Neurocomputing 267 (2017) 436–446

Contents lists available at ScienceDirect

Neurocomputing

journal homepage: www.elsevier.com/locate/neucom

Weighted contourlet binary patterns and image-based fisher linear discriminant for face recognition☆

Weifeng Li[a,b,*], Yichuan Wang[a,b], Zhen Xu[a,b], Yinyan Jiang[a,b], Zongqing Lu[a,b], Qingmin Liao[a,b]

[a] Department of Electronic Engineering/Graduate School at Shenzhen, Tsinghua University, China
[b] Shenzhen Key Laboratory of Information Science and Technology, Guangdong, China

ARTICLE INFO

Article history:
Received 28 May 2016
Revised 7 April 2017
Accepted 20 June 2017
Available online 29 June 2017

Communicated by Su-Jing Wang

Keywords:
NonSubsampled contourlet transform
Local binary patterns
Gabor wavelet
Fisher separation criterion

ABSTRACT

We propose a novel face representation model, called the weighted Contourlet binary patterns (WCBP), based on the NonSubsampled Contourlet Transform (NSCT), for face recognition. The decomposition using NSCT can capture rich image information at multiple scales, orientations, and frequency bands. This guarantees its robustness to illumination and expression variations. The weighting scheme embeds different discriminative powers of each NSCT-decomposed image. We also propose to carry out a subsequent Fisher linear Discriminant (FLD) on each decomposed image (named as WCBP+FLD) for dimension reduction of features. Our extensive experiments on the public FERET, CAS-PEAL-R1 and LFW databases demonstrate that the non-weighted Contourlet binary patterns performs better than local Gabor binary patterns. WCBP further improves the recognition rates. WCBP+FLD can achieve much competitive or even better recognition performance compared with the state-of-the-art Gabor feature based face recognition methods.

1. Introduction

Automatic face recognition (AFR) has been an active research topic during the last several decades. Considerable efforts have been devoted to the AFR tasks [1]. Although AFR in controlled environments has achieved excellent performance, most existing algorithms are still far away from surpassing human-level face recognition in real-world face images. Under uncontrolled circumstances, the large variations occurring in face images such as expression, illumination, pose, lead to the recognition difficulty because the inter-class variations are quite small while the intra-class variations could be considerably large.

AFR usually involves three crucial problems - preprocessing, face representation and classifier design. Numerous researches focus on seeking effective face representations to enhance face recognition performance. The underlying philosophy of extracting suitable features is to minish the intra-class variations, and meanwhile to provide the enough discriminative power to enlarge inter-class distances. 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. Subspace learning takes advantage of the whole face images and constructs a linear subspace using the subspace learning methods such as principal component analysis (PCA) [2], Fisher's linear discriminant (FLD) [3], independent component analysis (ICA) [4], local preserving projection (LPP) [5], kernel counterparts [6], and so on. Kan et al. [7] proposed multi-view discriminant analysis (MvDA) to learn a single discriminant common subspace for cross-view recognition. In recent years, using deep neural networks to learn effective feature representations has become popular in face recognition, such as DeepID [8] and DeepFace [9]. Through training on large-scale face images, they could learn discriminative deep face representation. Although the methods based on deep leaning have nearly achieved the best performance, the huge amount of training data restricts their application in many practical cases.

The second category is the local appearance feature extraction, which focuses on capturing useful information in local neighborhoods. Different from holistic features, local appearance features are more robust to local variations such as expression, and occlusion. These features are extracted using local descriptors, such as local binary pattern (LBP) [10], local XOR pattern (LXP) [11]. Tan

☆ This work was supported in part by the National Science Foundation of China under Grant no. 61271393.

* Corresponding author.
E-mail addresses: Li.Weifeng@sz.tsinghua.edu.cn (W. Li), wangyich14@mails.tsinghua.edu.cn (Y. Wang), xuz15@mails.tsinghua.edu.cn (Z. Xu), jiangyy_thu@163.com (Y. Jiang), luzq@sz.tsinghua.edu.cn (Z. Lu), liaoqm@tsinghua.edu.cn (Q. Liao).

http://dx.doi.org/10.1016/j.neucom.2017.06.045

et al. [12] proposed local ternary patterns (LTP) to alleviate the sensitivity to the near-uniform image regions. Dominant local binary pattern (DLBP) [13] method made use of the most frequently occurred patterns to capture descriptive textural information. Chen et al. [14] applied Weber's Law and proposed Weber local descriptor (WLD). Local derivative pattern (LDP) [15] is a generalization of LBP which aims at capturing higher-order local information. Guo et al. [16] made full use of both the signs and the magnitudes of local differences and proposed an associated completed LBP (CLBP) scheme. Three-Patch LBP (TPLBP) and Four-Patch LBP (FPLBP) were proposed by Wolf et al. [17] to encode similarities between neighboring image patches and to exploit the complementary information of the pixel-based LBP descriptor. However, features in the aforementioned methods, are extracted directly in the pixel-intensity domain and may not be able to provide robustness to variations due to uneven illumination. Therefore, some methods combining local features and subspace learning were proposed. Liu et al. [18] tried extracting random features from local image patches with random projection. Pairwise rotation invariant co-occurrence local binary pattern (PRICoLBP) [19] was proposed to capture both the spatial context co-occurrence information and rotation invariance. Lu et al. [20] proposed compact binary face descriptor (CBFD) to learn a feature mapping to project pixel difference vectors into low-dimensional binary vectors for recognition. Inspired by deep learning methods, Xi et al. [21] combined LBP with PCA and proposed LBPNet to extract dense feature.

