned shortcoming, Cunha *et al.* [8] proposed nonsampled contourlet transform (NSCT) by adopting nonsampled pyramid (NSP) and nonsampled filter banks (NSFB). An overview of the NSCT is displayed in Fig. 1(a), and the frequency plane in the subbands split by the filter banks is illustrated in Fig. 1(b). Note that the NSCT example illustrated in Fig. 1 consists of two scales and 4, 8 directions in the scales from coarser to finer, respectively. As can be seen, in the first scale the facial components are well represented, while in the second scale it offers a high degree of directionality and anisotropy. Due to its multiscale, multidirection, anisotropy and shift-invariance, NSCT is more efficient than other multiresolution analysis tools and has found its application in image denoising and image enhancement [8]. To our best knowledge, few work has been done to explore the potential of NSCT to solve face pattern recognition problems. However, thanks to the rich basis functions oriented at various directions at multiple scales, NSCT can effectively capture smooth contours which are the dominant features in facial images. In this paper, we investigate its potential for face feature extraction and compare its performance with Gabor wavelet experimentally.

3. PROPOSED FACE REPRESENTATION APPROACH

3.1. Local NSCT Binary Patterns (LNSCTBP)

The LBP operator was originally defined by encoding each pixel with 8 bit code, each of which is determined by thresholding the 3×3 neighborhood with the center pixel. Formally, it can be described as follows:

$$LBP(x_c, y_c) = \sum_{n=0}^7 2^n s(I_n - I_c), \quad (1)$$

in which (x_c, y_c) is the location of the center pixel, I_c and I_n are the intensity of the central pixel and its n -th neighbor, and $s(u)$ is 1 for $u \geq 0$ and 0 otherwise. LBP was first successfully applied to face recognition by Ahonen *et al.* [2]. To encode both texture and structure information for human face, the LBP map of a face image is divided into several nonoverlapping blocks and histograms computed in each block are concatenated together to form the final representation.

In [6], the authors pointed out that by applying LBP operator to further encode Gabor magnitude images with different scales and directions, one could gain robust feature sets for face images. In [3], by encoding the Gabor phase images with local XOR pattern (LXP), the

authors displayed the discriminative power of Gabor phases. These methods actually could be customized to a general feature extraction framework for face images: a face image is first decomposed by multiresolution analysis tools, and then further encoded by local descriptors. The underlying reason making it successful is two-fold: firstly, multiresolution analysis tools could capture more richer image features at multiple scales and orientations, which makes it robust to illumination and expression variations; secondly, encoding face images in the form of spatial histograms by local descriptors could avoid the side-effect of misalignment. In a word, the combination of multiresolution analysis tools and a proper type of local descriptor could result in a more robust local descriptor.

Following the aforementioned rule, we propose a novel feature extraction approach by combining NSCT and LBP. More specifically, a face image is first decomposed by NSCT and then we encode the NSCT coefficient images with LBP operator. LBP histograms computed for coefficient images of different scales and directions are concatenated together to form the final representation. The obtained feature sets are called local NSCT binary patterns (LNSCTBP). The whole encoding process is illustrated in Fig. 2.

3.2. Block-based Kernel FLD (BKFLD)

Although the proposed feature extraction approach (i.e. LNSCTBP) could be directly applied to face recognition, the obtained dimensionality of the feature sets is rather high due to the multiscale and multidirectional property of NSCT. Thus we apply feature selection technology to reduce the feature dimensionality.

Fisher linear discriminant (FLD) proves to be a successful approach for discriminative feature selection of face images. Due to its limitation of linearity, FLD fails to perform well for nonlinear problems. However, for practical applications, the subspace formed by face images is usually nonlinear due to the expression and illumination variations. Therefore, KFLD, which is a nonlinear extension of FLD with kernel trick, has been proposed [9] and could arguably provide better representation and achieve lower error rates for face recognition.

