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

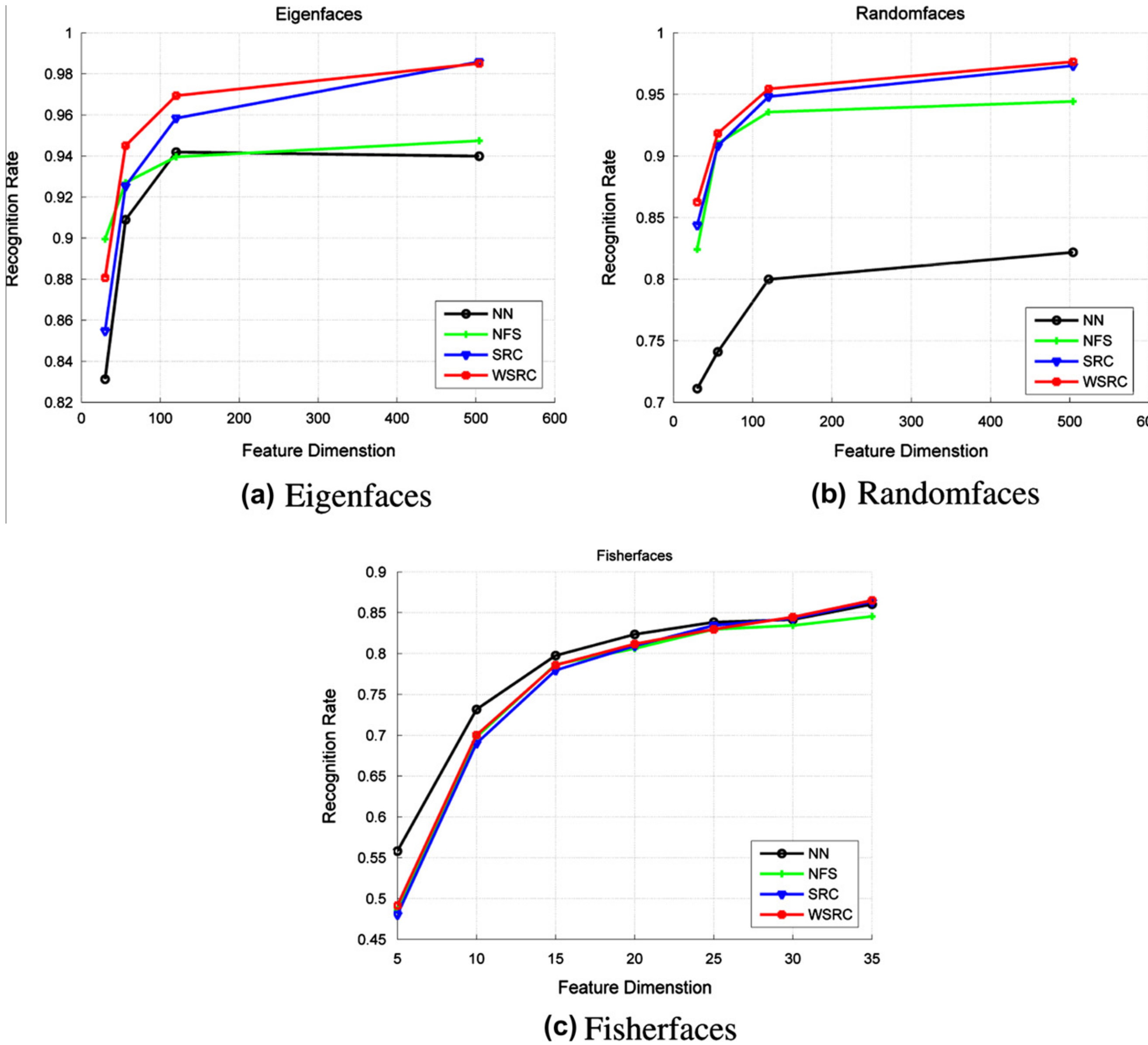

Fig. 1. Accuracy versus reduced dimensionality on Extended Yale B database: (a) Eigenfaces; (b) Randomfaces; (c) Fisherfaces.