The third category attempts to extract robust features in the transformation domain, such as the discrete Fourier transform (DFT) [22], discrete wavelet transform (DWT) [23], Gabor wavelet [11,24] and discrete shearlet transform (DST) [25,26]. 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 representations and achieve promising face recognition results. However, the time-consuming feature extraction process convolving a face image with the Gabor filter banks and the relatively high feature dimensionality restrict its application to real-time face recognition. For example, the Gabor feature extraction of a 128×128 image with the filter banks of five scales and eight orientations as typically used will result in 40 convolutions and a feature vector with the dimensionality of $40 \times 128 \times 128 = 655,360$.

Recently, applying LBP with a subsequent special histogram model to stable pixel attributes rather than the pixel intensity has been studied to seek more effective face representations. For example, Zhang et al. [24] proposed the local Gabor binary patterns (LGBP) by performing LBP on the Gabor magnitudes. Zhang et al. [11] presented a histogram of the Gabor phase pattern (HGPP) which extracts the global and local Gabor phase pattern to encode the phase information. The local Gabor XOR patterns (LGXP) is extracted by performing local XOR patterns (LXP), a variant of LBP, on Gabor phases. Xie et al. [27] demonstrated impressive results on FERET and FRGC-204 with the fusion of LGBP and LGXP. However, these methods still suffer from computational inefficiency. More recently, Vu et al. [28] extended the pixel-based self-similarity in LBP to the patch-based self-similarity, in which a central pixel is represented by the accumulated gradient magnitudes across different orientations with the help of the histogram of gradient (HOG) in a local patch centered around it. Ding et al. [29] extracted dual-cross patterns from multi-directional multi-level transformations of origin face images (MDML-DCP) for robust face recognition. Besides aforementioned methods, there have been some works on learning discriminant attributes rather than pre-defined ones. For example, Lei et al. [30] proposed to learn a set of image filters using the LDA criterion.

In our previous work [31], we proposed a novel face representation method via the nonsubsampled contourlet transform (NSCT). Different from traditional wavelets which can capture only discontinuities across edges, Contourlet can also capture the smoothness along contours. By overcoming the non-shift-invariance, NSCT is more flexible and efficient than other multi-resolution analysis tools and has been applied to image denoising and enhancement [32]. Due to its properties of being multiscale, multidirection, anisotropy and fully shift-invariant, NSCT can thereby provide better subband decomposition, which captures both the geometric details and directional information in the decomposed images. This important characteristic of NSCT leads to an efficient representation of facial images.

In this paper we further exploit the discriminative information embedded in NSCT, and propose a weighting scheme to NSCT-decomposed images for improving its face recognition performance. We call the scheme as the weighted Contourlet binary patterns (WCBP). For each NSCT-decomposed image, WCBP first encodes the local binary patterns in the NSCT magnitude and obtain a histogram map. A weighting scheme based on the Fisher Separation Criterion (FSC) is then designed to incorporate the importance of each decomposed image in terms of object recognition. The weighting scheme is necessary and effective because it assigns different weight values to each NSCT-decomposed image. These weights embed a special discriminative power at a particular scale, orientation, and frequency band. The proposed method can thus be used for the recognition task by selecting suitable NSCT-decomposed images and ignoring the others, which are redundant or even harmful for the recognition task. A single-image-based Fisher Linear Discriminant (FLD) is then employed to extract the low-dimensional discriminative features from the joint LBP histogram for each NSCT-decomposed image. Finally, we fuse the scores from the extracted low-dimensional NSCT feature maps for face classification. Extensive experimental results upon the FERET, CAS-PEAL-R1 and LFW databases verify the effectiveness of the proposed WCBP based method, achieving competitive face recognition performance compared with the Gabor-feature based methods [11,24,27].

The rest of our paper is organized as follows: Section 2 briefly reviews the Nonsubsampled Contourlet Transform. Section 3 introduces the feature extraction algorithm of the weighted contourlet binary patterns (WCBP). Section 4 proposes the scheme of the WCBP based face recognition. Section 5 presents and discusses our experimental results and finally Section 6 concludes the paper.

2. Nonsubsampled contourlet transform

Traditional wavelets decompose an image into only three orientations, namely, the horizontal, vertical and diagonal directions within each scale. Two-dimensional wavelets are only good at capturing point singularities and limited directional information but failed to catch the smoothness along contours. However, contours corresponding to smooth boundaries of objects usually contain intrinsic geometrical structures which are crucial features in visual information processing.

In order to represent images with smooth contours, Do and Vetterli [33] proposed the Contourlet Transform (CT) as an improvement of wavelets. Contourlets provide a high degree of anisotropy and directionality, which allows contourlets to represent smooth contours more sparsely. CT is able to decompose an input image into several directional sub-bands at multiple scales. CT at each scale consists of two stages-the Laplacian Pyramid (LP) and the directional filter bank (DFB). Within each scale, LP is first applied to decompose the image into two pyramidal levels. DFB is subsequently performed on the high frequency level to partition it into several directional sub-bands. The number of directions at each