However, in high dimensional space, KFLD suffers heavily from “small sample size (SSS)” problem, which means that the sample size is much smaller than the feature dimensionality. To address this problem, we adopt the “*divide and conquer*” strategy similar to [10] and propose the BKFLD method. For a specific face image, we first extract the NSCT features, and then divide the NSCT coefficient images into several nonoverlapping blocks and extract LNSCTBP features for each block. Finally, we conduct KFLD block-wisely. Cosine distance is adopted as the similarity measure, and the similarity

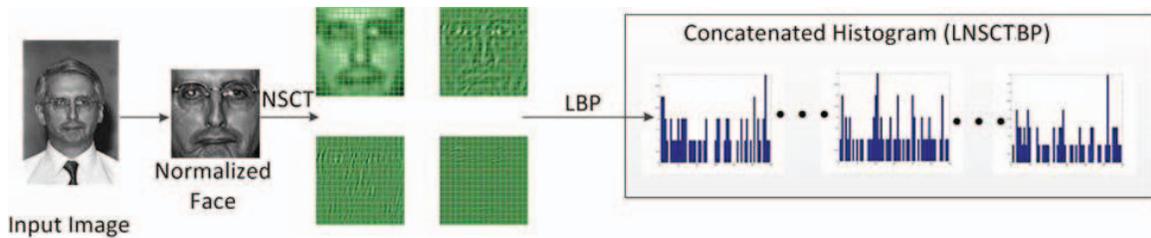


Fig. 2. Illustration of the computation of LNSCTBP.

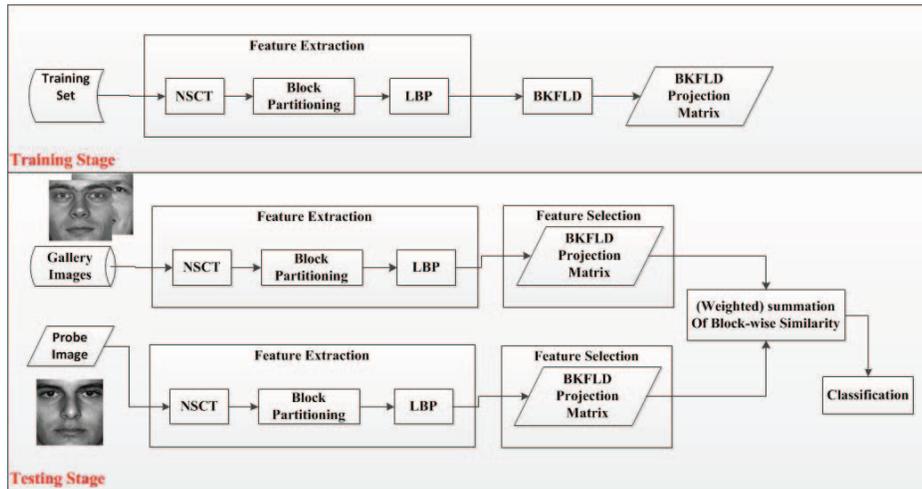


Fig. 3. Framework of the proposed face recognition approach.

of any two face images could be defined as the summation of the similarity for each block.

Moreover, much work has shown that different facial areas contribute unequally in term of recognition, which implies that the features extracted from different facial areas have different discriminative information and should be assigned with different weights. We adopt weighting scheme based on Fisher separation criteria (FSC) [3] to compute the weight for each block area. When such a weighting scheme is incorporated, the similarity of any two face images is defined as the weighted summation of the similarity for each block. The framework of our proposed face recognition approach based on LNSCTBP and BKFLD is illustrated in Fig. 3.

4. EXPERIMENTAL RESULTS

4.1. Experiment Setting

In this section, experiments are conducted on one publicly available face database, namely, FERET database to illustrate the effectiveness of our proposed method. All face images are properly aligned, cropped and resized to 128×128 with the centers of the eyes fixed at (29,34) and (99,34). No further preprocessing is performed.