$$(\ell^0): \quad \hat{\alpha}_0 = \arg\min ||\alpha||_0 \quad \text{subject to } y = X\alpha,$$

where $||\cdot||_0$ denotes the ℓ_0-norm, which counts the number of nonzero entries in a vector. However, the problem of finding the sparsest solution of an underdetermined system of linear equations is NP-hard and difficult even to approximate [19]. The theory of compressive sensing [20,21] reveals that if the solution α is sparse enough, the solution of the ℓ^0-minimization problem is equal to the following ℓ^1-minimization problem:

$$(\ell^1): \quad \hat{\alpha}_1 = \arg\min ||\alpha||_1 \quad \text{subject to } y = X\alpha.$$

In order to deal with occlusion, the ℓ^1-minimization problem is extended to the following stable ℓ^1-minimization problem:

$$(\ell^1_s): \quad \hat{\alpha}_1 = \arg\min ||\alpha||_1 \quad \text{subject to } ||y - X\alpha|| \leqslant \varepsilon,$$

where $\varepsilon > 0$ is a given tolerance.

After obtaining the sparsest solution $\hat{\alpha}_1$, SRC uses the same decision rule as NFCs, see Table 1.

Different from the traditional pattern recognition methods, SRC is not critical to the choice of an "optimal" feature transformation. If the solution $\hat{\alpha}_1$ is sparse enough, then with overwhelming probability, it can be correctly recovered via ℓ^1-minimization from any sufficiently large number m of linear measurements. Even random projections or downsampled images should be performed as well as any other carefully engineered features. However, if the number m of linear measurements is not large enough, SRC may perform worse than NN or NFS. This indicates that the discriminative information from the linearity structure of data in lower dimensional feature subspaces is not enough for SRC. We propose an extension of SRC by imposing the locality constraint on the sparsity regularized reconstruction problem, which is described in next section.

3. Weighted Sparse Representation based Classification

In this section, we present the Weighted Sparse Representation based Classification or WSRC method. NN or kNN utilizes the locality structure of data while NFS and SRC use the linearity structure of data. Due to the mechanism of ℓ^1-minimization, the sparse coding coefficients may vary a lot even for similar test samples. The sparse coding may reconstruct a test sample by training images which are

