FACE RECOGNITION WITH LOCAL CONTOURLET COMBINED PATTERNS

Yichuan Wang[†], *Shilian Yu*[†], *Weifeng Li*[†], *Longbiao Wang*[*] *and Qingmin Liao*[†]

[†]Department of Electronic Engineering/Graduate School at Shenzhen, Tsinghua University, China
[†]Shenzhen Key Laboratory of Information Science and Technology, Guangdong, China
[*]Nagaoka University of Technology, Nagaoka, Niigata 940-2188, JAPAN

ABSTRACT

This paper proposes a novel face image descriptor called local contourlet combined patterns (LCCP), based on the Non-Subsampled Contourlet Transform (NSCT), for face recognition. NSCT is a multiresolution analysis tool and can capture image information at multiple scales, orientations, and frequency bands. To adapt to the NSCT filter bank, a new encoding method named mean-based contrast patterns (MCP) is presented. We apply LBP and MCP to different levels' NSCT coefficient images respectively and then combine them to obtain a robust representation. Futhermore, block-based kernel Fisher linear discriminant (BKFLD) is used to select the most discriminative feature sets. Face recognition experiments on FERET database demonstrate the effectiveness of our proposed approach.

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

1. INTRODUCTION

Automatic face recognition has been studied as one of the most typical applications of computer vision during the last several decades. Numerous effective approaches aimed at this problem have been proposed [1]. However, most existing algorithms are still far away from surpassing human-level face recognition under uncontrolled circumstances due to the large variations occurring in face images such as expression, illumination, pose, aging and so on. These variations lead to the result that the interpersonal variations are relatively small, but the intrapersonal variations could be large.

Face representation is acknowledged to be a key issue to face recognition. Extracting suitable feature sets can minish the intrapersonal variations, and meanwhile provide enough discriminative power for different persons. Basically, feature extraction methods can be divided into three categories. The first category is the holistic feature extraction, which is usually based on the subspace learning technologies such as principal component analysis (PCA) [2], Fishers linear discriminant (FLD) [3], and so on. The second category considers a face image as a specific texture pattern and then extracts the features by using local descriptors, such as local binary pattern (LBP) [4], local XOR pattern (LXP) [5], local ternary patterns (LTP) [6]. The third category makes use of the transformation features in the frequency domain or wavelet domain, such as discrete fourier transform (DFT) [7], discrete wavelet transform (DWT) [8], and Gabor wavelet [5][9]. Compared to features in pixel-intensity domain, transformation-domain-based features have proven to be much more robust to different variations. Among them, local features in the Gabor wavelet domain provide a multiscale and multi-orientation representation and achieve promising face recognition results.

In our previous work [10], we proposed a novel face representation method LNSCTBP via the nonsubsampled contourlet transform (NSCT) and LBP. 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 of images [11]. These important characteristics of NSCT lead to an efficient representation of facial images.

In this paper, based on local binary patterns, we propose a new encoding method named mean-based contrast patterns (MCP) which is geared to the multidirectional filter bank in NSCT structure. According to their characteristics, we seperately apply the MCP and traditional LBP to different levels. Then we combine them to obtain a robust feature descriptor, named as local contourlet combined patterns (LCCP). Futhermore, we introduce the block-based kernel Fisher linear discriminant (BKFLD) used in [10] to select the most discriminative feature sets and to reduce the dimension of our proposed descriptor. Experimental results on FERET database demonstrate the effectiveness of our method, achieving competitive face recognition performance compared with the LNSCTBP method and some other state-of-the-arts.

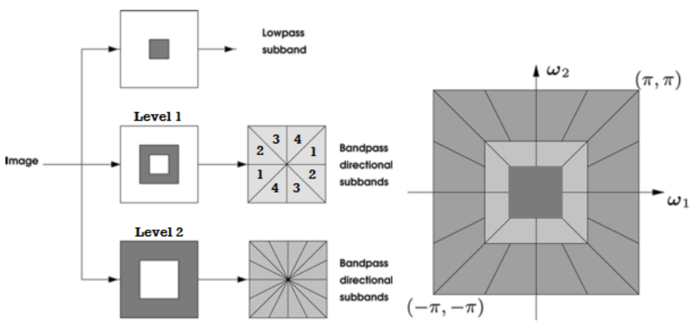
The remainder of the paper is organized as follows. The overview of NSCT is given in Section 2. Then in Section 3, we introduce the proposed encoding method MCP and LCCP descriptor. In Section 4, experiments are given to illustrate the effectiveness of our method, and some analysis on the results is stated. Section 5 concludes the paper.

