Collaborative Representation Classifier based on K Nearest Neighbors for Classification

Jiangshu Wei, Xiangjun Qi, Mantao Wang
Sichuan Agricultural University
China
bas298@126.com

ABSTRACT: *The sparse representation based classifier (SRC) is a classical representation method for classification. The solution of SRC is obtained by l_1 norm minimization, which can not obtain the closed form solution. Thus, the computational complexity of SRC is a little high. The collaborative representation classifier (CRC) is another classical method for classification. The solution of CRC is obtained by l_2 norm minimization, from the l_2 norm minimization, it can obtain the closed form solution, which makes the computational complexity of CRC is much lower than SRC. Although CRC is effective for classification, there are also some problems about CRC. Under some conditions, some test samples may be misclassified by CRC. This paper proposes a local CRC method, which is called KNN-CRC. This method firstly chooses K nearest neighbors of a test sample from all the training samples, then given a test sample, the test sample is represented by these K training samples. The solution of KNN-CRC is obtained by l_2 norm minimization, and the size of K is much smaller than the total number of all training samples. Thus, the computational complexity of KNN-CRC is much lower than SRC and CRC. Furthermore, the extensive experiments show that the proposed KNN-CRC can obtain very competitive classification results compared with other methods.*

Keywords: SRC, KNN-CRC, Recognition rate, Computational complexity, Classification

Received: 3 October 2015, Revised 12 November 2015, Accepted 21 November 2015

1. Introduction

With the rapid development of computer network technology and information technology, the research of big data is becoming a hot research field recently. Data mining is very important for handling big data application problems. However, the research of data mining includes many aspects. Classification is one of the most important aspects in the field.

Given some training samples from multiple classes, the aim of classification is to assign class labels to a test sample. Classification has been widely used in many fields, such as data mining, computer vision, machine learning, pattern recognition, etc [1-4]. There are many conventional methods for handling classification problems, such as Linear SVM, Nearest Subspace Classifier, K Nearest Neighbor, etc. Recently, there has been an increasing interest in the research of representation theory. With the rapid development of l_0 and l_1 norm minimization algorithms [5-6], sparse representation has been applied for solving many data mining problems. These sparse representation methods have been studied in many literatures. Sparse representation based classifier (SRC) is a classical sparse representation among all the representation methods. SRC was proposed by Wright [1], which is an interesting and effective method for handling the pattern classification problems. A test sample is first sparsely represented by all the training samples, and then the classifier computes the residual for every class, if the *i*th residual is the smallest, the SRC method will judge that the test sample belongs to the *i*th class. SRC is a classical representation method, which boosts the research of representation theory and applications. Based on SRC, many other representation methods are proposed, a lot of application problems, such as face recognition, digit recognition, signal processing, etc are also solved by representation methods.

Although SRC is effective for handling pattern classification problems [7-16], there are also some problems about SRC [2], [4]. First, the SRC method only looks for the sparsest representation of a test sample, however, the sparsest representation does not mean obtaining the highest right classification rate. Under some conditions, some test samples are misclassified by SRC. Second, the solution of SRC is obtained by l_1 norm minimization, which cannot obtain the closed form solution. Thus, the computational complexity of SRC is a little high.

In view of the advantages and disadvantages of SRC, some papers proposed many other representation methods, Chi and Porikli[2] proposed a Collaborative Representation Optimized Classifier (CROC). The CROC method combines the advantages of Nearest Subspace Classifier (NSC) and Collaborative Representation Classifier (CRC). Elhamifar and Vidal [17], [18] proposed a Block-Sparse representation for face recognition. They casted the classification problem as a structured sparse representation problem, they much emphasized the structured property of training samples. Zhang et al. [4] argued that not the sparse representation, but the usage of collaborative representation is more important for the success of the SRC. They proposed a kind of Collaborative Representation Classifier (CRC) method, by using l_2 norm minimization, CRC method can obtain closed form solution. Combined the advantages of KNN and SRC, Zhang and Yang proposed the KNN-SRC method[19]. First, for every test sample, KNN-SRC chooses the K nearest neighbors form all the training samples, Then, KNN-CRC makes these K nearest neighbors as the training samples. Finally, the classifier computes the residual for every class, if the *i*th residual is the smallest, the classifier will judge that the test sample belongs to the *i*th class.

