KERNEL COLLABORATIVE REPRESENTATION-BASED CLASSIFIER FOR FACE RECOGNITION

Biao Wang[1,3,4], *Weifeng Li*[1,3,4], *Norman Poh*[2], *and Qingmin Liao*[1,3,4]

1. Department of Electronic Engineering/Graduate School at Shenzhen, Tsinghua University, China
2. Department of Computing, University of Surrey, United Kingdom
3. Tsinghua-PolyU Biometric Joint Laboratory, Tsinghua University, China
4. Shenzhen Key Laboratory of Information Science and Technology, Guangdong, China

ABSTRACT

Recent research has shown that collaborative representation-based classifier (CRC) can lead to promising results for the classification of face images. However, CRC is conducted in the original image space rather than the nonlinear high dimensional feature space in which features belonging to the same class are better grouped together and thus can be easily separable. To address this problem, this paper presents a novel classifier, Kernel Collaborative Representation-based Classifier (KCRC), by incorporating the kernel trick into the framework of CRC. Extensive experiments on both the AT&T and the FERET face databases demonstrate the priority of KCRC to CRC and several state-of-the-art methods.

Index Terms— Face recognition, classifier, collaborative representation, sparse representation, kernel trick.

1. INTRODUCTION

During the last several decades, automatic face recognition has been extensively studied in computer vision communities [1]. And it used to be well-acknowledged that a well-designed feature extractor, such as Principal Component Analysis (PCA) [2] or Linear Discriminant Analysis (LDA) [3], is crucial for a successful face recognition system. However, very recently, a debate on the significance of feature extractor is aroused by the pioneering work of Wright et al. [4], in which they conclude that what really matters is not the choice of the feature space but the design of the classifier and the dimensionality of feature space. This claim is supported by the surprising experimental results that even with unorthodox features like downsampled images and random projections, Sparse Representation-based Classifier (SRC) can outperform several state-of-the-art face recognition schemes, e.g., the nearest neighbor classifiers and the nearest subspace classifiers with carefully selected features.

At the very beginning, researchers believed that it's the underlying l_1-norm sparsity in face images that improves the robustness and accuracy of the classification, which boosts the research of sparsity-based face classification. Yang et al. [5] give a comprehensive evaluation of several l_1-norm minimization techniques. Wagner et al. [6] designed a real-world face recognition system based on SRC. Also, several kernel SRC schemes [7][8][9] have been proposed to address the nonlinearity of real-world face images. A recent survey of sparse representation for computer vision and pattern recognition tasks can be found in [10]. Of late, in [11], Shi et al. argued that the sparsity assumption underpinning much of the aforementioned work is not supported by the data, which means SRC cannot be guaranteed to get the desired performance. Furthermore, Zhang et al. [12] demonstrated both theoretically and empirically that it's not the l_1-norm sparsity constraint but the collaborative representation (CR) that truly improves the FR performance. Inspired from this, CRC is cast into a regularized least square problem, which is much more efficient than SRC without sacrificing the performance.

In this paper, we propose a novel kernel collaborative representation-based classifier (KCRC) by introducing the kernel trick, which can better group and separate the data with intrinsic nonlinear structures, e.g., face images [13]. In KCRC, the data in the input image space are implicitly mapped into a high or even infinite dimensional kernel feature space by utilizing some nonlinear mapping associated with a kernel function. However, it's not practical for us to solve the corresponding regularized least square problem in the feature space since we can only access the feature space by the kernel functions. To make this optimization problem feasible, we take advantage of kernel principal component analysis (KPCA) to conduct dimensionality reduction in the feature space. Experimental results on both the AT&T and the FERET databases demonstrate the priority of KCRC to several state-of-the-art algorithms in terms of both accuracy and efficiency.

The remainder of this paper is organized as follows. A brief overview of CRC is given in Section 2. Then in Section 3, we describe the details of the proposed KCRC method, including both the objective function, its implementation and

This work was supported by the Shenzhen-Hongkong Innovation Circle Project No. ZYB200907070030A.

the relationship to prior work. In Section 4, we use KCRC for face recognition and list the performance comparisons between KCRC and CRC on both the AT&T and the FERET face databases. Section 5 concludes the paper.

