Neurocomputing 129 (2014) 41–48

Contents lists available at ScienceDirect

Neurocomputing

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

Robust recognition of face with partial variations using local features and statistical learning

Jeongin Seo, Hyeyoung Park*

Kyungpook National University, Daegu, Republic of Korea

ARTICLE INFO

Article history:
Received 1 February 2012
Received in revised form
17 June 2012
Accepted 11 September 2012
Available online 3 August 2013

Keywords:
Face recognition
Local features
Statistical learning
SIFT
Facial image variations

ABSTRACT

Despite the enormous interest in face recognition in the field of computer vision and pattern recognition, it still remains a challenge because of the diverse variations in facial images. In order to deal with variations such as illuminations, expressions, poses, and occlusions, it is important to find a discriminative feature that is robust to the variations while keeping the core information of original images. In this paper, we attempt to develop a face recognition method that is robust to partial variations through statistical learning of local features. By representing a facial image as a set of local feature descriptors such as scale-invariant feature transform (SIFT), we expect to achieve a representation robust to the variations in typical 2D images, such as illuminations and translations. By estimating the probability density of local feature descriptors observed in facial data, we expect to absorb typical variations in facial images, such as expressions and partial occlusions. In the classification stage, the estimated probability density is used to define the weighted distance measure between two images. Through computational experiments on benchmark data sets, we show that the proposed method is more robust to partial variations such as expressions and occlusions than conventional face recognition methods.

1. Introduction

In the recent decades, face recognition has been an active topic in the field of pattern recognition and machine learning [7,22]. Though there have been a number of works on face recognition, it is still a challenging topic owing to the highly nonlinear and unpredictable variations in facial images. As shown in Fig. 1, facial images of the same subject have many types of variations. While some variations such as illumination and translation are common in usual images, there are also face-specific variations such as expressions and partial occlusions. To deal with these variations efficiently, it is important to develop a robust feature extraction method that not only keeps essential information but also excludes unnecessary variational information.

Statistical feature extraction methods such as principal component analysis (PCA), linear discriminant analysis (LDA), non-negative matrix factorization (NMF), and their various extensions [6,12,17,19,20] can give efficient low-dimensional features by learning the variational properties of a data set. However, since these statistical approaches consider a sample image as a data point (i.e. a random vector) in the input space, they mainly capture the entire appearance as features, and it becomes difficult to handle local variations in image data. This property often leads to poor performance in recognizing faces with partial variations.

On the other hand, local feature extraction methods, such as Gabor filter [9], local binary pattern (LBP) [1], and SIFT [10], have been widely used for visual pattern recognition and image processing. By using local features, we can represent an image as a set of local patches and can treat the local variations more effectively. In addition, some local features such as SIFT are originally designed to be robust to image variations including scale and translations [10]. However, since most local feature extractors are previously determined at system developing stages, they cannot absorb distributional variations in a given data set. In particular, in the case of facial images, there are many types of face-specific occlusions caused by sun-glasses, scarfs, and so on, which cannot be considered in designing local features.

In this paper, we propose a robust face recognition method by introducing a statistical learning process for local features. First, we represent a facial image using SIFT, which is known to have robust property against local variations for face recognition [4,8]. Once all the given training images are represented by a set of local feature descriptors, we carry out learning process to estimate the probability distribution of the local features observed in training facial images. Through the learning process, we can obtain a probability distribution function (pdf) of SIFT features for normal facial images, and the obtained pdf can be utilized for detecting abnormal SIFT features in the newly observed test images. This kind of combination of SIFT features and its statistical learning was primarily tried in [16], and the present study is an extension of this work to develop a more general framework for the estimation of the probability density and the utilization of the estimated values in obtaining distance measures.

* Corresponding author. Tel.: +82 53 9507357.
E-mail addresses: hypark@knu.ac.kr, hyzoon@gmail.com (H. Park).

0925-2312/$ - see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.neucom.2012.09.048

Fig. 1. Variations in facial images: expression, illumination, pose, and occlusions. Images are obtained from public data [2], which are originally captured from well-known films.