These representation methods are all effective for handling classification problems, however, they all have advantages and disadvantages. For the Block-Sparse representation method, the solution of it is also obtained by l_1 norm minimization, the computational complexity of Block-Sparse representation is still a little high. For the CRC method, under some conditions, some test samples are also misclassified by CRC method. For the KNN-SRC method, the solution of KNN-SRC is also obtained by l_1 norm minimization, its computational complexity is also a little high. Furthermore, under some conditions, some test samples are also misclassified by KNN-SRC.

KNN is a conventional method for classification, which is familiar to us. And CRC is an another representation method for classification, the solution of CRC is obtained by l_2 norm minimization, thus, the computational complexity of CRC is low. In this paper, combined the KNN and CRC, a new classification method is proposed, namely KNN-CRC method. This method combines the advantages of KNN and CRC, KNN-CRC firstly chooses the K nearest neighbors for every test sample as the training samples, then the solution of KNN-CRC is obtained by l_2 norm minimization, it is a closed form solution, and the size of K is much smaller than the total number of all training samples.

Thus, the computational complexity of KNN-CRC is much lower than SRC, KNN-SRC and CRC. Furthermore, the extensive face and digit classification experiments clearly show that the proposed method can obtain very competitive results compared with other methods.

This paper is organized as follows. In Section 2, We review the KNN and CRC methods. In Section 3, the KNN-CRC method is proposed and the performance evaluation of the proposed method is given. Many face and digit classification experiments are conducted in Section 4. Section 5 concludes this paper.

2. KNN and CRC methods

2.1 Multi-Class Classification Problem

Assume there are *L* know classes, for every class, there are n_i training samples $\{b_{ij} \in R^m\}_{j=1}^{n_i}$ in the *i*th class, which formed a matrix as $B_i \square [b_{i1}, b_{i2}, ..., b_{in_i}] \in R^{m \times n_i}$. *B* is denoted by the collection of all training samples: $B = [B_1, B_2 ... B_L]$. If given a test sample $y \in R^m$, the aim of multi-class classification task is to judge the test sample **y** belongs to which class[1], [2], [4], [20-26].

2.2 K Nearest Neighbors (KNN) Classification Method

The nearest neighbor classifier was proposed by Hart and Cover for solving the classification problems. It was improved to the K Nearest Neighbors classifier immediately [19]. Assume there are K classes, for every class, there are n_i training samples $\{b_{ij} \in R^m\}_{j=1}^{n_i}$ in the ith class, $B_i = [b_{i1}, b_{i2}, ..., b_{in_i}] \in R^{m \times n_i}$ constitute the training samples of the ith class. Given a test sample y, it is easy to find its nearest neighbors b_{ir} in every class. The square of Euclidean distance between **y** and the ith class is obtained as:

$$d_i(y) = \|y - b_{ir}\|_2^2 \ (1 \le i \le L)$$

Suppose the distance between **y** and the *i*th class is the minimal distance, then the1-NN classifier will identify that the test sample y belongs to the *i*th class. 1-NN is a conventional and easy method for classification. K Nearest Neighbor classification (KNN) method is the improvement of the 1-NN algorithm. First, given a test sample y, the KNN classifier chooses the K nearest neighbors between the test sample and all the training samples. Second, assume that there are k_i samples from the *i*th class, If $k_j = \max_i k_i (1 \le i \le L)$, the KNN classifier will identify that the test sample y belongs to the *j*th class. KNN is also a conventional method for classification. However, the extensive experiments show that under many conditions, the accurate classification rates of KNN are not high.