978-1-4799-9988-0/16/$31.00 ©2016 IEEE

2. NONSAMPLED CONTOURLET TRANSFORM

Traditional wavelets for image processing are actually the tensor product of 1D wavelet and have only three directions, namely, horizontal, vertical and diagonal. Although wavelets are optimal to capture point singularities, 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 [12]. It utilized laplacian pyramid (LP) and directional filter bank (DFB). Compared to wavelets, CT holds improved directional elements and better ability to represent two dimensional singularities, which means that it's suitable to discover the 2D geometry of any digital image. However, being not shift-invariant is the main disadvantage of CT.



(a) (b)

Fig. 1. Nonsubsampled contourlet transform. (a) NSFB structure that implements the NSCT. (b) Idealized frequency partitioning obtained with the proposed structure.

To overcome the aforementioned shortcoming, Cunha *et al.* [11] proposed nonsubsampled contourlet transform (NSCT) by adopting nonsubsampled pyramid (NSP) and nonsubsampled 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). We can see that the NSCT example illustrated in Fig. 1 consists of two levels and 4, 8 directions in the levels from coarser to finer, respectively. Customarily the coefficient image from the lowpass subband processing is called "the lowpass image" and is not regarded as one level here.

3. PROPOSED METHOD

3.1. Review of Local Binary Patterns

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:

$$LBP(x_c, y_c) = \sum_{n=1}^{8} 2^{n-1} s(I_n - I_c) \tag{1}$$

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. With the uniformity measure "u2" [4], 8 bit code is mapped to 59 bins. LBP was first successfully applied to face recognition by Ahonen *et al.* [4]. 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.

3.2. Mean-based Contrast Patterns

Suppose that a filter bank of N directions is used in a certain level of NSCT and we will have the decomposition results $I_k, k = 1, ..., N$. Like LBP, we divide each image I_k into M blocks and denote the pixel set of the j-th block as $I_k^j, k = 1, ..., N, j = 1, ..., M$. Then we define

$$thr_k^j = \frac{1}{W_k^j} \sum_{(x,y)\in I_k^j} I_k(x, y) \tag{2}$$

where (x, y) is the location of pixels in I_k, and W_k^j is the number of pixels in I_k^j. It's obvious that thr_k^j is the intensity mean of the pixels in I_k^j. Now we can calculate the MCP of pixel (x, y) in the j-th block as follows:

$$MCP(x, y) = \sum_{k=1}^{N} 2^{k-1} s(I_k(x, y) - thr_k^j) \tag{3}$$

The format of (3) is very similar to (1). The "u2" is also applied here. We can compute a histogram for each block and concatenate them to form a representation of the level. Note that the LBP operator calculates a representation for each image while the MCP operator calculates a representation for a whole level.

3.3. Local Contourlet Combined Patterns

Firstly, the NSCT structure is used to process face images. We set the level number as 3 and divide these levels to 4, 8, 8 directions respectively, just the same as those in our previous work [10]. Secondly we use the LBP operator to encode the four NSCT coefficient images of the first level and the MCP operator to encode the second and the third level, and then obtain feature sets of images or levels. Finally, the four LBP feature vectors for level 1 and the two MCP feature vectors for level 2 and 3 are concatenated to form the final representation. Fig. 2 illustrates the framework of our proposed approach.

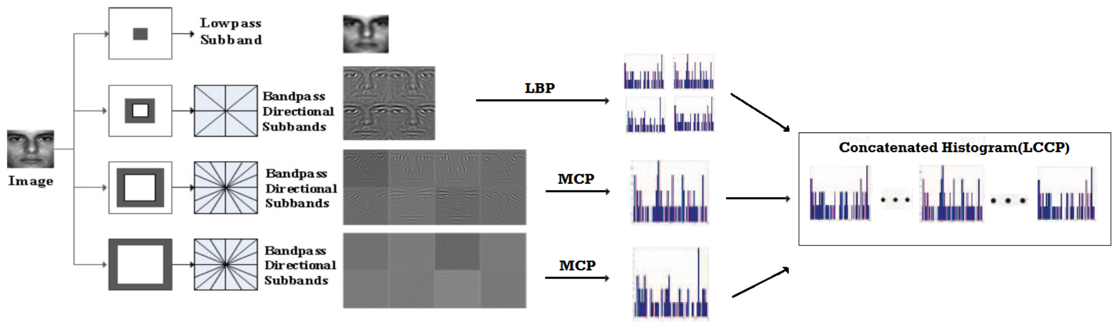


Fig. 2. Illustration of the computation of LCCP.

