Expert Systems With Applications 45 (2016) 352–358

Contents lists available at ScienceDirect

Expert Systems With Applications

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

Sample pair based sparse representation classification for face recognition

Hongzhi Zhang [a,*], Faqiang Wang [a], Yan Chen [c], Weidong Zhang [b], Kuanquan Wang [a], Jingdong Liu [d]

[a] School of Computer Science and Technology, Harbin Institute of Technology, Harbin, China
[b] Shanghai Jiao Tong University, Shanghai, China
[c] College of Information and Computer Engineering, Northeast Forestry University, Harbin, China
[d] Harbin Vicog Intelligent Systems Co., Ltd, Harbin, China

ARTICLE INFO

Keywords:
Sparse representation classification
Pattern recognition
Face recognition

ABSTRACT

Sparse representation classification, as one of the state-of-the-art classification methods, has been widely studied and successfully applied in face recognition since it was proposed by Wright et al. In this study, we proposed a method to generate virtual available facial images and modified the well-known linear regression classification (LRC) and collaborative representation based classification (CRC) for face recognition. The new method integrates the original and virtual symmetry facial images to form a training sample set of large size. Experimental results show that the proposed method can achieve better performance than most of the competitive face recognition methods, e.g. LRC, CRC, INNC, SRC, RCR, RRC and the method in Xu et al. (2014). This promising performance is mainly attributed to the fact that the sample combination scheme used in the new method can exploit limited original training samples to produce a large number of available training samples and to convey sufficient variations of the original training samples.

1. Introduction

Face recognition is an important task in computer vision, and has been widely used in authentication applications (Wei, Jian-qi, & Xiang, 2011). There are many methods proposed for face recognition in the past decades (Gumus, Kilic, Sertbas, & Ucan, 2010; Naseem, Togneri, & Bennamoun, 2010; Sun, Chen, Lo, & Tien, 2007; Wright, Yang, Ganesh, Sastry, & Ma, 2009; Zhang, Yang, & Feng, 2011). The subspace methods, such as eigenfaces (Jolliffe, 1986; Turk & Pentland, 1991) and Fisherfaces (Belhumeur, Hespanha, & Kriegman, 1997), have been widely applied in face recognition. These methods aim to learn a linear projection matrix to project the high-dimensional face samples into a low-dimensional subspace. Apart from subspace methods, many neural networks based approaches have been proposed for face recognition (Lawrence, Giles, Tsoi, & Back, 1997; Lin, Kung, & Lin, 1997; Sun, Chen, Wang, & Tang, 2014; Sun, Wang, & Tang, 2014). However, the training process of these methods are time consuming. Also, these methods are likely to result in overfitting when the training set is small. Some local descriptors, e.g. Gabor (Liu & Wechsler, 2002), local binary patterns (LBP) (Ahonen, Hadid, & Pietikainen, 2006), SIFT (Lowe, 2004) and histogram of oriented gradients (HOG) (Dalal & Triggs, 2005), have also been proposed to extract discriminative feature for face recognition.

To increase the robustness toward the occlusion and corruption of face images, sparse representation classification (SRC) based face recognition method has been proposed by Wright, Yang et al. (2009). Sparse representation was first proposed in the community of signal processing (Huang & Aviyente, 2006). Since it was introduced by Wright et al. to the community of pattern recognition and computer vision as a representation and classification method (Wright, Ma et al., 2009; Wright, Yang et al., 2009), it has been applied to various problems such as face recognition, super-resolution reconstruction (Yang, Wright, Huang, & Ma, 2008), image alignment (Wagner et al., 2012) and image de-noising (Dong, Li, Zhang, & Shi, 2011). Among the applications of SRC, its performance in face recognition is very good.

SRC can be viewed as a method that identifies the testing sample to the class with the strongest representation ability of this sample. SRC first constructs the relation between the testing sample and all the training samples from the viewpoint of constraint regression analysis, then tries to represent the testing sample by using a linear combination of all the training samples and usually use a method

* Corresponding author at: Research Center of Computational Perception and Cognition, School of Computer Science and Technology, Harbin Institute of Technology, Harbin 150001, China. Tel.: +86 451 86412871.