2.3 Collaborative Representation Classifier (CRC)

The Sparsest solution to $y = Bx$ can be obtained by solving the following problem:

$$P_{l_0}: \quad \min \|x\|_0 \quad \text{s.t.} \quad y = Bx.$$

However, the P_{l_0} problem is a NP hard problem. Fortunately, Bruckstein et al. [23] proved that P_{l_0} problem can be substituted by P_{l_1} problem. From l_1 norm minimization, the sparsest solution $y = Bx$ of can be also obtained. The P_{l_1} program is as follows:

$$P_{l_1}: \quad \min \|x\|_1 \quad \text{s.t.} \quad y = Bx \ .$$

However, from l_1 norm minimization, it can not obtain the closed form solution. The computational complexity of l_1 norm minimization is a little high. Thus, some papers proposed other method by using l_2 norm minimization. Collaborative Representation Classifier (CRC) is proposed by Zhang and Yang [4], which is a typical method by l_2 norm minimization. The steps of CRC algorithm are as follows [4]:

Task: Find the solution of $P_{l_2}: \min_x \|x\|_2 \ \text{s.t.} \ y = Bx$

1)**Input:**

A matrix concatenated by training samples $B = [B_1, B_2, ...B_k] \in R^{m \times n}$ for k classes, a test sample $y \in R^m$.

2) Solve the l_2 norm minimization problem:

$x_2 = \arg\min_x \|x\|_2 \ \text{s.t.} \ y = Bx$

3) Compute the residuals:

$r_i(y) = \|y - B\delta_i(x_1)\|_2^2, \ \text{for } i=1,2,\ldots,k.$

4) **Output:**

$identity(y) = \arg\min_i r_i(y)$.

From l_2 norm minimization, a closed form solution can be obtained, which gives $x=(B^T B+\lambda I)^{-1} B^T y$. Its computational complexity is low. Although CRC is useful for classification, it also has some problems. Extensive experiments show that, under some conditions, some test samples are misclassified by CRC. Thus, combined the advantages of KNN and CRC, this paper proposes a local method, namely Collaborative Representation Classifier based on K Nearest Neighbors (KNN-CRC) method. The computational complexity of KNN-CRC is lower than SRC and CRC. Furthermore, under many conditions, the classification results of KNN-CRC are better than KNN and CRC.

3. Collaborative Representation Classifier based on K Nearest Neighbors

3.1 KNN-CRC Method

Combined the advantages of KNN and CRC, this paper proposes a local CRC method, namely KNN-CRC method. The basic steps of KNN-CRC are as follows. First, given a test sample, find its K nearest neighbors from all the training samples. Second, make the K nearest neighbors as the dictionary, represent the test sample with the dictionary. It can also obtain the closed solution from l_2 norm minimization. Third, compute the residual between the test sample and the each class in the dictionary. Finally, the KNN-CRC classifier can identify the test sample belongs to which class.

Specifically, the representation using KNN-CRC method is as follows:

$$x=\arg\min_x \|x\|_2 \quad \text{s.t. } y=\tilde{B}x, \tag{1}$$

where **y** denotes a test sample. $\tilde{B}$ denotes the dictionary, which is composed of the K nearest neighbors chosen from all the training samples. Unlike CRC, the KNN-CRC method uses the K nearest neighbors as the dictionary. However, CRC method uses all the training samples as the dictionary. The size of K is much smaller than the total number of all training samples, thus, the computational complexity of KNN-CRC is much lower than SRC and CRC. The extensive experiments also clearly show that the proposed method can obtain very competitive classification results compared with other methods.