3.4. Block-based Kernel Fisher Linear Discriminant

Block-based kernel Fisher linear discriminant was first proposed in [10]. It conducts KFLD block-wisely to solve KFLD's "small sample size (SSS)" problem. We introduce this approach here to reduce the feature dimensionality and strengthen our feature's discriminative power.

4. EXPERIMENTAL RESULTS

4.1. Experiment Setting

Our 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, 195 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).Fig. 3 shows samples of the same person from the five sets.

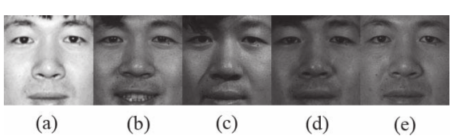


Fig. 3. Sample face images from the FERET database. (a) Fa (b) Fb (c) Fc (d) Dup I (e) Dup II.

4.2. Evaluation of LCCP

We firstly use the proposed LCCP descriptor for face recognition directly. In our experiments, each NSCT coefficient image is divided into 8×8 blocks. To check the effectiveness of our proposed MCP, we compare it with some LBP-based methods. For classification, histogram intersection is adopted as the similarity measure. The result is illustrated in Table 1.

Table 1. Recognition rates (%) of LCCP and different LBP-based methods on FERET database. Avg denotes the average rate across the four probe sets.

Methods	Fb	Fc	Dup I	Dup II	AVG
Gray Image + LBP [4]	93	51	61	50	75
Gabor Magnitude + LBP [9]	94	97	68	53	82
Gabor Phase + LBP [5]	93	92	65	59	81
LNSCTBP [10]	94	96	72	72	85
LCCP	90	98	81	80	87

Then 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. Each face image is divided into $16(4 \times 4)$ nonoverlapping blocks and BKFLD is performed block- wisely. Just the same as the BKFLD in [10], we retain 200 dimensions for each block and choose Gaussian kernel, $k(x, y) = exp(-||x - y||^2/\sigma)$, where the parameter σ is set as 8×10^5. We compare the combination of LCCP and BKFLD with other state-of-the-art methods reported in literatures. Cosine distance is adopted as the similarity measure. The result is shown in Table 2.

Table 2. Performance comparison of proposed method with several state-of-the-art approaches on FERET database. Avg denotes the average rate across the four probe sets.

Methods	Fb	Fc	Dup I	Dup II	AVG
FERET97 Best [13]	96	82	59	52	79
LBP, weighted [4]	97	79	66	64	83
LGBPHS [9]	98	97	74	71	88
HGPP [5]	98	99	78	76	90
Fusion of (Gabor and LBP) [14]	98	98	90	85	94
LNSCTBP+BKFLD [10]	99	99	89	79	94
LCCP+BKFLD	98	99	95	90	96

From Table 1 and Table 2, we can see that our proposed LCCP is a rather effective feature extraction approach. Compared with the direct LBP method and Gabor-based methods, NSCT-based methods preform better when illumination and aging variations exist. Compared with LBP-based LNSCTBP, our LCCP method improves the robustness to aging variations greatly. However, the behavior of LCCP is just OK on Fb, which means our method seems not so good at dealing with expression variations. In the following subsection, we will analyze the results and give a reasonable explain.

4.3. Analysis on Experimental Results

Our proposed MCP use block means as thresholds in encoding process. As is known to all, calculating averages usually means smooth filtering in signal processing, whose main function is to remove the high frequency information and keep the low frequency information. For a face image, the high frequency information generally refers to the small wrinkles or flecks, while the low frequency information refers to the contours of main organs such as eyes, nose and mouth. Thus the MCP operator has the ability to ignore the small texture interference and keep the basic contour information.

Different from MCP, LBP is a pixel-based encoding method and uses central pixels as thresholds. Its advantage is the ability to capture the most details of a face image, which makes it one of the most widely-used encoding operator. However, excessive attention to details is not always proper for recognition if small variations happen.

Fig. 2 displays the NSCT coefficient images in three levels. As can be seen, in the first level the coefficient images keep most basic information and the contours of the eyes, nose and mouth are still clear. The third level captures much small texture information (wrinkles). Between them, the second level still represents some contours, which makes us able to see the blurry shape of faces, meanwhile, some details also appear in the coefficient images of this level.

The aging variations mainly lead to some wrinkles in face images. These wrinkles enlarge the difference between images of the same person and become the main obstacle of recognition on the Dup sets. We know that wrinkles belong to high frequency information and will be captured by the last two levels in NSCT. Then the interference will be removed and the contour information will be reserved in the encoding step due to that LCCP uses the MCP operator to encode these levels instead of LBP. That's why our LCCP method performs much better than LBP-based methods on the Dup sets.