2. COLLABORATIVE REPRESENTATION-BASED CLASSIFIER

For face recognition tasks, we denote the gallery samples as $\mathbf{A} = [\mathbf{A}_1, \mathbf{A}_2, \cdots, \mathbf{A}_c] \in \mathbf{R}^{d\times N}$, in which $d = w \times h$ is the image dimension, c is the number of subjects, $\mathbf{A}_i = [\mathbf{a}_{i,1}, \mathbf{a}_{i,2}, \cdots, \mathbf{a}_{i,n_i}]$ is the collection of the gallery samples for subject i, and $N = \sum_{i=1}^{c} n_i$ is the total amount of samples. In SRC [4], a new test sample $\mathbf{y} \in \mathbf{R}^d$ can be sparsely coded by the following l_1-minimization optimization problem:

$$\hat{\boldsymbol{\alpha}} = \arg\min_{\boldsymbol{\alpha}} ||\mathbf{y} - \mathbf{A}\boldsymbol{\alpha}||_2^2 + \lambda\,||\boldsymbol{\alpha}||_1\,, \tag{1}$$

where λ is a regularized scalar. The decision is made in favor of the class with minimum reconstruction error:

$$Identity(\mathbf{y}) = \arg\min_{i} ||\mathbf{y} - \mathbf{A_i}\delta_i(\hat{\boldsymbol{\alpha}})||_2\,, \tag{2}$$

where $i = 1, ..., c$, and $\delta_i(\hat{\boldsymbol{\alpha}}) \in \mathbf{R}^c$ is a vector whose only nonzero entries in $\hat{\boldsymbol{\alpha}}$ are associated with subject i.

Wright et al. [4] attributed that the success of SRC to the l_1-norm which ensures that the resultant coding coefficients $\boldsymbol{\alpha}$s are sparse. In contrast, the authors of [12] argued both theoretically and empirically that collaborative representation-based classification (CRC) is crucial and regularizing α using l_2 norm can result in similar performance. The coding model of CRC can be expressed by:

$$\hat{\alpha} = \arg\min_{\boldsymbol{\alpha}} ||\mathbf{y} - \mathbf{A}\boldsymbol{\alpha}||_2^2 + \lambda\,||\boldsymbol{\alpha}||_2^2\,. \tag{3}$$

The decision of CRC is ruled as follows:

$$Identity(\mathbf{y}) = \arg\min_{i}\{||\mathbf{y} - \mathbf{A_i}\delta_i(\hat{\boldsymbol{\alpha}})||_2\,/\,||\delta_i(\hat{\boldsymbol{\alpha}})||_2\}. \tag{4}$$

Although the l_1-norm minimization problem is extensively studied and lots of fast numerical algorithms have been proposed, it is still very computationally demanding. Compared with the l_1-norm regularized SRC, the l_2-norm regularized CRC is significantly more computationally efficient without sacrificing the face recognition accuracy. Extensive experimental results in [12] demonstrate that CRC is up to 1600 times faster than SRC with similar recognition rate.

3. PROPOSED METHOD

3.1. The Kernel Trick

In machine learning field, the kernel trick is a well-known technique which can generalize a linear algorithm to its nonlinear counterpart without ever having to compute the mapping explicitly. The motivation is that after such nonlinear mapping, the samples in the high dimensional feature space can better group together and thus can be easily separable. The kernel trick has been successfully applied to several algorithms, such as SVM [14], KPCA and KLDA [13].

