Knowledge-Based Systems 84 (2015) 78–88

Contents lists available at ScienceDirect

Knowledge-Based Systems

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

Pose-invariant face recognition using facial landmarks and Weber local descriptor

Zheng Zhang [a,1], Long Wang [a,*,1], Qi Zhu [b], Shu-Kai Chen [c], Yan Chen [a,d]

[a] Shenzhen Graduate School, Harbin Institute of Technology, Shenzhen, China
[b] College of Computer Science and Technology, Nanjing University of Aeronautics and Astronautics, Nanjing 210016, China
[c] Shenzhen ZKTeco Inc., Shenzhen, China
[d] Shenzhen Sunwin Intelligent Corporation, Shenzhen, China

ARTICLE INFO

Article history:
Received 26 December 2014
Received in revised form 27 March 2015
Accepted 1 April 2015
Available online 7 April 2015

Keywords:
Face recognition
Facial landmarks
Weber local descriptors
Pose invariant
Feature extraction

ABSTRACT

Face recognition across pose is a popular issue in biometrics. Facial rotations caused by pose dramatically enlarge the intra-class variations, which considerably obstructs the performance of the face recognition algorithms. It is advisable to extract more discriminative features to overcome this difficulty. In this paper, we present a simple but efficient feature extraction method based on facial landmarks and multi-scale fusion features. We first extract local features by using Weber local descriptors (WLD) and multi-scale patches centered at predefined facial landmarks, and then construct fusion features by randomly selecting parts of local features. Finally, the classification result is obtained by decision fusion of all local features and fusion features. The proposed method has the following two characteristics: (1) local features around landmarks can well describe the similarity between two images under pose variations and simultaneously reduce redundant information and (2) fusion features constructed by randomly selecting local features from predefined regions further alleviate the influence of pose variations. Extensive experimental results on public face datasets have shown that the proposed method greatly outperforms the previous state-of-the-art algorithms.

1. Introduction

Face recognition has been studied for the few decades [1]. Most of face recognition algorithms perform well under restricted environment. However, face images captured from practical application are usually influenced by the variations of illuminations, poses and facial expressions. Therefore, extracting robust features is significant for the practical face recognition system. Generally, the feature extraction methods in face recognition could be broadly summarized into two categories, i.e., holistic methods and local methods.

The holistic methods generally extract features from a facial image by treating the image as a whole. For example, Turk and Alex [2] applied principal component analysis to calculate a group of eigenvalues and eigenvectors and form the Eigenfaces for face identification. The Fisherfaces method in [3] attempted to construct a subspace, which could maximize the between-class differences and minimize the intra-class differences. The representation based classification method has also been proved to be a powerful method in the pattern recognition field [4,5]. Especially, the sparse representation classification (SRC) [6] achieved very high accuracy in face recognition owing to its robustness and discriminative capability. However, these methods still cannot perfectly address the problem of facial pose variation and the classification capability of these holistic methods will degrade when images are captured under unrestricted environment.

The local methods usually consider several regions or sets of isolated points, from which features for classification are extracted. Local binary pattern (LBP) [7,8] and Gabor feature [9,10] are two most typical local features used in image analysis and face recognition. Elastic bunch graph matching in [11] applied Gabor wavelets to extract facial semantic components (e.g., eyes, nose, and mouth) and exploited the graph to perform image matching. Weber local descriptors (WLD) [12] is a simple but powerful local descriptor, which simulates the human visual perception. WLD has comparably computational efficiency with LBP, but it is more discriminative than LBP. Li et al. [13] applied the WLD operator to face recognition and achieved good experimental results. In his work, non-linear quantization rules of differential excitations and orientations took the place of the original operations in [12]. Additionally, some other

* Corresponding author.
E-mail address: patrick8863@gmail.com (L. Wang).
[1] These authors contributed equally to this work.

http://dx.doi.org/10.1016/j.knosys.2015.04.003
0950-7051/

effective local methods also have been proposed. Scale-invariant feature transform (SIFT) [14] as a local features descriptor has been widely used in many fields, e.g. object recognition, 3D modeling, face recognition and video tracking [15–17]. Xu et al. [18] proposed a local sparse representation method called TPTSR, which first figured out some nearest samples of the test sample, and then employed the selected samples to make final classification. Wang et. al. [19] developed an efficient image classification method by exploiting local-constrained linear coding (LLC), which projected local constraints into local-coordinate system and employed local representation result to perform the classification task. A number of learning-based descriptors, e.g. LE [20] and FRD [21], were also proposed. These methods try to learn the features that can reduce intra-class variances and enhance the inter-class variances.