On the contrary to aging variations, changes in expression usually lead to difference in contours. When a man is smiling, his mouth rises with his eyes narrowed, while if he is angry, the corners of his mouth goes down and his eyes become wide. On this condition, LBP-based methods can make full use of the details rather than the contours to recognize faces correctly. But in LCCP, the MCP used in the second level keep the contour information well, which may be an advantage in other cases but a negative factor here. That's why our LCCP method performs not better than LBP-based methods, even we still use LBP in the first level.

Taken together, to respond to all kinds of conditions that may occur in real-world face images, our combination of LBP and MCP is still the best choice.

5. CONCLUSION

This paper has proposed an encoding method MCP geared to multidirectional filter bank. We combine it with LBP to process NSCT coefficient images and obtain the LCCP descriptor for face recognition. Thanks to the characteristics of NSCT and the excellent combination of MCP and LBP, our descriptor performs great robustness to condition variations, especially aging variations. To further reduce the feature dimensionality, we adopt KFLD to select the most discriminative feature sets. 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] Wenyi Zhao, Rama Chellappa, P Jonathon Phillips, and Azriel Rosenfeld, "Face recognition: A literature survey," *ACM computing surveys (CSUR)*, vol. 35, no. 4, pp. 399–458, 2003.

[2] Matthew Turk and Alex Pentland, "Eigenfaces for recognition," *Journal of cognitive neuroscience*, vol. 3, no. 1, pp. 71–86, 1991.

[3] Peter N Belhumeur, João P Hespanha, and David J Kriegman, "Eigenfaces vs. fisherfaces: Recognition using class specific linear projection," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 19, no. 7, pp. 711–720, 1997.

[4] Timo Ahonen, Abdenour Hadid, and Matti Pietikainen, "Face description with local binary patterns: Application to face recognition," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 28, no. 12, pp. 2037–2041, 2006.

[5] Baochang Zhang, Shiguang Shan, Xilin Chen, and Wen Gao, "Histogram of gabor phase patterns (hgpp): A novel object representation approach for face recognition," *Image Processing, IEEE Transactions on*, vol. 16, no. 1, pp. 57–68, 2007.

[6] Xiaoyang Tan and Bill Triggs, "Fusing gabor and lbp feature sets for kernel-based face recognition," in *Analysis and Modeling of Faces and Gestures*, pp. 235–249. Springer, 2007.

[7] Jian Huang Lai, Pong C Yuen, and Guo Can Feng, "Face recognition using holistic fourier invariant features," *Pattern Recognition*, vol. 34, no. 1, pp. 95–109, 2001.

[8] Jen-Tzung Chien and Chia-Chen Wu, "Discriminant waveletfaces and nearest feature classifiers for face recognition," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 24, no. 12, pp. 1644–1649, 2002.

[9] Wenchao Zhang, Shiguang Shan, Wen Gao, Xilin Chen, and Hongming Zhang, "Local gabor binary pattern histogram sequence (lgbphs): A novel non-statistical model for face representation and recognition," in *Computer Vision, 2005. ICCV 2005. Tenth IEEE International Conference on*. IEEE, 2005, vol. 1, pp. 786–791.

[10] Biao Wang, Weifeng Li, and Qingmin Liao, "Face recognition based on nonsubsampled contourlet transform and block-based kernel fisher linear discriminant," in *Acoustics, Speech and Signal Processing (ICASSP), 2012 IEEE International Conference on*. IEEE, 2012, pp. 1533–1536.

[11] Arthur L Da Cunha, Jianping Zhou, and Minh N Do, "The nonsubsampled contourlet transform: theory, design, and applications," *Image Processing, IEEE Transactions on*, vol. 15, no. 10, pp. 3089–3101, 2006.

[12] Minh N Do and Martin Vetterli, "The contourlet transform: an efficient directional multiresolution image representation," *Image Processing, IEEE Transactions on*, vol. 14, no. 12, pp. 2091–2106, 2005.

[13] P Jonathon Phillips, Hyeonjoon Moon, Syed Rizvi, Patrick J Rauss, et al., "The feret evaluation methodology for face-recognition algorithms," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 22, no. 10, pp. 1090–1104, 2000.

[14] Xiaoyang Tan and Bill Triggs, "Enhanced local texture feature sets for face recognition under difficult lighting conditions," *Image Processing, IEEE Transactions on*, vol. 19, no. 6, pp. 1635–1650, 2010.