E-mail addresses: zhanghz@hit.edu.cn, zhanghz0451@gmail.com (H. Zhang), tshfqw@163.com (F. Wang), cheny0451@126.com (Y. Chen), wdzhang@sjtu.edu.cn (W. Zhang), wangkq@hit.edu.cn (K. Wang), musicrainie@163.com (J. Liu).

http://dx.doi.org/10.1016/j.eswa.2015.09.058
0957-4174/

similar to regression analysis to obtain the coefficients of the linear combination. It then calculates the deviation between the testing sample and the weighted sum of the training samples of each class (the coefficient is used as the weight) and ultimately assign the testing sample to the class with the minimum deviation (Xu et al., 2013; Xu, Zhang, Yang, & Yang, 2011).

Many face recognition methods have been developed based on the idea of SRC. Among these methods, linear regression classification (LRC) (Naseem et al., 2010; Naseem, Togneri, & Bennamoun, 2012) algorithm is quite simple and is very easy to be implemented. It has achieved good performance in face recognition. Because of this, it has received increased attention. Collaborative representation classification (CRC) is another notable method with a simple algorithm (Shi, Eriksson, Hengel, & Shen, 2011; Zhang et al., 2011). Both LRC and CRC uses the l_2-norm sparsity to formulate the optimization problems, because they can be solved much faster than the l_1-norm based problem, and l_2-norm sparsity can achieve no worse results than l_1-norm sparsity (Zhang et al., 2011). CRC and LRC not only perform well in face recognition but can also obtain analytical solutions by solving linear equations. Compared with CRC and LRC, the other similar methods usually need to design an elaborate alternative algorithm to obtain their solution, because the other similar methods are usually constructed on the basis of the l_1-norm minimization (Yang, Zhang, Xu, & Yang, 2012).

It should be pointed out that face recognition is a challenging problem and the main challenge is that the facial image is a deformable object. For example, facial images of the same class may be very different owing to various poses, illuminations and facial expressions (Yang, Sun, & Zhang, 2011). This problem limits the potential of SRC and its similar methods. In order to overcome this problem, a method to generate virtual facial images to represent possible variations of the original facial image was used. Previous studies showed that virtual facial images can reflect possible variations in poses and illuminations of the original facial image (Xu et al., 2013). The performance of SRC may be improved by exploiting virtual facial images. For example, Xu et al. proposed a method to exploit the mirror image and symmetrical facial image of the original facial image to improve SRC and obtained a very promising performance (Xu, Li, Yang, & Zhang, 2014). This method was based on the fact that the face usually has a symmetrical structure and the original facial image usually does not. As a result, Xu et al. used the original facial image to generate the symmetrical facial image and exploited both of them to perform face recognition. The symmetrical structure of the face has also been applied to face detection (Saber & Tekalp, 1998; Saha & Bandyopadhyay, 2007; Su & Chou, 1999).

In this study, we proposed a novel method for face recognition. This method can be viewed as an improvement of the method in Xu et al. (2014). It also combines the original facial image and the virtual facial images for training and test. However, when there are few training samples in each class, the scale of training set of Xu et al. (2014) is not large enough. Here we combine two samples from the same class to form a sample pair. Each sample pair may include two original facial images, two virtual facial images or one original and one virtual facial image. By this strategy, we can enlarge the scale of training instances from $O(cn)$ samples to $O(cn^2)$ sample pairs, where c is the number of classes, and n is the number of samples in each class. The difference of the proposed method and Xu et al. (2014) is that the method in Xu et al. (2014) constructs the training and testing set by each single original and virtual facial image, while the proposed method form the sample pairs based on the original and virtual facial images, and then uses the sample pairs to construct the training and testing set. Therefore, the scale of training samples in Xu et al. (2014) is $O(cn)$, while the scale of the training samples in the proposed method is $O(cn^2)$.

As the collaborative representation coefficients of LRC and CRC are complementary, and the combination of their residuals achieves higher classification accuracy (Zhang et al., 2014), we use the linear combination of the residuals of LRC and CRC for face recognition. For LRC, the testing sample pair is represented by a linear combination of the training sample pair of each class, respectively, and its residual is defined as the deviation between the testing sample pair and its representations by the training sample pairs of each class. For CRC, the testing sample pair is represented by a linear combination of all the training sample pairs, and its residual definition is similar to LRC. So we use the weighted sum of the residuals of LRC and CRC for classification.