Although the local methods discussed above have promising performance, they still confront with many limitations and challenges, such as the small sample size (SSS problem) [22] and face misalignment caused by pose variations. Due to the variations caused by pose variation and facial expression, the intra-class differences may be greater than inter-class differences so that the classification accuracy will drop dramatically. In such case, one intuitive solution to this challenge is to use multi-scale patches, because patches from different scales provide complementary information for classification. Li et al. [13] proposed using the multi-scale patches to increase the number of training samples for promoting classification accuracy. Literature [23] demonstrated that the large fixed patch size on each pyramid layer not only described the detailed appearance around landmarks but also captured the shape of face in relatively large scale that contained facial shape information. Zhu et al. [24] implemented the CRC method on multi-scale patches, and then combined the recognition outputs of all patches. However, the methods discussed above still have some drawbacks and limitations. The methods in literature [13,24] simply divided the face image into several overlapped or non-overlapped multi-scale patches, which might lead to semantic mismatch of corresponding patches owing to the image misalignment. In addition, considering that too many sub-images are utilized in classification, the algorithms are very time-consuming. Chen et al. [23] constructed high-dimensional features by extracting multi-scale patches based on dense facial landmarks, and achieved good performance on account of the merits of landmark features, which were very discriminative.

In this paper, a novel face image feature extraction method is proposed, which utilizes critical points based on facial landmarks and multi-scale image patches to extract pose-invariant features., which is driven by the following motivations. (1) The main challenge on pose-invariant face recognition is how to overcome the misalignment by differences between corresponding semantic facial regions under different poses. Although face alignment is important, only a few studies on it have been performed for 2D face recognition or verification. Face landmark localization [25–28] has made huge progress in recent years and becomes an important tool for face analysis. Thus, we believe that facial landmarks are beneficial to solve face misalignment problem. (2) Regions around facial landmarks contain rich facial information. Features extracted from these sub-images are more discriminative and can greatly decrease redundant information such that the algorithm is more computationally efficient and tractable. (3) Fusion features as complementary information of local features are favorable for classification. A feature vector formulation method by randomly selecting local features from predefined regions is developed to further alleviate the influence of pose variations, resulting in a simple yet effective fusion feature based feature extraction algorithm. In our work, fusion feature is constructed from local features which are extracted from different facial regions. Thus, the proposed method can obtain more discriminative and effective features for face recognition.

The rest part of our paper is organized as follows. Section 2 introduces the Weber local descriptors. Section 3 presents the three phases of the proposed method in details. The analysis of the proposed method and parameter discussion is presented in Section 4. In Section 5, we show the experimental results on some widely used databases. Finally, we give our conclusion in Section 6.

2. Weber local descriptor

In this section, the Weber local descriptor used in our method will be explicitly presented. Chen et al. proposed the basic WLD motivated by Weber's Law, which is demonstrated to be a simple but very powerful local feature descriptor. In literature [13], Li et al. proposed a further modified WLD by using a non-linear quantization method and applied it to face recognition, which achieved inspiring experimental results on some public databases. In this work, we prefer the improved non-linear quantization WLD which is more efficient and robust for face recognition. Generally, WLD includes two components, i.e. differential excitation and orientation, which construct a WLD histogram to represent the image.

2.1. Differential excitation

Differential excitation is a ratio, which concerns the differences of the center pixel against its neighborhood (i.e., 3×3 region) and the current pixel itself. It can be calculated by using

$$\begin{aligned} V_1 &= I * f_1, \\ V_2 &= I * f_2. \end{aligned} \tag{1}$$

where I represents the original image, $*$ is the convolution operator and V_i is the output of filter f_i. Fig. 1 shows some specific filters. Differential excitation G_1 is computed by combining V_1 and V_2, and then is projected into $\left[-\frac{\pi}{2}, \frac{\pi}{2}\right]$ by employing the following arctangent function

$$\alpha = arctan(G_1) = arctan(V_1/V_2). \tag{2}$$

α is further non-linearly quantized into T_1 dominant differential excitation as follows:

$$\{\xi_i\} = floor\left(\frac{\alpha + \pi/2}{\pi/T_1}\right), \quad i = 0, 1, 2, \ldots, T_1 - 1 \tag{3}$$

where $b = floor(a)$ rounds the value of a to a nearest integer, which is less than or equal to a. Eq. (3) allows the different excitations located within $[(-1/2)\pi + ((i-1)/T_1)\pi, (-1/2)\pi + (i/T_1)\pi]$ to be quantized to ξ_i.

2.2. Orientation

The orientation component of WLD is calculated by using

$$\theta = arctan(G_2) = arctan(V_3/V_4) \tag{4}$$

$$\underset{f_1}{\begin{bmatrix} 1 & 1 & 1 \\ 1 & -8 & 1 \\ 1 & 1 & 1 \end{bmatrix}} \quad \underset{f_2}{\begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}} \quad \underset{f_3}{\begin{bmatrix} 1 & 2 & 1 \\ 0 & 0 & 0 \\ -1 & -2 & -1 \end{bmatrix}} \quad \underset{f_4}{\begin{bmatrix} 1 & 0 & -1 \\ 2 & 0 & -2 \\ 1 & 0 & -1 \end{bmatrix}}$$

Fig. 1. Filters used to calculate the differential excitation and gradient orientation.