Optik 125 (2014) 5017–5024

Contents lists available at ScienceDirect

Optik

journal homepage: www.elsevier.de/ijleo

Producing virtual face images for single sample face recognition

Tao Zhang [a,*], Xianfeng Li [a], Rong-Zuo Guo [b]

[a] Engineering Lab on Intelligent Perception for Internet of Things (ELIP), School of Electronic and Computer Engineering, Peking University, Shenzhen, China
[b] Department of Computer Science, Sichuan Normal University, Chengdu 610068, China

ARTICLE INFO

Article history:
Received 3 August 2013
Accepted 28 January 2014

Keywords:
Single sample
Face recognition
Symmetrical face
Linear combination

ABSTRACT

For single sample face recognition, there are limited training samples, so the traditional face recognition methods are not applicable to this problem. In this paper we propose to combine two methods to produce virtual face images for single sample face recognition. We firstly use a symmetry transform to produce symmetrical face images. We secondly use the linear combination of two samples to generate virtual samples. As a result, we convert the special single sample problem into a non-single sample problem. We then use the 2DPCA method to extract features from the samples and use the nearest neighbor classifier to perform classification. Experimental results show that the proposed method can effectively improve the recognition rate of single sample face recognition.

1. Introduction

Face recognition is one of the most challenging branches of biometrics recognition [1]. Due to the fact that face recognition are influenced by light, gesture, expression, etc., the existing processing methods of face recognition, such as linear feature extraction methods [2–6], nonlinear feature extraction methods [7–15] usually require a sufficient number of representative training samples for obtaining good results. But in some special occasions, the available samples are limited. For example, the id card, student card, passport, diploma, admission ticket, or employee's card and so on can provide only a single photo, so we can only use this picture to perform training to achieve the goal of face recognition, which is known as single sample face recognition. The research of single sample face recognition is very meaningful, because single sample face recognition can effectively reduce the training samples collection costs, storage costs, and accelerate the processing speed of face recognition system. But single sample face recognition not only has the interference factors such as illumination, posture, facial expression, but also has the following problems. First, face recognition is a typical small sample problem, and the single sample problem is the extreme situation. Second, a single training sample does not have enough representative information on the face. Third, the scattering matrix of intra-class for single sample does not exist, so some classic methods, such as the LDA method [16] cannot be implemented. Moreover, as the intra-class distribution cannot be estimated out, the probability based methods are also not applicable. Fourth, the inter-class variation is overestimated in single sample problem. The inter-class variations measure the differences between images that have different class labels. As there is only one image per person in single sample problem, all the variations are inter-class variations. Though feature extraction methods can try their best to maximize the inter-class variations, they cannot perform well in single sample problem.

Due to the severe challenging of single sample face recognition problem and itself important meanings, it has become one of the most active branches of face recognition. In recent years, many researchers put forward many important methods [17]. In literature [18,19] $(PC)^2A$ method is put forward, which uses the original image and integral projection to form a new image, and then reuses the PCA for identification. In literature [20] sparse PCA (SPCA) method is proposed, which uses the singular value decomposition of original images to reconstruct and to enhance the image, performing special pretreatment of face samples. In literature [21] the authors put forward a method based on the three-layer virtual image generation, using the singular value disturbance prominent characteristics. Literature [22] proposed a method based on sparse representation, first exploiting the transformation process for the single sample, then using the sparse representation (SR) for classification. In literature [23] uniform pursuit (UP) approach is proposed to utilize whitening transform and locality dispersing projection to uniform the pair wise distance of prototypes in PCA space. Deng et al. [24] proposed Extend SRC(ESRC) method, takes advantage of the pair wise difference images from a generic database to construct the intra-class variation dictionary, which is further incorporated into the framework of SRC to cover the variations between gallery and probe samples. In literature [25] the author proposed a local

* Corresponding author.
E-mail addresses: zhtao123abc@163.com (T. Zhang), lixianfeng@pkusz.edu.cn (X. Li), rongzuoguo@126.com (R.-Z. Guo).

http://dx.doi.org/10.1016/j.ijleo.2014.01.171
0030-4026/

probabilistic approach, where the subspace of each individual is learned and represented by a separate Gaussian distribution. In literature [26] the authors proposed a method by separating texture and shape information and projecting them into separate PCA spaces. After the treatment, they construct separate eigenspaces for texture and the invariant shape features. However, the most of the methods have been proposed in the literatures did not make full use of inter-class information of samples. Due to the particularity of the single sample problem, some classic face recognition algorithms, such as principal component analysis (PCA) [27] and local preserve projection (LPP) [28] cannot obtain good effect. In single sample face recognition problem, if can find a certain way to increase the number of training samples and convert the special single sample problem into a general face recognition issue, we might obtain good results.

2. Introduction of related methods

To convert special problems into general issues is a kind of very effective method to solve the problem. In order to convert single sample face recognition, the extreme small sample problem, into a general face recognition issue, literature [29,30] put forward some effective methods of generating the virtual samples as follows.