We use the standard FERET protocol to conduct our experiments. The gallery set consists of 1,196 images of 1,196 subjects. There are four probe sets: Fb (different expressions with gallery, 1,195 images of 1,196 subjects), Fc (different illumination conditions with gallery, 194 images of 194 subjects), Dup I (images taken later in time, 722 images of 243 subjects), Dup II (images taken at least 18 months after the corresponding gallery, 234 images of 75 subjects). Figure 4 shows samples of the same person from the five sets.

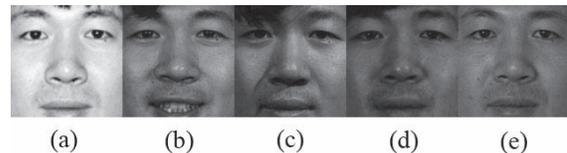


Fig. 4. Sample face images from the FERET database. (a) Fa (b) Fb (c) Fc (d) Dup I (e) Dup II.

4.2. Evaluation of LNSCTBP

In this section, we want to check whether the proposed LNSCTBP descriptor is suitable for face recognition. We compare it with three other classical LBP-based methods: Gray image plus LBP [2], Gabor Magnitude plus LBP [6] and Gabor Phase plus LBP [3].

In order to keep more spatial information, in our experiments we empirically extract the LBP histograms in 8×8 blocks with 59 bins, each corresponding to a uniform pattern. For NSCT, we use three scales and 4, 8, 8 directions in the scales from coarser to finer, respectively. The parameters of Gabor wavelet are set according to [3]. For classification, histogram intersection is adopted as the similarity measure, because it proves to be a good choice for histogram matching [2].

The result is illustrated in Table 1, from which we could see that LNSCTBP is a rather effective feature extraction approach. Compared with the method which directly applies LBP to gray images, Gabor and NSCT based methods are much more robust to illumination variations. Compared with Gabor based methods, LNSCTBP performs better when aging variations exist, which is probably due to the fact that NSCT could well capture the local features (wrinkles) that are outstanding in aged faces [11].

Table 1. Recognition rates (%) of different LBP-based methods on FERET database.

Methods	Fb	Fc	Dup I	Dup II
Gray Image + LBP	93	51	61	50
Gabor Magnitude + LBP	94	97	68	53
Gabor Phase + LBP	93	92	65	59
LNSCTBP (NSCT + LBP)	94	96	72	72

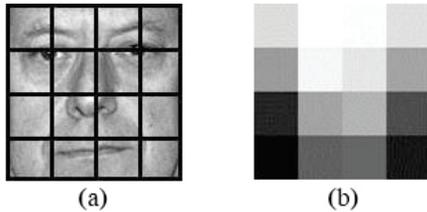


Fig. 5. (a) Partition the face image into 16 nonoverlapping blocks. (b) Visualization of FSC weights for each block. The whiter, the larger the weight is; the blacker, the smaller the weight is.

4.3. Evaluation of BKFLD

In this subsection, we take advantage of the standard training set, which consists of 1,002 frontal images of 429 subjects, to learn the KFLD projection matrix for each block. We divide each face image into 16 (4×4) nonoverlapping blocks and perform BKFLD block-wisely. We empirically retain 200 dimensions for each block, and thus the final dimensionality for a specific face image is 3,200. As for the choice of the kernel, we select Gaussian kernel, $k(x, y) = \exp(-||x - y||^2/\sigma)$, where the parameter σ is set as 8×10^5 empirically. Note that here the cosine distance is adopted as similarity measure as indicated in Section 3.2.

The partitioning scheme is illustrated in Fig. 5(a), and the corresponding weights for different block areas are visualized in Fig. 5(b). As can be seen, the eye areas play a more important role in face recognition, which agrees with the conclusions in [2] and [3].

To further demonstrate the effectiveness of our proposed method, we compare it with other state-of-the-art methods reported in literatures. The results are summarized in Table 2. As can be seen, combination of LNSCTBP and BKFLD could achieve high accuracy in all four probe sets, and the block weighting scheme could further improve the performance. Our best result is comparable to those reported in the state-of-the-arts.

Table 2. Performance comparison of proposed method with several state-of-the-art approaches on FERET database.