The trick to avoid explicit mapping is to take advantage of algorithms which only require inner products for vectors in feature space $\mathscr{F} \in \mathbf{R}^D$, and the mapping is chosen such that these high-dimensional inner products can be calculated in the original space $\mathscr{O} \in \mathbf{R}^d (d << D)$ in terms of a kernel function. Mercer kernel, which is a continuous, symmetric and positive semidefinite function, is popular in kernel methods. A Mercer kernel $k(\cdot,\cdot) : \mathscr{O} \times \mathscr{O} \to \mathscr{F}$ can be expressed as

$$k(\mathbf{x}, \mathbf{y}) =< \Phi(\mathbf{x}), \Phi(\mathbf{y}) >= \Phi(\mathbf{x})^T\Phi(\mathbf{y}), \tag{5}$$

where $\mathbf{x}$ and $\mathbf{y}$ are any two sample points in $\mathscr{O}$, $\Phi(\cdot) : \mathscr{O} \to \mathscr{F}$ is the the implicit nonlinear mapping function corresponding to the kernel function $k(\cdot,\cdot)$. In kernel methods, $\Phi(\cdot)$ is typically unknown and the only way to access the feature space is via $k(\cdot,\cdot)$. The commonly used Mercer kernels include the linear kernel, the polynomial kernels, the Gaussian radial basis function (RBF) kernel, and the χ^2 kernel [7].

3.2. Kernel Collaborative Representation-Based Classifier (KCRC)

Suppose there exists a nonlinear feature mapping function $\Phi(\cdot) : \mathbf{R}^d \to \mathbf{R}^D (d << D)$. It maps the test sample $\mathbf{y}$ and the collaborative representation dictionary $\mathbf{A}$ to the high dimensional feature space: $\mathbf{y} \to \Phi(\mathbf{y})$, $\mathbf{A} = [\mathbf{a}_{1,1}, \mathbf{a}_{1,2}, \cdots, \mathbf{a}_{1,n_1}, \cdots\cdots, \mathbf{a}_{c,1}, \mathbf{a}_{c,2}, \cdots, \mathbf{a}_{c,n_c}] \to \Phi(\mathbf{A}) = [\Phi(\mathbf{a}_{1,1}), \cdots, \Phi(\mathbf{a}_{c,n_c})]$. We substitute the mapped test sample and the dictionary to the formulation of CRC (Eq. (3)) and arrive at the kernel collaborative representation classifier (KCRC):

$$\hat{\boldsymbol{\alpha}} = \arg\min_{\alpha} ||\Phi(\mathbf{y}) - \Phi(\mathbf{A})\boldsymbol{\alpha}||_2^2 + \lambda\,||\boldsymbol{\alpha}||_2^2\,. \tag{6}$$

In our work, we take advantage of the RBF kernel due to its excellent performance reported in current literature [13][14][15]:

$$k(\mathbf{x}, \mathbf{y}) = exp(-\gamma\,||\mathbf{x} - \mathbf{y}||^2), \tag{7}$$

where $\gamma > 0$ is a parameter for the RBF kernels.

Obviously, we cannot easily solve the optimization problem for Eq. (6) directly due to the implicity of the mapping function $\Phi(\cdot)$. Fortunately, as indicated in [9], such an optimization problem can be addressed by resorting to kernel-based dimensionality reduction methods. In our work, we choose KPCA because it can even address the single-sample per person problem, which is a common face recognition application. In the following, we will discuss how KPCA makes Eq. (6) feasible.

Eq. (6) is the l_2-norm regularized solution to the following equation:

$$\Phi(\mathbf{y}) = \Phi(\mathbf{A})\boldsymbol{\alpha}. \tag{8}$$

Denote $\mathbf{P} = [\mathbf{P}_1, \mathbf{P}_2, \cdots, \mathbf{P}_d] \in \mathbf{R}^{D \times m}$ as a transformation matrix which projects the high dimensional sample points in $\mathscr{F}$ into a low dimensional subspace with dimensionality m. Recall that $\mathbf{A} = \left[\mathbf{a}^1, \mathbf{a}^2, \cdots, \mathbf{a}^N\right]$. For KPCA, each projection vector $\mathbf{P}_j$ can be deemed as a linear combination of samples in $\mathscr{F}$, i.e.,

$$\mathbf{P}_j = \Phi(\mathbf{A})\boldsymbol{\beta}_j, \tag{9}$$

where each $\boldsymbol{\beta}_j \in \mathbf{R}^N, j = 1, 2, \cdots, m (m < N)$[1] is a normalized eigenvector associated with the j-th largest eigenvalues corresponding to the following Eigenvalue problem [15]:

$$N\lambda\boldsymbol{\beta} = \boldsymbol{K}\boldsymbol{\beta}, \tag{10}$$

where $\boldsymbol{K} = (\Phi(\mathbf{A}))^T\Phi(\mathbf{A})$ is the kernel Gram matrix and $K_{ij} = k(\mathbf{a}^i, \mathbf{a}^j)$. Denote $\mathbf{V} = [\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \cdots, \boldsymbol{\beta}_m]$, from Eq. (9) we can obtain

$$\mathbf{P} = \Phi(\mathbf{A})\mathbf{V}, \tag{11}$$

Performing dimensionality reduction with $\mathbf{P}$ to both sides of Eq. (8) we can get

$$\mathbf{P}^T\Phi(\mathbf{y}) = \mathbf{P}^T\Phi(\mathbf{A})\boldsymbol{\alpha}. \tag{12}$$

Substitute Eq. (11) into Eq. (12) we have

$$\mathbf{V}^T\mathbf{k}(\cdot, \mathbf{y}) = \mathbf{V}^T\mathbf{K}\boldsymbol{\alpha}. \tag{13}$$

where $\mathbf{k}(\cdot, \mathbf{y}) = [k(\mathbf{a}^1, \mathbf{y}), \cdots, k(\mathbf{a}^N, \mathbf{y})]^T = \Phi(\mathbf{A})^T\Phi(\mathbf{y})$. Now, we can see, by taking advantage of KPCA, given a kernel function $k(\cdot, \cdot)$, $\mathbf{V}$, $\mathbf{K}$ and $\mathbf{k}(\cdot, \mathbf{y})$ are easy to be calculated, which makes Eq. (13) feasible. The corresponding l_2-norm regularized optimization problem is equivalent to:

$$\hat{\boldsymbol{\alpha}} = \arg\min_{\alpha} \left|\left|\mathbf{V}^T\mathbf{k}(\cdot, \mathbf{y}) - \mathbf{V}^T\mathbf{K}\boldsymbol{\alpha}\right|\right|_2^2 + \lambda \left|\left|\boldsymbol{\alpha}\right|\right|_2^2. \tag{14}$$

The solution can be easily determined according to [16]:

$$\hat{\boldsymbol{\alpha}} = ((\mathbf{V}^T\mathbf{K})^T(\mathbf{V}^T\mathbf{K}) + \lambda \cdot \mathbf{I})^{-1} \cdot (\mathbf{V}^T\mathbf{K}) \cdot (\mathbf{V}^T\mathbf{k}(\cdot, \mathbf{y})). \tag{15}$$

3.3. Implementation and Relationship to Related Work

The proposed KCRC algorithm is summarized in Algorithm 1. In Eq. (15), let $\mathbf{Q} = ((\mathbf{V}^T\mathbf{K})^T(\mathbf{V}^T\mathbf{K}) + \lambda \cdot \mathbf{I})^{-1} \cdot (\mathbf{V}^T\mathbf{K})$. Clearly, $\mathbf{Q}$ is independent of $\mathbf{y}$ such that it could be pre-calculated as a projection matrix. When a query sample $\mathbf{y}$ comes, we should just simply project $(\mathbf{V}^T\mathbf{k}(\cdot, \mathbf{y}))$ onto $\mathbf{Q}$, which makes KCRC very fast.

KCRC is a direct extension of CRC [12] by incorporating the kernel trick, therefore is more suitable for samples with inherent nonlinear structures. Compared with SRC [4] [6] and its kernel extensions [7] [8] [9], KCRC inherits the merits of CRC and avoids solving the time-demanding l_1-norm minimization problem, thus should be more efficient and easy to implement.

[1] m is the number of the reserved KPCA dimension.

Algorithm 1 Kernel Collaborative Representation Classifier (KCRC)

Input: A test image vector $\mathbf{y} \in \mathbb{R}^{d \times 1}$, dictionary $\mathbf{A} = [\mathbf{a}^1, \mathbf{a}^2, \cdots, \mathbf{a}^N] \in \mathbf{R}^{d \times N}$, the regularization parameter λ, and the reserved KPCA dimensionality m.