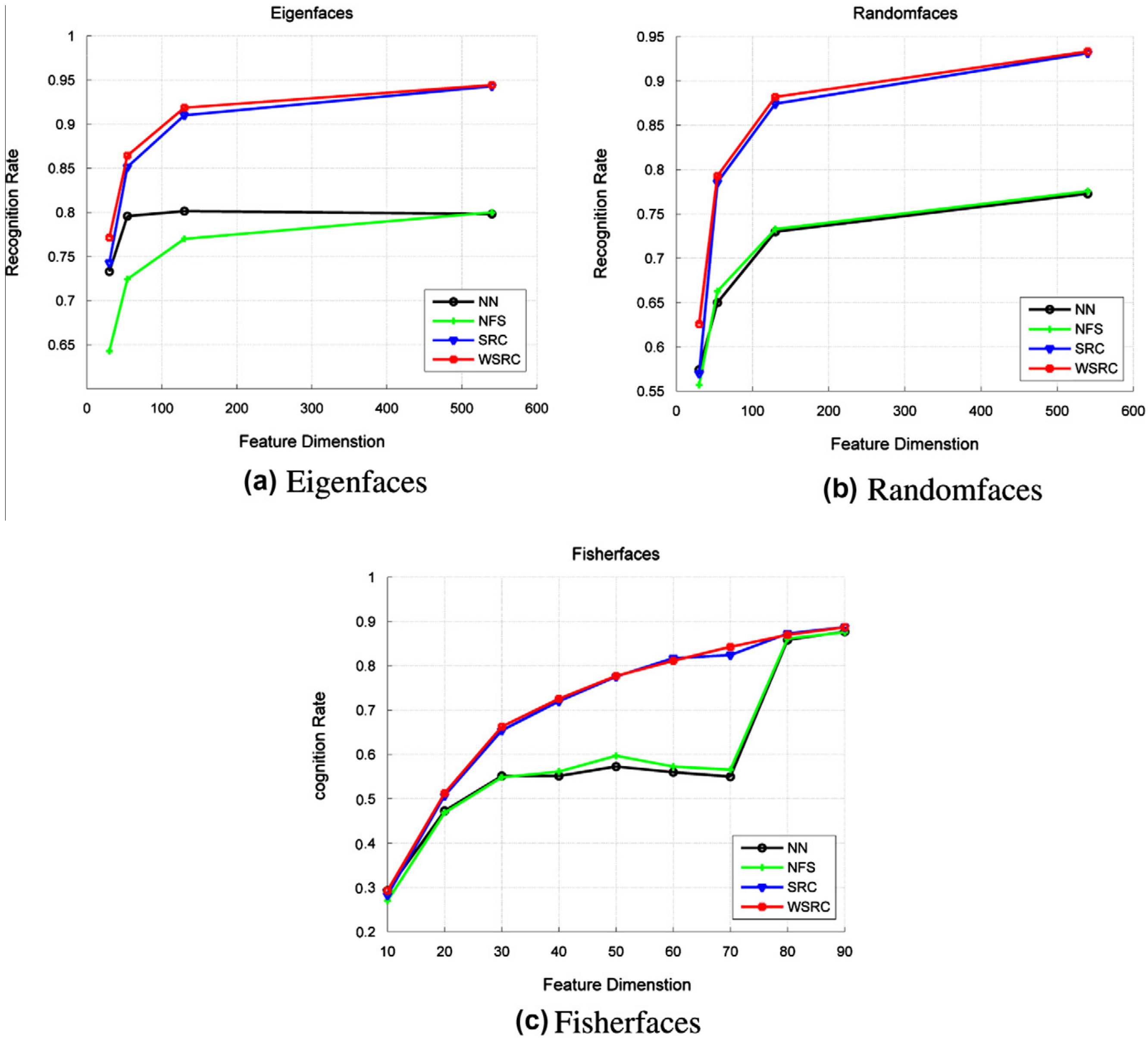

Fig. 2. Accuracy versus reduced dimensionality on AR database: (a) Eigenfaces; (b) Randomfaces; (c) Fisherfaces.

far from the test sample and thus produce unstable classification results. Thus SRC tends to lose locality information, but Yu et al. [22,23] pointed out that under certain assumptions locality is more essential than sparsity. It also has been shown that encouraging the coding to be local is effective for image classification [16] and video annotation [24]. To overcome the drawback of sparse coding, we propose the more robust weight sparse representation method which integrates both sparsity and data locality structure into a unified formulation. WSRC preserves the similarity between the test sample and its neighboring training data while seeking the sparse linear representation. Similar to SRC, we use all the training data as dictionary, and impose the locality on the ℓ^1 regularization. WSRC solves the following weighted ℓ^1-minimization problem:

$$(\text{Weighted } \ell^1): \quad \hat{\alpha}_1 = \arg\min \|W\alpha\|_1 \quad \text{subject to } y = X\alpha,$$

where W is a block-diagonal matrix, which is the locality adaptor that penalizes the distance between y and each training data. Specifically,

$$\operatorname{diag}(W) = [\operatorname{dist}(y, x_1^1), \ldots, \operatorname{dist}(y, x_{n_C}^C)]^T,$$

where $\operatorname{dist}(y, x_i^c) = \|y - x_i^c\|^s$, s is the locality adaptor parameter. When $s = 0$, WSRC degenerates to SRC. We further normalize $\operatorname{diag}(W)$ by dividing $\max(\operatorname{diag}(W))$ from $\operatorname{diag}(W)$. A larger $\operatorname{dist}(y, x_i^c)$ indicates a farther distance between y and x_i^c, it can well characterize the similarity between the test sample and training data. Our formulation of WSRC can generate more discriminative sparse codes which can be used to represent the test sample more robustly. The coding coefficient of WSRC tends to be local in linear representation.

