J. Vis. Commun. Image R. 24 (2013) 111–116

Contents lists available at SciVerse ScienceDirect

J. Vis. Commun. Image R.

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

Face recognition via Weighted Sparse Representation

Can-Yi Lu [a,b,*], Hai Min [a,b], Jie Gui [a], Lin Zhu [a,b], Ying-Ke Lei [c]

[a] Hefei Institute of Intelligent Machines, Chinese Academy of Sciences, Hefei, China
[b] Department of Automation, University of Science and Technology of China, Hefei, China
[c] Electronic Engineering Institute, Hefei, China

ARTICLE INFO

Article history:
Available online 9 May 2012

Keywords:
Face recognition
Weighted Sparse Representation
Nearest Feature Classifiers
Locality
Linearity
Sparse Representation
Classification
Local representation

ABSTRACT

Face recognition using Sparse Representation based Classification (SRC) is a new hot technique in recent years. SRC can be regarded as a generalization of Nearest Neighbor and Nearest Feature Subspace. This paper first reviews the Nearest Feature Classifiers (NFCs), including Nearest Neighbor (NN), Nearest Feature Line (NFL), Nearest Feature Plane (NFP) and Nearest Feature Subspace (NFS), and formulates them as general optimization problems, which provides a new perspective for understanding NFCs and SRC. Then a locality Weighted Sparse Representation based Classification (WSRC) method is proposed. WSRC utilizes both data locality and linearity; it can be regarded as extensions of SRC, but the coding is local. Experimental results on the Extended Yale B, AR databases and several data sets from the UCI repository show that WSRC is more effective than SRC.

1. Introduction

Face recognition has become one of the most intensively investigated topics in biometrics. Likewise in other fields in pattern recognition, the identification of faces has been addressed from different approaches according to the chosen representation and the design of the classification method. Over the past two decades, industrial interests and research efforts in face recognition have been motivated by a wide range of potential applications such as identification, verification, posture/gesture recognizers and intelligent multimodal systems. The real world face images are typically with significant lighting, expression, pose, etc. variations. The robust face recognition remains a very challenging task.

Beyond the preprocessing (face detection and face alignment), a common face recognition system consists of two stages: (i) feature extraction: numerous methods have been proposed to project data to a low dimensional feature subspace, e.g. PCA [1], LDA [2] and LPP [3] and (ii) classifier construction and label prediction. Usually Nearest Neighbor (NN) [4] and Nearest Feature Subspace (NFS) [5–7] are used. However, NN classifies the query image by only using its Nearest Neighbor in the training data; therefore it can easily be affected by noise. NFS approximates the query image by using all the images belonging to the same class, and predicts the image to the class which minimizes the reconstruction error. But NFS may fail for the case where classes are highly correlated to each other. To overcome these problems, a Sparse Representation based Classification (SRC) [8] method was proposed. A query image is first sparsely coded over the template images, and then the classification is performed by checking which class yields the least coding errors. SRC is robust to occlusion, illumination and noise, and achieves excellent performance. It boosted the research of sparsity based face recognition. Elhamifar and Vidal [9] proposed a more robust classification method using structured sparse representation, while Gao et al. [10] introduced a kernelized version of SRC. Qiao et al. [11] proposed a sparsity preserving projection method which was unsupervised while Lu [12] provided a supervised dimensionality reduction method for SRC. Cheng et al. [13] discussed the ℓ^1-graph based image analysis. A recent review of sparse representation based machine learning can be found in [14].

For general pattern classification problems such as dimensionality reduction, classification, clustering, etc., the locality structure of data has been observed to be critical [15,16]. NN utilizes the locality structure of data, while NFS and SRC uses the linearity structure of data. It has been shown that in some case locality is more essential than sparsity but the original sparse coding does not guarantee to be local which lead to unstable. In order to overcome this problem, we present an extension of SRC, called Weighted Sparse Representation based Classification (WSRC). WSRC integrates the locality structure of data into sparse representation in a unified formulation.

The remainder of this paper is organized as follows: Section 2 reviews the SRC algorithm, and a serious of related work Nearest Feature Classifiers (NFCs). Section 3 presents the WSRC method and discusses the relationships between WSRC, SRC and NFCs. The experimental results are presented in Section 4. Finally, we conclude this paper in Section 5.

* Corresponding author at: Hefei Institute of Intelligent Machines, Chinese Academy of Sciences, Hefei, China.

E-mail addresses: canyilu@gmail.com, canyilu@mail.ustc.edu.cn (C.-Y. Lu).

1047-3203/$ - see front matter
http://dx.doi.org/10.1016/j.jvcir.2012.05.003

Table 1
A general formulation of NFCs, SRC and WSRC.