2.1. Generation of symmetry face samples

In this subsection we show how to use every original training sample to generate two symmetrical face training samples. Let $x_i \in R^{p\times q}$ represents the i-th training samples in the form of image matrix. Let ${y_i}^1$ and ${y_i}^2$ respectively stand for the first and second symmetrical face training samples generated form x_i. The details of generating symmetry face are as follows. The left half columns of ${y_i}^1$ is set to the same as that of x_i and the right half columns of ${y_i}^1$ is the mirror image of the left half columns of ${y_i}^1$. However, the right half columns of ${y_i}^2$ is set to the same as that of x_i and the left half columns of ${y_i}^2$ is the mirror image of right half columns of ${y_i}^2$. The mirror image S of an arbitrary image R is defined as $S(i, j) = R(i, V - j + 1)$, $i = 1, \ldots, U$, $j = 1, \ldots, V$. U and V stand for the numbers of the rows and columns of R, respectively. $S(i,j)$ denotes the pixel located in the i-th row and j-th columns of S. Thus, each single training sample is produced two symmetry face of virtual sample ${y_i}^1$ and ${y_i}^2$. As shown in Fig. 1, we randomly selected some original samples from the ORL face database to generate the symmetry face of virtual samples.

2.2. The linear combination of inter-class samples

Each sample image can be seen as a point in the high-dimensional face space. Due to the changes of pose, illumination and expression, the same person's face images are different, so we use different points to denote them respectively. Since these images are from the same person, so they have something in common. We will classify them as the same class. Assume image x and y are taken from class 1 and class 2 respectively, regard them as two points in face space. We can perform a linear fitting. The fitting formula is as follow:

$$z = \lambda x + (1 - \lambda)y, \quad 0 \le \lambda \le 1 \tag{1}$$

Note that not every fitting out image is real image, the middle part of the linear fitting out images are apparently different from original images, but the images on both ends of linear fitting are very similar to original images, so the selection of parameters is very important.

In Fig. 2, we randomly choose two face images from the ORL face database and synthesize nine novel images using (1) by setting the parameter λ = 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9. As can be seen from Fig. 2, the synthesized images are quite similar to y when the parameter is 0.1 or 0.2, and these are similar to x when the parameter is 0.8 or 0.9. Thus, we can take some of synthesized face images as variations of the real face images, and use them to enlarge the training dataset.

By using formula (1) to synthesize different classes x and y to enlarge the number of training samples, we need to set the parameter value scope. We confines this parameter into the union of two sets $s_1 = [0, 1/3)$ and $s_2 = (2/3, 1]$. If $\lambda \in s_1$, formula (1) synthesizes a variation for y. If $\lambda \in s_2$, formula (1) synthesizes a variation for x. If $\lambda = 0$ or 1, formula (1) synthesizes a variation for y or x original image. In the set consists of the original images and the ones synthesized using (1), we can prove that the intra-class variation is smaller than the inter-class variation in terms of Euclidean distance.

Suppose two images z_1 and z_2 are synthesized using (1) respectively corresponding to parameter λ_1 and λ_2, as follows

$$z_1 = \lambda_1 x + (1 - \lambda_1)y \tag{2}$$

$$z_2 = \lambda_2 x + (1 - \lambda_2)y \tag{3}$$

The distance between them can be computed.

$$\begin{aligned} d^2(z_1, z_2) &= d^2(\lambda_1 x + (1 - \lambda_1)y, \lambda_2 x + (1 - \lambda_2)y) \\ &= (\lambda_1 - \lambda_2)^2 (x - y)^T (x - y) = (\lambda_1 - \lambda_2)^2 d^2(x - y) \end{aligned} \tag{4}$$

Analysis: If z_1 and z_2 are synthesized images for the same image y(or x), both λ_1 and λ_2 are from the same set s_1(or s_2). In this set s_1(or s_2), the difference between these two parameters is smaller than 1/3. Thus,

$$d^2(z_1, z_2) = (\lambda_1 - \lambda_2)^2 d^2(x, y) < \frac{1}{9} d^2(x, y) \tag{5}$$

If z_1 and z_2 are synthesized images for two different images, λ_1 and λ_2 are from two different sets s_1 and s_2. Thus, the difference between these two parameters is larger than 1/3. Thus,

$$d^2(z_1, z_2) = (\lambda_1 - \lambda_2)^2 d^2(x, y) > \frac{1}{9} d^2(x, y) \tag{6}$$

Based on the analysis, we know that all the intra-class variations are smaller than $(1/3)d^2(x, y)$ and all the inter-class variations are larger than $(1/3)d^2(x, y)$. Thus, the intra-class variations are smaller than the inter-class variations. According to this we can conduct identification and classification.

2.3. Two-dimensional principal component analysis (2DPCA)

Principal component analysis (PCA) is a method of data analysis which was proposed by K. Pearson in more than a century ago. It is a mathematical procedure that uses an orthogonal transformation to convert a set of observations of possibly correlated variables into a set of values of linearly uncorrelated variables [31]. But the traditional PCA methods [32–43] need to transform the image into a one-dimensional vector. However, the dimension of converted data is very high, the computational complexity is also increased a lot. Therefore, Yang et al. [44] proposed two-dimensional principal component analysis (2DPCA) method, which uses original image to construct covariance matrix directly, do not need to convert image matrix into a one-dimensional vector.

According to the method presented in reference [44], let X denote an n-dimensional unitary column vector. Regard image A as a $m \times n$ random matrix, project A onto X by the following linear transformation [45,46],

$$Y = AX \tag{7}$$

Thus, we obtain an m-dimensional projected vector Y, which is called the projected feature vector of image A. The total scattering