In order to deal with occlusion, the weighted ℓ^1-minimization problem can be extended to the following stable ℓ^1-minimization problem:

$$(\text{Weighted } \ell_s^1): \hat{\alpha}_1 = \arg\min \|W\alpha\|_1 \quad \text{subject to } \|y - X\alpha\| \leqslant \varepsilon,$$

where $\varepsilon > 0$ is a given tolerance.

Once the sparse coefficient obtained, it is used to recognize the query image as the class with the lowest reconstruction error. The decision rule is the same as NFCs and SRC, which is shown in Table 1.

WSRC combines both linearity and locality information for improving recognition. From the perspective of linearity, WSRC is a direct extension of SRC. From the perspective of locality, WSRC can be regarded as an extension of kNN, while WSRC is a neighborhood adaptive algorithm. Actually, considering the sparsity, kNN can named as *k sparse neighborhood representation*, while SRC is based on *sparse linear representation*, and WSRC strikes a balance between them.

4. Experimental verification

In this section, we investigate the performance of our proposed WSRC method for face representation and recognition. We perform two sets of experiments. In the first set we compare the recognition accuracy on the different classifiers for face recognition. The Extended Yale B [25] and AR [26] are used to test the performance of WSRC and its competing methods, including NN, NFS and SRC. Since the dimension of the original face vectors is large, we use several conventional holistic face features, namely Eigenfaces [1], Fisherfaces [2] and Randomfaces [8]. In the second set of experiments, we test the WSRC algorithms on some benchmark databases from the UCI machine learning repository [27]. We choose 15 databases from UCI that do not have missing values.

The matlab code of WSRC will be available from our homepage and the SPAMS package [28,29] is used to solve the stable weighted ℓ^1-minimization problem. In our experiments, we set $\varepsilon = 10^{-4}$ and $s = 1.5$ in WSRC.

4.1. Extended Yale B database

The Extended Yale B database consists of 2414 frontal face images of 38 subjects under various lighting conditions. For each subject, we randomly select half of the images for training and the remainder for test.

Fig. 1 shows the recognition performance for the various features, in conjunction with four classifiers: NN, NFS, SRC and WSRC. Table 2 shows the detail recognition accuracy of the methods considered.

From Fig. 1 and Table 2, we can see that WSRC outperforms SRC by using Eigenfaces and Randomfaces. Such improvements become more significant in lower dimensional subspaces. This implies, in lower dimensional subspaces, the data linearity is not enough to separate data from different subspaces, and the imposed data locality is better preserved which help improve recognition performance. However, WSRC is only slightly better than SRC using Fisherfaces. That is because after the projection of LDA, the training data belong to the same subject are very close, the original SRC coding is already local, which is similar to WSRC.

4.2. AR database

The AR database consists of over 4000 frontal images for 126 subjects. In this experiment, we use a subset (with only illumination and expression changes) contains 50 male subjects and 50 female subjects. For each subject, the seven images from session 1 were used for training, with the other seven images from session 2 for test. The comparison of competing methods is given in Fig. 2 and Table 3.

Similar to the results on Yale B database, WSRC outperforms SRC, NN and NFS by using Eigenfaces and Randomfaces, especially in lower dimensional subspaces. The improvement comes from the integrating of data linearity and locality. But they make no obvious difference on Fisherfaces.