Using the proposed framework, we expect to develop a face recognition system that is relatively robust to partial face-specific variations.

Recently, similar works have been carried out on the partial variations. Martinez [13] proposed a probabilistic approach that is able to compensate for imprecisely localized, partially occluded, and expression-variant faces. To deal with the partial occlusion problem, each facial image was divided into local patches and analyzed separately. Though this work effectively showed the importance of local approach in the face recognition problem, it uses a PCA-based method for extracting features from a local region. Inspired by this pioneering work on face-specific partial variations, Oh et al. [15] proposed the utilization of the local NMF (LNMF) method instead of PCA for representing local characteristics of facial images. Though the local property of LNMF was well exploited in the selection of basis based on the results of occlusion detection module, it entails the limitations that the statistical features treat an image as a whole vector. More recently, Chen et al. [3] and Min et al. [14] proposed the utilization of LBP features for handling occlusion variations. Through a separate training process for detecting the occlusion region by using support vector machine (SVM), LBP features in an occluded region can be detected and excluded from the recognition stage. However, the training module for detecting occlusion needs training data that includes occlusions, and it is difficult to achieve robustness to various types of partial variations that have not been shown in the learning stage.

While these related works [3,14,15] mainly focus on partial occlusion and involve a specifically designed module of detecting and excluding occlusions, our previous work [16] suggests a similarity measure that combines the distance between features and the weights of features corresponding to the appearance of abnormal partial distortions. Similarly, Zhang et al. [21] proposed a combined similarity measure using local Gabor binary pattern (LGBP) features and the possibility of occlusions. The proposed methods can be considered as a general framework for evaluating the importance of local features and using it to define a new similarity measure that is robust against partial variations.

2. Representation of facial images using SIFT

In order to represent a facial image using local features, we use SIFT feature [10], which is one of the most popular local features for complex visual pattern recognition owing to its robust properties to scale and translations. In this section, we briefly describe SIFT feature extraction for face recognition. There are two main stages of computation to obtain a set of SIFT features for an image. First, we need to determine how to select an interesting point from a whole image. We call the selected interesting pixel *keypoint*. Second, we need to calculate the descriptor for the selected keypoints so that it can represent the meaningful properties of the corresponding local patches. We call it *keypoint descriptor*, and each image is represented by a set of keypoints with descriptors.

SIFT uses scale-space difference-of-Gaussian (DOG) to detect keypoints in images. For an input image, $I(x, y)$, the scale space is defined as a function, $L(x, y, \sigma)$ produced from the convolution of a variable-scale Gaussian $G(x, y, \sigma)$ with the input image. The DOG function is defined as follows:

$$\begin{aligned} D(x,y,\sigma) &= (G(x,y,k\sigma)-G(x,y,\sigma))*I(x,y) \\ &= L(x,y,k\sigma)-L(x,y,\sigma) \end{aligned} \tag{1}$$

where k represents a multiplicative factor. The local maxima and minima of $D(x, y, \sigma)$ are computed on the basis of its eight neighbors in the current image and nine neighbors in the scale above and below.

In the original work on SIFT, a number of keypoints are selected on the basis of the value of keypoint descriptors, and thus the number of keypoints and location depends on each image. In the case of face recognition, however, the original work has a limitation in that only a few number of keypoints are extracted due to the lack of textures of facial images. To solve this problem, Dreuw [5] proposed the selection of keypoints at regular image grid points to obtain a dense description of the image content, which is usually called dense SIFT (DSIFT).

Each keypoint extracted by the SIFT method is represented as a descriptor that is a 128 dimensional vector $\boldsymbol{\kappa}$ composed of four parts: locus (location in which the feature has been selected), scale (σ), orientation, and magnitude of gradient. The magnitude of gradient $m(x, y)$ and the orientation $\Theta(x, y)$ at each keypoint located at (x, y) are computed as follows:

$$m(x,y) = \sqrt{(L(x+1,y)-L(x-1,y))^2 + (L(x,y+1)-L(x,y-1))^2} \tag{2}$$