Fig. 1. Five images and their virtual images of one person in the ORL dataset. The original images are shown in the first row. The virtual images are shown in the second row.

The rest parts of the paper are organized as follows. Section 2 presents the proposed method. Section 3 describes the rationales of the proposed method. Section 4 shows the experimental results and Section 5 presents the conclusion.

2. The proposed method

2.1. Construction of training and test sample pairs

Suppose the resolution of each training and test image is $D \times R$ pixels. There are c classes and each class has n original facial images for training. Let $\mathbf{x}^{\text{ori}}_{i,k} \in \mathbb{R}^{DR\times 1}$ $(k = 1, \ldots, n,\ i = 1, \ldots, c)$ be the kth original training sample of the ith class. Let $\mathbf{z}^{\text{ori}}$ be the original testing sample. $\mathbf{x}^{\text{vir}}_{i,k}$ and $\mathbf{z}^{\text{vir}}$ are the virtual facial images of $\mathbf{x}^{\text{ori}}_{i,k}$ and $\mathbf{z}^{\text{ori}}$. $\mathbf{x}^{\text{vir}}_{i,k}$ and $\mathbf{z}^{\text{vir}}$ are obtained using $[\mathbf{x}^{\text{vir}}_{i,k}]_{m,s} = [\mathbf{x}^{\text{ori}}_{i,k}]_{m,R-s+1}$ and $[\mathbf{z}^{\text{vir}}]_{m,s} = [\mathbf{z}^{\text{ori}}]_{m,R-s+1}$, where $m = 1, \ldots, D$, $s = 1, \ldots, R$, $[\mathbf{x}^{\text{vir}}_{i,k}]_{m,s}$ and $[\mathbf{z}^{\text{vir}}]_{m,s}$ denote the entry at the mth row and sth column of matrix $\mathbf{x}^{\text{vir}}_{i,k}$ and $\mathbf{z}^{\text{vir}}$, respectively. Five facial images and their virtual images of one person in the ORL dataset are shown in Fig. 1.

We have four ways to generate training sample pairs:

(1) Any two samples in $\{\mathbf{x}^{\text{ori}}_{i,1}, \mathbf{x}^{\text{ori}}_{i,2}, \ldots, \mathbf{x}^{\text{ori}}_{i,n}\}$ can formulate a training sample pair $\mathbf{x}^{\text{ori}}_{i,p,q} = [\begin{smallmatrix}\mathbf{x}^{\text{ori}}_{i,p}\\ \mathbf{x}^{\text{ori}}_{i,q}\end{smallmatrix}] (1 \le p, q \le n)$. So we can construct a set $\mathcal{X}_1$ of n^2 sample pairs.

(2) Any two samples in $\{\mathbf{x}^{\text{vir}}_{i,1}, \mathbf{x}^{\text{vir}}_{i,2}, \cdots, \mathbf{x}^{\text{vir}}_{i,n}\}$ can formulate a training sample pair $\mathbf{x}^{\text{vir}}_{i,p,q} = [\begin{smallmatrix}\mathbf{x}^{\text{vir}}_{i,p}\\ \mathbf{x}^{\text{vir}}_{i,q}\end{smallmatrix}] (1 \le p, q \le n)$. So we can construct a set $\mathcal{X}_2$ of n^2 sample pairs.

(3) For any $1 \le p < q \le n$, we can formulate a training sample pair $\mathbf{x}^{\text{o/v}}_{i,p,q} = [\begin{smallmatrix}\mathbf{x}^{\text{ori}}_{i,p}\\ \mathbf{x}^{\text{vir}}_{i,q}\end{smallmatrix}]$. So we can construct a set $\mathcal{X}_3$ of $n(n-1)$ sample pairs.

(4) For any $1 \le p < q \le n$, we can formulate a training sample pair $\mathbf{x}^{\text{v/o}}_{i,p,q} = [\begin{smallmatrix}\mathbf{x}^{\text{vir}}_{i,p}\\ \mathbf{x}^{\text{ori}}_{i,q}\end{smallmatrix}]$. So we can construct a set $\mathcal{X}_4$ of $n(n-1)$ sample pairs.