4.3. Data sets from UCI repository

In order to test the effectiveness of WSRC on other pattern recognition problems, we conduct this experiment on 15 datasets

Table 4
Mean recognition rates (%) and standard deviations on 15 data sets from the UCI machine learning repository.

Datasets	NN	NFS	SRC	WSRC
Air	94.29 ± 4.84	87.14 ± 2.47	**95.14** ± 0.78	**95.14** ± 2.78
Austra	**78.82** ± 4.05	76.76 ± 4.54	78.38 ± 6.20	76.03 ± 5.80
Breast	72.59 ± 11.48	70.37 ± 0.00	71.48 ± 5.25	**74.44** ± 6.86
Breast_gy	70.37 ± 7.41	**70.74** ± 3.24	68.89 ± 7.45	70.00 ± 7.70
German	67.20 ± 6.20	70.00 ± 0.00	71.90 ± 4.56	**73.10** ± 5.82
Glass	31.50 ± 6.02	26.50 ± 5.76	33.50 ± 6.75	**39.00** ± 9.12
Heart	76.30 ± 6.34	**76.67** ± 4.64	71.85 ± 5.84	74.44 ± 5.91
Ionosphere	64.71 ± 0.00	40.29 ± 3.41	89.41 ± 5.75	**91.18** ± 5.00
Iris	96.00 ± 4.66	94.00 ± 6.63	95.33 ± 4.50	**97.33** ± 3.44
Sonar	84.00 ± 8.76	79.50 ± 11.89	84.50 ± 6.43	**85.50** ± 5.99
Vote	91.19 ± 2.76	76.43 ± 5.55	93.57 ± 4.05	**93.81** ± 4.52
Vowel	98.18 ± 1.44	63.18 ± 7.48	97.27 ± 1.44	**98.41** ± 1.53
WBC	**94.93** ± 2.46	76.42 ± 4.38	91.94 ± 5.27	92.39 ± 4.31
Wine	95.63 ± 6.62	83.13 ± 9.34	95.63 ± 5.15	**96.88** ± 4.42
X8D5K	**100.00** ± 0.00	92.20 ± 3.46	**100.00** ± 0.00	**100.00** ± 0.00

Table 3
Recognition rate of NN, NFS, SRC and WSRC on AR database: (a) Eigenfaces; (b) Randomfaces; (c) Fisherfaces.

Dimension	30	54	130	540
(a) Eigenfaces				
NN	73.29	79.57	80.14	79.81
NFS	64.29	72.43	77.00	80.00
SRC	74.29	85.14	91.00	94.29
WSRC	77.14	86.43	91.86	94.43

Dimension	30	54	130	540
(b) Randomfaces				
NN	57.43	65.00	73.00	77.29
NFS	55.71	66.29	73.29	77.57
SRC	57.00	78.57	87.43	93.14
WSRC	62.57	79.29	88.20	93.33

Dimension	10	20	30	40	50	60	70	80	90
(c) Fisherfaces									
NN	29.29	47.29	55.14	55.14	57.29	56.00	55.00	85.86	87.71
NFS	27.00	46.86	54.86	56.14	59.71	57.29	56.57	86.14	87.57
SRC	28.29	50.71	65.43	72.00	77.57	81.71	82.43	87.29	88.71
WSRC	29.29	51.29	66.29	72.57	77.71	81.14	84.29	87.00	88.71

from the UCI machine learning repository. For each data set, we perform 10-fold cross validation and record the mean of accuracies and standard deviation. Experimental results are shown in Table 4. The first column shows the name of the data set. The second column gives the classification results from the NN. The results of NFS are presented in the third column. The fourth column gives the classification results from SRC. The final column of Table 4 gives the results from our proposed WSRC.

In Table 4, the best results for each dataset are highlighted in bold. We can see that WSRC performs better than the competing methods on most of these data sets. Furthermore, WSRC also obtains good performance on many data sets which implies that it not only works well for face recognition, but also many other practical pattern recognition problems.