From KNN-CRC method, the solution to (1) is

$$x=(\tilde{B}^T\tilde{B}+\lambda I)^{-1}\tilde{B}^T y.$$

Using Lagrange multiplier, a relaxed form of (1) can be obtained as:

$$x=\arg\min_x \left\|y-\tilde{B}x\right\|_2^2+\lambda\left\|x\right\|_2^2. \tag{2}$$

Taking derivative of (2) with respect to *x*, it can be easily obtained that:

$$-2\tilde{B}^T y+\tilde{B}^T\tilde{B}x+\tilde{B}^T\tilde{B}x+2\lambda x=0, \tag{3}$$

from (3), then the solution is obtained as:

$$x=(\tilde{B}^T\tilde{B}+\lambda I)^{-1}\tilde{B}^T y.$$

3.2 KNN-CRC Classifier

From the KNN-CRC method, the solution can be obtained. The solution is represented as $x=[x_{\lambda_1},\ldots,x_{\lambda_j},\ldots,x_{\lambda_k}]$, where x_{λ_j} is the part of coefficients corresponding to the j^{th} class in **x**. The *j*th block of **y** is defined as $y_j=\tilde{B}_{\lambda_j}x_{\lambda_j}, 1\le j\le L$.

For $j=1,\ldots,k$, this classifier will compute the residual $r_j=\left\|y-y_j\right\|_2^2=\left\|y-\tilde{B}_{\lambda_j}x_{\lambda_j}\right\|_2^2$.It can find the smallest residual easily. If the *j*th residual is the smallest, it will identify that the test sample **y** belongs to the *j*th class.

4. Experiments

In this section, some experiments on digit recognition and face recognition are presented to show the results of classification. These experiments focus on the property evaluation of KNN-CRC and other methods on the digit recognition and face recognition. Three databases, including AR[1], [2], [4], Extended Yale-B[1], [2], [4] and MNIST Handwritten Digits database[2], are used to test the performance of KNN-CRC and other methods, including NN, SRC and CRC.

KNN-CRC is proposed in this paper. NN is the nearest neighbors algorithm, which is a conventional method for classification. SRC is the classical sparse representation method [1]. CRC is the collaborative representation based classification with regularized least square[4]. Our experiments focus on the property comparison of KNN-CRC and other methods.

4.1 Face Recognition

The KNN-CRC and other methods are tested for comparing the recognition rate. Recognition rate denotes how many test samples can be classified correctly for all the test samples. Higher recognition rate means the property of this method is better. In our experiments, the Eigenface is used as preprocessing in feature extraction.

4.1.1 AR Database

The AR database contains about 4000 frontal images for 126 individuals[1], [2], [4].

These images are captured under different facial disguises, illuminations and expressions. The images are cropped to size 60 × 43. A few samples of AR database are shown in Figure 1. Table 1 and Figure 2 show the recognition rates versus feature dimension by NN, SRC, CRC and KNN-CRC.

Figure 1. Some training samples from the AR database

Dimension	45	50	55	60	65	70	75	80
NN	65.52%	66.38%	66.81%	67.24%	67.38%	67.38%	67.38%	68.10%
SRC	78.97%	79.26%	81.55%	84.12%	84.26%	84.12%	85.69%	86.70%
CRC	76.40%	79.26%	80.40%	83.26%	83.41%	84.26%	86.55%	86.84%
KNN-CRC (K=150)	80.83%	82.26%	83.12%	84.84%	85.27%	85.69%	85.84%	86.41%
KNN-CRC (K=200)	80.69%	81.69%	82.69%	84.55%	85.55%	86.41%	86.84%	87.55%
KNN-CRC (K=300)	81.16%	81.83%	83.98%	85.12%	86.12%	87.84%	87.84%	88.70%

Table 1. The face recognition rates of different methods on the AR database

4.1.2 Extended Yale-B Database

The Extended Yale-B database contains 2414 frontal face images of 38 individuals [1], [2], [4]. These samples were cropped and normalized to 54 × 48. A few samples of Extended Yale-B database are shown in Figure 3. Table 2 and Figure 4 shows the recognition rates versus feature dimension by NN, SRC, CRC and KNN-CRC.

