Digital Signal Processing ••• (••••) •••–•••

Contents lists available at ScienceDirect

Digital Signal Processing

www.elsevier.com/locate/dsp

Singular value decomposition based sample diversity and adaptive weighted fusion for face recognition

Guiying Zhang [a,b], Wenbin Zou [c], Xianjie Zhang [b], Xuefeng Hu [d], Yong Zhao [d,*]

[a] College of Computer Science and Technology, Guizhou University, Guiyang, China
[b] Department of Medical Information Engineering, Zunyi Medical University, Zunyi, China
[c] Shenzhen Key Lab of Advanced Telecommunication and Information Processing, College of Information Engineering, Shenzhen University, Shenzhen, China
[d] The Key Lab of Integrated Microsystems, Peking University Shenzhen Graduate School, Shenzhen, China

ARTICLE INFO

Article history:
Available online xxxx

Keywords:
Adaptive weight fusion
Virtual sample
Singular value decomposition (SVD)
Face recognition
Collaborative representation

ABSTRACT

The performance and robustness of face recognition are largely determined by the data samples used for model training. To obtain more representative samples of a face, this paper proposes a novel approach to acquire two groups of virtual samples from the right singular vectors as well as from left singular vectors via singular value decomposition (SVD) for each class of training samples. The generated virtual images not only enrich training samples but also obtain more representative information of faces, therefore higher accuracy of face recognition is achieved. Furthermore, we propose a simple and effective method that automatically determines adaptive weight without any manual intervention for three groups of scores, including the original samples and two groups of virtual samples. The weighted score fusion scheme is able to offer more supplementary information from multiple sources and obtain better performance in face recognition. Experiments on three benchmark datasets demonstrate that our proposed method is robust and obtains better accuracy for face recognition compared with previous methods.

1. Introduction

Face recognition is an active research topic in the fields of pattern recognition, artificial intelligence and digital image processing owing to its wide range of applications and important theoretical research value. In the past several years, various methods for face recognition have been proposed. For example, principal component analysis (PCA) [1–4] is an excellent method for face recognition, and Fisher discriminant analysis (FDA) [5–7] is also the most typical representative algorithms for face recognition. When sufficient training samples are provided for modeling, the state-of-the-art face recognition approaches, e.g. linear regression based classification (LRC), are able to achieve good performance. Unfortunately, the number of face images is limited in some real face recognition applications and lead to obvious performance decrease. For example, by using only a few training samples per class for modeling, sparse representation based classification (SRC) [9,10] and collaborative representation based classification (CRC) [11], both of them play an important role in face recognition [12–17], usually generate unreliable representative residual error and thus have relatively weak performance for face recognition. Besides these methods, other face recognition method such as [18–21], demands enough training samples as well.

To improve the robustness of face recognition, many methods have been proposed to incorporate existing methods. Among them, producing virtual samples from original samples to expand the size of the set of the training samples is an intuitive and effective approach. The methods for generating virtual samples can be categorized into three groups. The first kind of virtual samples are produced via symmetrical characteristic of faces. For instance, Xu et al. [22,23] used symmetrical faces and mirror images of original images as virtual images for face recognition. Wu et al. [24] and Song et al. [25] also applied the symmetrical face to improve the performance of face recognition. The second kind of virtual samples are generated by exploiting the variation of illuminations, poses and expressions. For example, Boom et al. [26] proposed the virtual illumination grid (VIG) approach to model unknown illumination conditions. Abdolali et al. [27], Ho and Chellappa [28] proposed similar methods as well to expand the training set. The third kind of virtual samples are obtained based on common structural characteristic of face samples. Ryu et al. [29] exploited the distribution of the training samples to produce virtual training samples of the face. Tang et al. [30] proposed a method based on prototype faces, an optic flow and expression ratio image to generate

* Corresponding author.
E-mail address: yongzhao@pkusz.edu.cn (Y. Zhao).

http://dx.doi.org/10.1016/j.dsp.2016.11.004
1051-2004/

more expressive facial expression, which are used to create virtual face images to enlarge the training sample set. Liu et al. [31,32] integrated the original and approximate face images to perform collaborative representation based classification, and fused synthesized training samples to perform face recognition, respectively. Besides these methods, some simple methods can also obtain reasonable virtual samples. Tang et al. [33] generated virtual samples by adding random noise to training samples. Xu et al. [34] expanded the training set by linearly combining the training samples and then selected the N nearest training samples to reduce the uncertain of face image data. In existing methods, score fusion is often ignored. Usually, scores fusion can enable more information from multiple sources to be exploited and allows higher accuracy to be achieved.

In the pattern recognition field, singular value decomposition (SVD) has attracted much interest of researchers. In early studies, singular values obtained using SVD was used as features [35–37]. However, the corresponding methods do not bring good recognition results. Probably the main reason is that singular values can stand for importance of the associated singular vectors but are not the dominant information of the data. The most information of the data is contained in the singular vectors. Owing to this reason, later studies almost pay attention to proper exploiting of singular vectors [38,39]. With the same reason, we also exploit them to design our method.