4.4. Discussion

Based on the results on the face databases (Extended Yale B and AR database) and several data sets from the UCI repository, we draw the following observations and discussions:

1. WSRC outperforms SRC by using Eigenfaces and Randomfaces on both face databases, and the improvement is more significant in lower dimensional subspaces. In order to explain this phenomenon, let us consider the ℓ^1-minimization problem. The lower the dimension is, the larger the feasible set is, and the sparser the coefficient is. But the sparse recovery may be not correct due to the insufficient linear measurements. In this situation, the data locality may contain more discriminative information than the data linearity. Then the locality weighted Sparse Representation results to a local sparse representation which is more discriminative. If the dimension is large enough, the feasible set is small, the representation coefficient does not have many choices, the imposed locality may make no obvious difference from SRC. That is why WSRC performs very similar to SRC in high dimensional subspace.
2. WSRC performs similar to SRC by using Fisherfaces. The projected data belong to the same class are very close by using Fisherfaces. The original sparse representation tends to be local, which is similar to WSRC. This also can explain why SRC performs better by using Fisherfaces than other face features. SRC actually utilizes the locality structure of data in lower dimensional subspaces.
3. The experimental results on 15 data sets from the UCI repository show that WSRC also outperforms the completing methods in most case. Face recognition is not the only application of our proposed method.

5. Conclusions

In this paper, we first review the Nearest Feature Classifiers, including NN, NFL, NFP, and NFS, and formulate the NFCs as general optimization problems, which provide a new perspective for understanding NFCs and their relationships. As a generalization of NN and NFS, SRC utilizes the linearity structure of data, it does not perform well in lower dimensional subspaces. A direct extension of SRC, Weighted Sparse Representation is proposed which integrates the locality to sparse coding. Empirical results confirm that WSRC outperforms SRC, especially in lower dimensional subspaces.

Acknowledgments

This work was supported by the grants of the National Science Foundation of China, Nos. 60975005, 61005010, 60873012, 60805021, 60905023, 31071168, 61133010, and the Knowledge Innovation Program of the Chinese Academy of Sciences, Y023A61121.