4.2 Digit Recognition

The MNIST database is used to test the property of these methods. The dimension of each image is 28 × 28. Every image, which is an 8 bit gray scale image from 0 to 9[2]. For the MNIST handwritten digits database, which has a training set of 60,000 samples, and a test set of 10,000 samples of each class. For our experiment, 10 training samples are randomly selected from each class, 10 test samples are also randomly selected from each class. A few images of MNIST database are shown in Figure 5. Table 3 and Figure 6 show the recognition rates versus feature dimension by NN, SRC, CRC and KNN-CRC

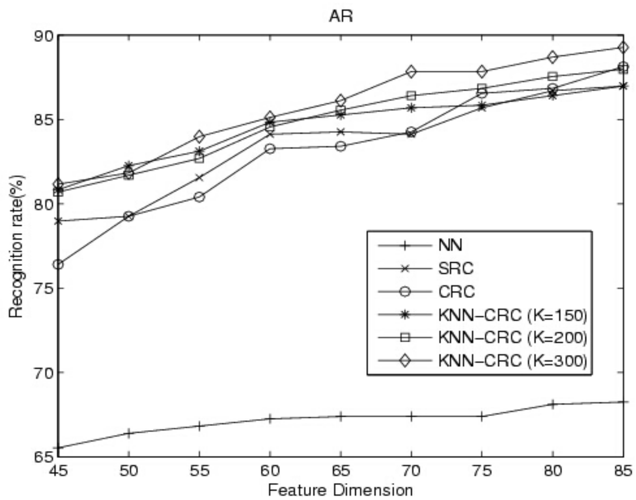

Figure 2. Recognition results on the AR database for different methods

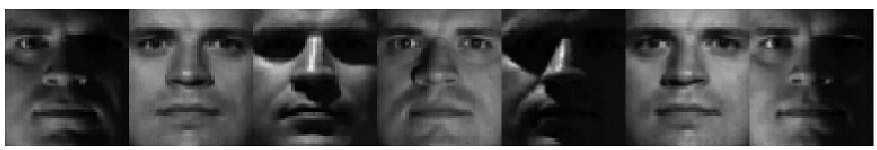

Figure 3. Some training samples from the Extended Yale-B database

Dimension	85	90	95	105	110	130	180
NN	70.33%	70.78%	71.32%	72.14%	72.60%	73.50%	75.05%
SRC	95.01%	95.46%	95.64%	95.46%	95.64%	95.82%	96.55%
CRC	95.01%	95.46%	95.46%	95.64%	95.83%	96.01%	96.82%
KNN-CRC (K=150)	95.19%	95.64%	95.74%	96.01%	96.01%	96.64%	96.73%
KNN-CRC (K=200)	95.55%	95.46%	95.83%	95.28%	96.28%	96.55%	97.19%
KNN-CRC (K=300)	95.1%	95.46%	95.46%	96.10%	96.37%	96.19%	96.92%

Table 2. The face recognition rates of different methods on the Extended Yale-B database

5. Conclusions

In this paper, a new classification method is proposed, which is different from SRC and CRC method, namely collaborative representation classifier based on K nearest neighbors (KNN-CRC) method. KNN-CRC combines the advantages of KNN and CRC methods, the solution of KNN-CRC is obtained from the l_2 norm minimization, it is a closed form solution. From the results of experiments, compared with other representation methods, KNN-CRC method has very competitive recognition rates. Furthermore, with the KNN-CRC method, the training samples of a test sample are from the K nearest neighbors of the test sample. The number of the K nearest neighbors is much fewer than all the training samples. A closed form solution is also obtained from KNN-CRC. Thus, the computational complexity of KNN-CRC is much lower than SRC and CRC.

From the digit recognition and face recognition experiments, it clearly showed that the proposed KNN-CRC method can obtain very competitive results compared with SRC, CRC and other representation methods.

