Neurocomputing 151 (2015) 471–480

Contents lists available at ScienceDirect

Neurocomputing

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

Nearest orthogonal matrix representation for face recognition

Jian Zhang [a,b], Jian Yang [a,*], Jianjun Qian [a], Jiawei Xu [c]

[a] Department of Computer Science, Nanjing University of Science and Technology, Nanjing 210094, PR China
[b] Department of Computer Engineering, Huaihai Institute of Technology, Lian yungang 222000, PR China
[c] School of Computer Science, University of Lincoln, UK

ARTICLE INFO

Article history:
Received 12 April 2014
Received in revised form
14 July 2014
Accepted 10 September 2014
Communicated by Ran He
Available online 28 September 2014

Keywords:
Face recognition
Image representation
Pattern classification
Singular vector decomposition (SVD)
Sketch image recognition

ABSTRACT

This paper presents a simple but effective method for face recognition, named nearest orthogonal matrix representation (NOMR). Specifically, the specific individual subspace of each image is estimated and represented uniquely by the sum of a set of basis matrices generated via singular value decomposition (SVD), i.e. the nearest orthogonal matrix (NOM) of original image. Then, the nearest neighbor criterion is introduced for recognition. Compared with the current specific individual subspace based methods (e.g. the sparse representation based classifier, the linear regression based classifier and so on), the proposed NOMR is more robust for alleviating the effect of illumination and heterogeneous (e.g. sketch face recognition), and more intuitive and powerful for handling the small sample size problem. To evaluate the performance of the proposed method, a series of experiments were performed on several face databases: Extended Yale B, CMU-PIE, FRGCv2, AR and CUHK Face Sketch database (CUFS). Experimental results demonstrate that the proposed method achieves encouraging performance compared with the state-of-the-art methods.

1. Introduction

In past few decades, face recognition has become a popular area of research in computer vision and one of the most successful applications of image analysis and understanding. So far, the state-of-the-art recognition technologies can strike high accuracy under controlled environment, such as frontal faces with comfortable lighting conditions [1]. However, most existing face recognition technologies are still far from the perfection in uncontrolled cases of larger illumination, occlusion, disguise etc. [2]

Recently, reconstruction-representation-based methods, utilizing known samples as the basis to linearly rebuild the non-ideal query sample to eliminate the obstruction, has aroused widespread concerns in the field of robust face recognition. Derived from sparse representation and robust principle component analysis (RPCA) [3], an important relevant work named sparse representation based classifier (SRC) was firstly proposed in [4], where the rarely known samples are selected by L1-optimizer to reconstruct the query sample. With huge computational cost, SRC shows strong ability in dealing with sparse random pixel corruption and block occlusion. Relatively low-cost, Shi et al. advocated the L2-optimizer based regression method selecting all known samples to reconstruct the query sample in [5]. Subsequently, Zhang et al. analyzed the working principle of SRC and asserted the importance of collaborative representation strategy than L1-norm based sparsity constraint [6]. Thus, a collaborative representation based classifier (CRC) was proposed with ridge regression (L2-norm).

On the other hand, above representation residuals are usually measured by L1-norm or L2-norm corresponding to Gaussian or Laplacian distribution respectively. However, the distribution of representation residuals is really complicated to suppress the performance of above mentioned methods [7,8]. To this end, Yang et al. borrowed the idea of robust regression and proposed a regularized robust coding method [7,8]. He et al. further presented a correntropy based sparse representation (CESR) algorithm using the correntropy induced robust error metric [9,10].

Actually, the strategy of seeking the best linear representation in known samples, i.e. minimum linear reconstruction representation residuals, can be traced back to the nearest neighbor line (NFL) [11], which aims to extend the capacity of prototype features by computing a linear function to interpolate and extrapolate each sample pair in the same class. Chien and Wu further extended NFL and proposed the nearest feature plane (NFP) and the nearest feature space (NFS) methods for pattern classification [12]. As a special case of NFS, Naseem proposed linear regression based classifier (LRC) using the whole class samples to construct the linear prototype [13].

* Corresponding author.
E-mail addresses: zhangjianhhit@gmail.com (J. Zhang), csjyang@mail.njust.edu.cn (J. Yang), qjjtx@126.com (J. Qian), jxu@lincoln.ac.uk (J. Xu).

http://dx.doi.org/10.1016/j.neucom.2014.09.019
0925-2312/

Essentially, all successful application methods mentioned above are underpinned by the following assumption [4–16]:

Assumption 1. Specific individual subspace assumption: *samples from a specific object class are known to lie in a specific linear subspace, i.e. any test imagey from a specific class i has a specific individual subspace spanned by* $A_i = [a_{i1}, a_{i2}, \ldots, a_{iN}]$ *such that*

$$y = A_i w. \tag{1}$$

In SRC and CRC, the specific individual subspace of query sample is achieved by the linear regression (L1-optimizer or L2-optimizer). While the specific individual subspace of NFL, NFP, NFS and LRC is directly spanned by the known samples in one class. Despite different strategies, the specific individual subspace of all above methods is composed of the known samples. Thus, to guarantee sufficient representation of a query sample, relevant methods depend on the holding of following assumption heavily:

Assumption 2. Large sample size assumption: *there are sufficient known samples for each class, such that any query sample can be sufficiently represented using only the known samples in the same class.*

However, Assumption 2 does not really hold well. This leads to larger representation residuals and lower recognition rates of above mentioned methods. For example, none of the above mentioned methods can handle the case of the single image per person problem well. Towards this end, a simple but effective basis-acquisition technique was proposed, coined nearest orthogonal matrix representation (NOMR) explicitly utilizing Assumption 1 rather than implicitly using known samples. Specifically, the specific individual subspace of each image is firstly estimated and represented uniquely by the sum of a set of basis matrices generated via singular value decomposition (SVD), the nearest orthogonal matrix (NOM) of original image. Then, a simple nearest neighbor based criterion is introduced for recognition. This idea behind NOMR is that the visual face images can be divided into two parts by SVD: the special intrapersonal subspace basis corresponding to essence identity and the disguising space associated with various appearance changes such as illumination etc. Consistent with the Assumption 1, the same face images should have similar NOMs.