References

[1] M.A. Turk, A.P. Pentland, Face recognition using eigenfaces, in: IEEE Computer Society Conference on Computer Vision and, Pattern Recognition, 1991, pp. 586–591.
[2] P.N. Belhumeur, J.P. Hespanha, D.J. Kriegman, Eigenfaces vs. fisherfaces: recognition using class specific linear projection, IEEE Transactions on Pattern Analysis and Machine Intelligence 19 (1997) 711–720.
[3] X.F. He, S.C. Yan, Y.X. Hu, P. Niyogi, H.J. Zhang, Face recognition using Laplacianfaces, IEEE Transactions on Pattern Analysis and Machine Intelligence 27 (2005) 328–340.
[4] T.M. Cover, P.E. Hart, Nearest neighbor pattern classification, IEEE Transactions on Information Theory 13 (1967) 21–27.
[5] S. Shan, W. Gao, D. Zhao, Face identification from a single example image based on face-specific subspace (FSS), in: IEEE International Conference on Acoustic, Speech and, Signal Processing, 2002, pp. II/2125–II/2128.
[6] S. Shan, W. Gao, X. Chen, J. Ma, Novel face recognition based on individual eigen-subspaces, in: International Conference on Signal Processing, IEEE, 2000, pp. 1522–1525.
[7] X. Liu, T. Chen, B.V.K.V. Kumar, Face authentication for multiple subjects using eigenflow, Pattern Recognition 36 (2002) 313–328.
[8] J. Wright, A.Y. Yang, A. Ganesh, S.S. Sastry, Y. Ma, Robust face recognition via sparse representation, IEEE Transactions on Pattern Analysis and Machine Intelligence 31 (2009) 210–227.
[9] E. Elhamifar, R. Vidal, Robust classification using structured sparse representation, in: IEEE Conference on Computer Vision and Pattern Recognition, 2011, pp. 1873–1879.
[10] S. Gao, I. Tsang, L.-T. Chia, Kernel sparse representation for image classification and face recognition, in: K. Daniilidis, P. Maragos, N. Paragios (Eds.), Computer Vision – ECCV 2010, Springer, Berlin/Heidelberg, 2010, pp. 1–14.
[11] L.S. Qiao, S.C. Chen, X.Y. Tan, Sparsity preserving projections with applications to face recognition, Pattern Recognition 43 (2010) 331–341.
[12] C.-Y. Lu, Optimized projection for sparse representation based classification, Advanced Intelligent Computing (2012) 83–90.
[13] B. Cheng, J. Yang, S. Yan, Y. Fu, T.S. Huang, Learning with l1-graph for image analysis, IEEE Transactions on Image Processing 19 (2010) 858–866.
[14] J. Wright, Y. Ma, J. Mairal, G. Sapiro, T.S. Huang, S.C. Yan, Sparse representation for computer vision and pattern recognition, Proceedings of the IEEE, Special Issue on Applications of Compressive Sensing & Sparse Representation 98 (2010) 1031–1044.
[15] S.T. Roweis, L.K. Saul, Nonlinear dimensionality reduction by locally linear embedding, Science 290 (2000) 2323–2326.
[16] J. Wang, J. Yang, K. Yu, F. Lv, T. Huang, Y. Gong, Locality-constrained linear coding for image classification, in: IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2010, pp. 3360–3367.
[17] S.Z. Li, J.W. Lu, Face recognition using the nearest feature line method, IEEE Transactions on Neural Networks 10 (1999) 439–443.
[18] J.-T. Chien, C.-C. Wu, Discriminant waveletfaces and nearest feature classifiers for face recognition, IEEE Transactions on Pattern Analysis and Machine Intelligence 24 (2002) 1644–1649.
[19] E. Amaldi, V. Kann, On the approximability of minimizing nonzero variables or unsatisfied relations in linear systems, Theoretical Computer Science 209 (1998) 237–260.
[20] D.L. Donoho, For most large underdetermined systems of linear equations the minimal l(1)-norm solution is also the sparsest solution, Communications on Pure and Applied Mathematics 59 (2006) 797–829.
[21] E.J. Candes, J.K. Romberg, T. Tao, Stable signal recovery from incomplete and inaccurate measurements, Communications on Pure and Applied Mathematics 59 (2006) 1207–1223.
[22] K. Yu, T. Zhang, Y. Gong, Nonlinear learning using local coordinate coding, Advances in Neural Information Processing Systems 22 (2009) 2223–2231.
[23] K. Yu, T. Zhang, Improved local coordinate coding using local tangents, in: International Conference on Machine Learning, 2010, pp. 1215–1222.
[24] M. Wang, X.-S. Hua, J. Tang, R. Hong, Beyond distance measurement: constructing neighborhood similarity for video annotation, IEEE Transactions on Multimedia 11 (2009) 465–476.
[25] A.S. Georghiades, P.N. Belhumeur, D.J. Kriegman, From few to many: illumination cone models for face recognition under variable lighting and pose, IEEE Transactions on Pattern Analysis and Machine Intelligence 23 (2001) 643–660.
[26] A.M. Martinez, The AR face database, CVC Technical, Report 24, 1998.
[27] A. Frank, A. Asuncion, UCI Machine Learning Repository, University of California, School of Information and Computer Science, Irvine, CA, 2010. <http://archive.ics.uci.edu/ml>.
[28] J. Mairal, F. Bach, J. Ponce, G. Sapiro, Online dictionary learning for sparse coding, in: International Conference on Machine Learning, 2009, pp. 689–696.
[29] J. Mairal, F. Bach, J. Ponce, G. Sapiro, Online learning for matrix factorization and sparse coding, Journal of Machine Learning Research 11 (2010) 19–60.