In this paper, we apply the SVD to produce virtual images of original images to enlarge the training samples set size and to improve the performance of classification. In our method, we first decompose the training sample matrix of a class into U, D and V, and use the obtained V only to construct a matrix. We refer to it as the first kind of virtual sample. We second decompose the training sample matrix of a class into U, D and V, and use the obtained U only to construct a matrix named as the second kind of virtual sample. It enables available information of original non-square sample matrix to be captured and conveyed by virtual images. Our proposed method can directly work for the sample matrix and extract information of original samples.

Fusion can offer more supplementary information from multiple scores and obtain better performance in image classification. Conventional scores fusion methods generally fall into three categories, i.e. transformation score fusion, classifier based scores fusion and density based fusion [40,41]. Other score fusion methods such as the receiver operating characteristic (ROC) based score fusion method is also feasible. This kind of methods includes the optimizing approach of area under the curve (AUC) [42], the least square error based framework [43], and the margin based ranking [44].

The existing methods showed that the weighted score fusion approach achieved better accuracy owing to more information obtained from different sources [45,46]. However, it is hard to determine optimal weights for the weighted score fusion, and the weight selection often depend on empirical knowledge. Some efforts have been made to improve weighted scores [47–49], but most of this kind of approach also relies on experience and cannot completely implement automatic determination of adaptive weight. In order to address this issue, Xu and Lu [50] designed a perfect adaptive weighted fusion approach, which automatically determines optimal weights without any manual setting. In this paper, we propose a very simple adaptive method based on [50] to automatically set up the weight of three kinds of sources, including the original samples and two groups of virtual samples.

In summary, the contributions of this paper are as follows:

First, we apply SVD to produce the right singular vectors and left singular vectors, then convert them into two kinds of virtual samples, which provide stable and supplement information of images and allow higher accuracy to be achieved.

Second, we integrate the original samples and two groups of virtual samples which allow more information of the same class object to be available.

Third, we propose a very simple adaptive method to automatically set up the weight of three groups of sources, including the original samples and two groups of virtual samples.

Last but not the least, the extensive experiments show that the proposed approach outperforms previous methods.

The remainder of this paper is organized as follows. Section 2 describes related works of this paper. Section 3 presents the proposed two kinds of virtual samples and the adaptive weight fusion approach. Section 4 discusses the experimental results. Finally, this paper is concluded in Section 5.

2. Related works

In this section we make a brief review of typical score fusion methods and collaborative representation based classification (CRC) for face recognition.

2.1. Score fusion

Chen et al. [51] proposed a novel method for synthesizing visual light images from near infrared images based on learning the mappings between images of different spectra. This approach can reduce the inter-spectral differences significantly and allow effective matching between face images taken under different imaging conditions. In [52], the authors presented a deep face system for unconstrained face recognition. Sun et al. [53] proposed a novel deep learning algorithm which can be well generalized to new classes and the verification task. In [54], Lei et al. proposed a method to learn a discriminant face descriptor (DFD) in a data-driven way, which learns the most discriminant local features by minimizing the difference of the features between images of the same person as well as maximizing that between images from different people. Zhang et al. [55] proposed the Pose Invariant PErson Recognition (PIPER) method to recognize face. The authors proposed a method named Riesz Binary Pattern (RBP) for face recognition in [56].

2.2. Collaborative Representation Based Classification (CRC)

We review the algorithm of CRC in this section. We suppose that there are c different classes, and each class has n_i training samples from i-th class, $i = 1, 2, \ldots, c$. Let $x_i^j \in \Re^{h*w}$ denote i-th class and j-th sample, where $i = 1, 2, \ldots, c$, $j = 1, 2, \ldots, n_i$, h and w represent the number of row and column of the sample, respectively. Suppose y is a test sample, $X = [X_1, X_2, \ldots, X_c]$ is the training set, where $X_i = [x_i^1, x_i^2, \ldots, x_i^{n_i}]$, $i = 1, 2, \ldots, c$.

The conventional CRC method is formulated by minimizing the following objective function

$$\|y - X\mu\|_2^2 + \lambda\|\mu\|_2^2, \tag{1}$$

where λ is a small positive constant. The solution of (1) is obtained by using

$$\mu = \left(X^T X + \lambda I\right)^{-1} X^T y, \tag{2}$$

here μ denotes the representation coefficient of all training samples.

Based on (2), the matching score of test sample y with regard to the i-th class can be calculated as follows

$$s_i = \left\| y - \sum_{k=1}^{n_i} \mu_i^k x_i^k \right\|_2^2, \quad i = 1, 2, \ldots, c, \tag{3}$$