Methods	Fb	Fc	Dup I	Dup II
FERET97 Best [12]	96	82	59	52
LBP [2]	97	79	66	64
LGBPHS [6]	98	97	74	71
HGPP [3]	98	99	78	76
Fusion of (Gabor and LBP) [13]	98	98	90	85
LNSCTBP + BKFLD	99	99	89	79
Weighted (LNSCTBP + BKFLD)	99	99	92	86

5. CONCLUSION

This paper has exploited the effectiveness of the NSCT representation of facial images with application to face recognition. As a newly developed multiscale geometric analysis tool, NSCT can extract both intrinsic geometrical structure and directional information in digital

images, thus making it suitable for feature extraction of face images. We have designed a robust descriptor named “LNSCTB” by encoding the NSCT coefficient images with LBP operator. To further reduce the feature dimensionality, we adopted KFLD to select the most discriminative feature sets. Block-based strategy has been introduced to address the small sample size problem, and weighting scheme based on Fisher separation criteria has been incorporated to further improve the performance. Experimental results on FERET database demonstrate that our proposed approach is suitable for face recognition applications and outperforms several state-of-the-arts.

6. REFERENCES

- [1] W. Zhao, R. Chellappa, A. Rosenfeld, and P.J. Phillips, “Face recognition: A literature survey,” *ACM Computing Surveys*, pp. 399–458, 2003.
- [2] A. Hadid, T. Ahonen, and M. Pietikäinen, “Face description with local binary patterns: Application to face recognition,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28(12), pp. 2037–2041, 2006.
- [3] B. Zhang, S. Shan, X. Chen, and W. Gao, “Histogram of gabor phase patterns(hgpp): a novel object representation approach for face recognition,” *IEEE Trans. Image Process.*, vol. 16(1), pp. 57–68, 2006.
- [4] J. Lai, P. C. Yuen, and G. Feng, “Face recognition using holistic fourier invariant features,” *Pattern Recognition*, vol. 34(1), pp. 95–109, 2001.
- [5] J. T. Chien and C. C. Wu, “Discriminant waveletfaces and nearest feature classifiers for face recognition,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24(12), pp. 1644–1649, 2002.
- [6] W. Zhang, S. Shan, W. Gao, X. Chen, and H. Zhang, “Local gabor binary pattern histogram sequence (lgbphs): a novel non-statistical model for face representation and recognition,” in *ICCV*, 2005, pp. 786–791.
- [7] M. N. Do and M. Vetterli, “The contourlet transform: an efficient directional multiresolution image representation,” *IEEE Trans. Image Process.*, vol. 14(12), pp. 2091–2106, 2005.
- [8] A. L. da Cunha, J. P. Zhou, and M. N. Do, “The nonsubsampling contourlet transform: theory, design, and applications,” *IEEE Trans. Image Process.*, vol. 15(10), pp. 3089–3101, 2006.
- [9] M. Yang, “Kernel eigenfaces vs. kernel fisherfaces: Face recognition using kernel methods,” in *Int. Conf. Automatic Face and Gesture Recognition*, 2002, pp. 215–220.
- [10] P. Hsieh and P. Tung, “A novel hybrid approach based on sub-pattern technique and whitened pca for face recognition,” *Pattern Recognition*, vol. 42(5), pp. 978–984, 2009.
- [11] K. Luu, K. Seshadri, M. Savvides, T. D. Bui, and C. Y. Suen, “Contourlet appearance model for facial age estimation,” in *Proceedings of the IEEE International Joint Conference on Biometrics*, 2011.
- [12] P. J. Phillips, H. Moon, P. Rizvi, and P. Rauss, “The feret evaluation method for face recognition algorithms,” *IEEE Trans. on Patt. Analysis and Machine Intell.*, vol. 22(10), pp. 1090–1104, 2000.
- [13] X. Tan and B. Triggs, “Enhanced local texture feature sets for face recognition under difficult lighting conditions,” *IEEE Trans. Image Process.*, vol. 19(6), pp. 1635–1650, 2010.