Methods	Objective function	Constrains	Decision rule
NN	$\min_{\{\alpha^c\}}\sum_{c=1}^{C}\|\|y-X^c\alpha^c\|\|$	$\|\|\alpha^c\|\|_0=1,\ 1^T\alpha^c=1,\ c=1,\ldots,C$	$\min_c\|\|y-X^c\alpha^c\|\|$
NFL		$\|\|\alpha^c\|\|_0=2,\ 1^T\alpha^c=1,\ c=1,\ldots,C$	
NFP		$\|\|\alpha^c\|\|_0=3,\ 1^T\alpha^c=1,\ c=1,\ldots,C$	
NFS		–	
SRC	$\min_\alpha\|\|\alpha\|\|_1$	$y = X\alpha$ or $\|\|y-X\alpha\|\|\leqslant\varepsilon$	
WSRC	$\min_\alpha\|\|W\alpha\|\|_1$		

2. Nearest Feature Classifiers and Sparse Representation based Classification

Given sufficient C classes training samples, a basic problem in pattern recognition is to correctly determine the class which a new coming (test) sample y belongs to. We arrange the n_c training samples from the cth class as columns of a matrix $X^c = [X_1^c,\ldots,x_{n_c}^c]\in R^{m\times n_c}$ where m is the dimension. Then we obtain the training sample matrix $X=[X^1,\ldots,X^c]\in R^{m\times n}$ where $n=\sum_{c=1}^{C}n_c$ is the total number of training samples.

2.1. Nearest Feature Classifiers

In a sense, SRC can be considered as a generalization of popular classifiers such as NN and NFS. It strikes a balance between NN and NFS, which is similar to Nearest Feature Line (NFL) [17] and Nearest Feature Plane (NFP) [18]. For the convenience of latter discussion, we briefly review the NFCs, including NN, NFL, NFP and NFS.

NN is the simplest nonparametric method for classification, it assigns the label of the test sample by its Nearest Neighbor. NFL is an extension of NN which classifies the test sample by assigning it the class label according to the Nearest Feature Line. NFP further uses a feature plane instead of feature line. NFS uses all the data points in each class to span a subspace and classifies the test sample to the nearest subspace.

Generally speaking, NFL classifier is supposed to handle more variations than NN, NFP should capture more variations of each class than NFL and NFS should handle more variations than NFP. So, it is expected that NFL outperforms NN, NFP performs better than NFL and NFS is more accurate than NFP. It was suggested that the improvement gained by using feature lines is due to their faculty to expand the representational ability of the available feature points, accounting for new conditions not represented by the original set.

The differences between different NFCs are their representational ability for query point. As shown in Table 1, we formulate the NFCs as the optimization problems, which is a new perspective for understanding NFCs and their relationships. NFCs have the same objective function and decision rule. They match well from optimization to classification. The differences between NFCs reflected in constrains. Different NFCs use different number of points of each class to represent the query point. NN uses only one, NFL uses two, NFP uses three, and NFS uses all the points of each class. In this sense, NN can be called *Nearest Feature Point*. Different number of points used in each class for representing the query point results to different representational ability. The more data points of each class are used, the more powerful of their representational ability, and the more variations they can capture.

2.2. Sparse Representation based Classification

SRC is based on the assumption that the training samples from a single class do lie on a subspace. Any new (test) sample y from the same class will approximately lie in the linear span of the training samples associated with object c

$$Y = X^c\alpha_0^c$$

Since c is unknown, the linear representation of y can be rewritten in terms of all training samples as

$$y = X\alpha_0,$$

where $\alpha_0=[0^T,\alpha_0^{cT},0^T]$ is a coefficient vector, the nonzero entries of which associated with the cth class. Motived by the sparse coefficient, SRC aims to solve the following ℓ^0-minimization problem:

Table 2
Recognition rates of NN, NFS, SRC and WSRC on Extended Yale B database: (a) Eigenfaces; (b) Randomfaces; (c) Fisherfaces.

Dimension	30	56	120	504
(a) Eigenfaces				
NN	83.12	90.89	94.19	93.99
NFS	89.95	92.70	93.96	94.74
SRC	85.48	92.54	95.84	98.59
WSRC	88.07	94.51	96.94	98.51

Dimension	30	56	120	504
(b) Randomfaces				
NN	71.11	74.10	79.98	82.18
NFS	82.42	91.02	93.56	94.43
SRC	84.38	90.82	94.82	97.33
WSRC	86.26	91.84	95.45	97.65

Dimension	5	10	15	20	25	30	35
(c) Fisherfaces							
NN	55.81	73.16	79.75	82.34	83.83	84.14	86.03
NFS	48.59	69.78	78.65	80.61	82.97	83.44	84.54
SRC	48.04	69.00	77.94	80.93	83.44	84.30	86.34
WSRC	49.14	70.02	78.57	81.16	82.97	84.46	86.50