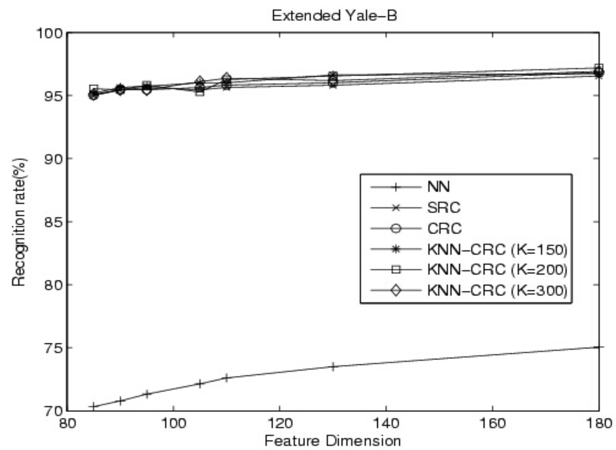

Figure 4. Recognition results on the Extended Yale-B database for different methods

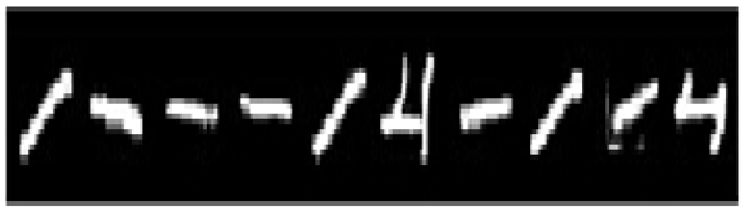

Figure 5. Some training samples from the MNIST database

Dimension	15	25	35	45	55	65	75	85	95
SRC	59%	60%	61%	55%	59%	58%	60%	58%	55%
CRC	63%	60%	61%	59%	59%	57%	59%	57%	56%
KNN-CRC (K=95)	61%	58%	58%	57%	57%	56%	56%	55%	54%

Table 3. The digit recognition results of different methods on the MNIST database

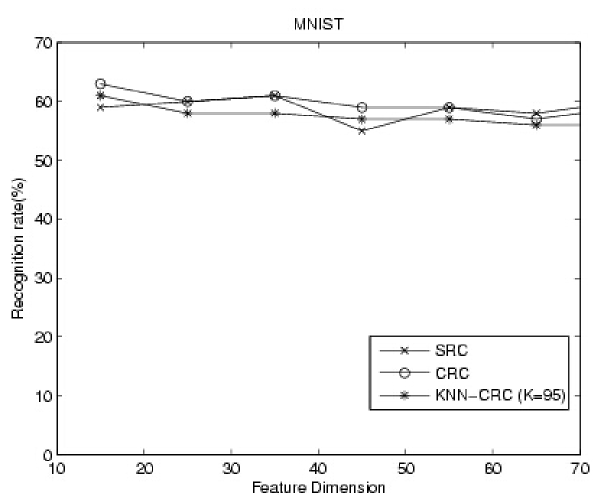

Figure 6. Recognition results on the MNIST database for different methods

Acknowledgements

This work was supported by the key fund project of Sichuan Educational department (NO.14ZA0005).

References

[1] Wright, J., Yang, A. Y., Sastry, S. S., Ma, Y. (2009). Robust face recognition via sparse representation. *Pattern Analysis and Machine Intelligence, IEEE Transactions on,* 31(2) 210–227.

[2] Chi, Y., F., Porikli. (2012). Connecting the dots in multi-class classiûcation: From nearest subspace to collaborative representation. *In*: 2012 IEEE Conference on, *Computer Vision and Pattern Recognition (CVPR)*, 3602–3609, .

[3] Roweis, S.T., Saul, L.K. (2000). Nonlinear dimensionality reduction by locally linear embedding. Science, 290 (5500), 2323–2326.

[4] Zhang, L., Yang, M., Feng, X. (2011). Sparse representation or collaborative representation: Which helps face recognition?. Computer Vision (ICCV), 2011 *IEEE International Conference on*, 471–478, 2011.

[5] Pati, Y. C., Rezaiifar, R., Krishnaprasad, P. (1993). Orthogonal matching pursuit: Recursive function approximation with applications to wavelet decomposition. Signals, Systems and Computers, 1993 Conference Record of The Twenty-Seventh Asilomar Conference on, 40–44, .