Compared with traditional specific individual subspace based methods, the proposed NOMR with precise, complete and expressive basis is more robust for alleviating the effect of illumination and heterogeneous (e.g. sketch face recognition), and more intuitive and powerful for handling the small sample size problem in face recognition.

In literatures, some methods with SVD for face recognition have been proposed but are quite different from ours. In [17], Hong et al. firstly proposed a singular value decomposition based face recognition method which uses the singular values of the original face image as the feature. In refs. [18–20], SVD based methods were further proposed for face recognition, which still used the singular values as the image representation features. In spite of perfect mathematical theory and good performance in small sample size databases, many experiments show that the above mentioned methods cannot get a satisfied recognition results on large databases with variations occlusion. In contrast to the methods with singular values, Tian et al. advocated that the left and right orthogonal matrix of SVD have more information for recognition [21]. They proposed a new feature extraction method which takes the projection coefficients with a proper selected orthogonal base of SVD (e.g. the average of one class images) as the feature for face recognition. This method probably leads to a better recognition performance with Bayesian classifier, but still falls into the category of using the singular value feature (i.e. the projection coefficients). In addition, the Non-negative Matrix factorization (NMF) is also established on the SVD. Lee et al. firstly use NMF to yield sparse representation of localized features to represent distributed parts over a face image in [22]. However, the NMF like Eigenface is established on the common subspace assumption rather than the specific individual subspace assumption.

The remainder of this paper is organized as follows. Section 2 presents the preliminary topics about the SVD. Section 3 develops the idea of the nearest orthogonal matrix representation for face recognition. Section 4 conducts extensive experiments to verify the validity of our approach. Section 5 offers our conclusions and future work.

2. Preliminaries of SVD

The singular value decomposition (SVD) is one of the matrix factorization methods, which can be described with the following Theorem in real field.

Theorem 1. ***SVD*** *[23]. If* $A \in R^{\mathbf{m} \times \mathbf{n}}$*, then there exist orthogonal matrix* $U = [u_1, \ldots, u_m] \in R^{m \times m}$ *and* $V = [v_1, \ldots, v_n]^T \in R^{n \times n}$ *such that*

$$A = U \Sigma V^T, \tag{2}$$

where $\Sigma = diag(\sigma_1, \sigma_2, \ldots, \sigma_p, 0, \ldots, 0)$ is an $m \times n$ rectangular diagonal matrix with nonnegative real numbers on the diagonal and $p = \text{rank}(A)$.

The diagonal entries σ_i are denoted as the singular values of A, by convention arranged in non-increasing order $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_p \geq 0$. The columns of U are termed left-singular vectors of A and the columns of V are called right-singular vectors of A.

In theory, the singular vector decomposition can be achieved by the eigen-decomposition. However, the SVD is really computed more efficiently. The specific algorithms can refer to the literature [24].

Naturally derived, the matrix $A \in R^{m \times n}$ can also be expressed as

$$A = \sum_{i=1}^{p} \sigma_i A_i = \sum_{i=1}^{p} \sigma_i u_i v_i^T \tag{3}$$

with the singular values and vectors. Here u_i and v_i are the ith columns of the corresponding singular vectors, σ_i are the ordered singular values, $p = \text{rank}(A)$. Generally, the Eq. (3) is coined as the separate model of SVD.

Removing the singular values from the Eq. (2), we will acquire

$$L: = U \otimes V^T = \sum_{i=1}^{p} u_i v_i^T. \tag{4}$$

Eq. (4) is usually called the *symmetric orthogonalization* of the matric$A$ and the symbol $\otimes$ denotes the outer product of two matrices. It should be noted that L is unique since any sequence of sign choices for the columns of V determines a sequence of signs for the columns of U. In addition, L is also the *nearest orthogonal matrix* (defined as an orthogonal matrix Q which has the minimum$\|Q - A\|_F$) of A, which is underpinned by the following theorem:

Theorem 2. ***NOM****. Over all orthogonal matrices* Q*,* $\|Q - A\|_F$ *is minimized if and only if* $Q = L$*.*

Theorem 2 is the key of the orthogonal procrustes problem which was originally solved by Peter Schonemann in [25,26]. A detailed proof is given in [27]. The classical OPP asks how closely a matrix $A \in R^{m \times n}$ can be approximated by a second given matrix $B \in R^{p \times n}$ multiplied by a matrix $\Omega \in R^{m \times p}$ with orthogonal columns in the sense of Frobenius norm. This problem is equivalent to finding the nearest orthogonal matrix (NOM) to a given matrix $M = BA^T$. In addition, this also intuitively makes sense because an