[6] Chen, S., Donoho Basis pursuit, D. (1994). Signals, Systems and Computers, 1994 Conference Record of the Twenty-Eighth Asilomar Conference on, 41–44.

[7] Elhamifar, E., (2011). Sparse manifold clustering and embedding. *Advances in Neural Information Processing Systems*, 24, 55–63.

[8] Elhamifar, E. Vidal, R. (2012). Sparse subspace clustering: Algorithm, theory, and applications. arXiv preprint arXiv:1203.1005.

[9] Fuchs, J. J. (2012). Recovery of exact sparse representations in the presence of bounded noise. *Information Theory, IEEE Transactions on*, 51 (10), 3601–3608.

[10] Guha, T., Ward, R. K. (2012). Learning sparse representations for human action recognition. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 34 (8), 1576–1588.

[11] Mairal, J., Bach, Ponce, J. Sapiro, G. (2009). Online dictionary learning for sparse coding, *In*: Proceedings of the 26^{th} Annual International Conference on Machine Learning, 689–696.

[12] Saul, L. K., Roweis, S. T. (2003). Think globally, fit locally: unsupervised learning of low dimensional manifolds, *The Journal of Machine Learning Research*, 4, 119–155.

[13] Soltanolkotabi, M., Elhamifar, E., Candes, E. J. (2013). Robust Subspace Clustering. arXiv preprint arXiv:1301.2603.

[14] Stojnic, M., Parvaresh, F., Hassibi, B. (2009). On the reconstruction of block-sparse signals with an optimal number of measurements, n*Signal Processing, IEEE Transactions on,* 57(8), 3075–3085.

[15] Tropp, J. A. (2004). Greed is good: Algorithmic results for sparse approximation. *Information Theory, IEEE Transactions on*, 50 (10), 2231–2242.

[16] Wright, J., Ma, Y., Mairal , J., Sapiro, G. ., Huang, T. S., Yan, S. (2010). Sparse representation for computer vision and pattern recognition. *In*: Proceedings of the IEEE, 98 (6), 1031–1044.

[17] Elhamifar, E., Vidal, R. (2011). Robust classiûcation using structured sparse representation.Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on, 1873–1879.

[18]. Elhamifar E, Vidal, R. (2012). Block-sparse recovery via convex optimization. *Signal Processing, IEEE Transactions on*, 60 (8) 4094–4107.

[19] Zhang, N., Yang, J. K. (2010). nearest neighbor based local sparse representation classiûer. Pattern Recognition (CCPR), 2010 Chinese Conference on, IEEE, 1-5.

[20] Adler, A., Elad, M., Y. Hel-Or Fast subspace clustering via sparse representations. Department of Computer Science,

Technion, Tech. Rep.

[21] Adler, A., Elad, M., Hel-Or. Y. (2011). Fast subspace clustering via sparse representations. Department of Computer Science, Technion, Tech. Rep.

[22] Aharon, M., Elad, M., Bruckstein , A. (2006). K-svd: an algorithm for designing overcomplete dictionaries for sparse representation. *Signal Processing, IEEE Transactions on*, 54 (11), 4311–4322.

[23] Bruckstein, A. M., Donoho, D. L., Elad, M. (2009). From sparse solutions of systems of equations to sparse modeling of signals and images. *SIAM Review*, 51 (1), 34–81.

[24] Cheng, B., Yang, J., Yan, S., Fu, Y., Huang, T. S. (2010). Learning with l1-graph for image analysis. *Image Processing, IEEE Transactions on*, 19 (4), 858–866.

[25] Cotter, S.F. (2010). Sparse representation for accurate classiûcation of corrupted and occluded facial expressions.Acoustics *Speech and Signal Processing (ICASSP), 2010 IEEE International Conference on*, 838–841.

[26] Elad, M., Aharon, M. (2006). Image denoising via sparse and redundant representations over learned dictionaries. *Image Processing, IEEE Transactions on*, 15 (12), 3736–3745.