Output: The identity of $\mathbf{y}$.

1: Compute the kernel Gram matrix $\mathbf{K}$ where $K_{ij} = k(\mathbf{a}^i, \mathbf{a}^j)$, and a vector $k(\cdot, \mathbf{y}) = [k(\mathbf{a}^1, \mathbf{y}), \cdots, k(\mathbf{a}^N, \mathbf{y})]^T$.
2: Compute the matrix $\mathbf{V} = [\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \cdots, \boldsymbol{\beta}_m]$ by solving the eigenvalue problem: $N\lambda\boldsymbol{\beta} = \boldsymbol{K}\boldsymbol{\beta}$.
3: Normalize the columns of $(\mathbf{V}^T\mathbf{K})$ and $(\mathbf{V}^T\mathbf{k}(\cdot, \mathbf{y}))$ to have unit l_2-norm.
4: Compute the KCRC coding coefficients $\hat{\boldsymbol{\alpha}} = ((\mathbf{V}^T\mathbf{K})^T(\mathbf{V}^T\mathbf{K}) + \lambda \cdot \mathbf{I})^{-1} \cdot (\mathbf{V}^T\mathbf{K}) \cdot (\mathbf{V}^T\mathbf{k}(\cdot, \mathbf{y}))$.
5: Compute the regularized residuals $r_i = \left|\left|(\mathbf{V}^T\mathbf{k}(\cdot, \mathbf{y})) - (\mathbf{V}^T\mathbf{K})_i\delta_i(\hat{\boldsymbol{\alpha}})\right|\right|_2 / \left|\left|\delta_i(\hat{\boldsymbol{\alpha}})\right|\right|_2$
6: Output the identity of $\mathbf{y}$ as $identity(\mathbf{y}) = \arg\min_i r_i(\mathbf{y})$.

4. EXPERIMENTAL RESULTS

In this section, experiments are conducted on two publicly available face databases, namely, the AT&T [17] and the FERET face databases [18] to illustrate the effectiveness of the proposed KCRC algorithm. We will also compare our algorithm with several state-of-the-arts: Linear Regression Classifier (LRC) [19], CRC [12], SRC [4] and KSRC [9]. We implement the KCRC algorithm using 64-bit Matlab platform on a PC with 2.13 GHz Intel I3 CPU and 4 GB memory. The regularization parameter λ is set as $1e$-4 as indicated in [12]. All the competing algorithms are implemented on the same platform with parameters set as the original paper recommended.

4.1. Results on the AT&T Face Database

The AT&T face database includes 40 subjects with 10 images per person. Several expression variations exist in this database, e.g., open or closed eyes, smiling or nonsmiling, with or without glasses. It also has a maximum rotation up to 20 degree with some scale variations of about 10 percent. Figure 1 shows all the samples of the same person.

A simple observation is that the performance of all the evaluated algorithms depend on the number of training samples for each subject. In our experiments, we randomly partition the 10 images into two non-overlapped parts, one part serves as the gallery and the other serves as the probe set. We varied the sample number of the gallery from 1 to 5. All experiments are conducted with images downsampled to an order of 10×5, and we set $\gamma = 2$ for the RBF kernel, $m = 80$. The average recognition rate versus the number of gallery samples is illustrated in Table 1, from which two useful conclusions can be drawn:

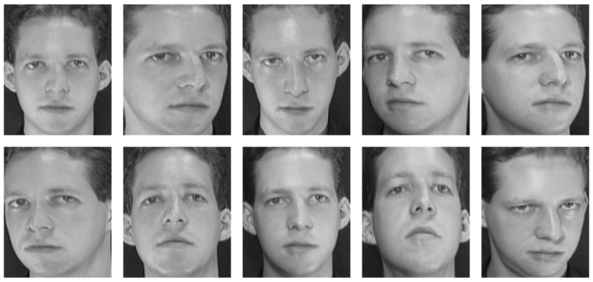

Fig. 1. Sample face images in the AT&T face database.

Table 1. Average recognition rates (%) versus the number of gallery sample per person on the AT&T Face Database.

Methods	1	2	3	4	5
LRC [19]	73.33	80.00	82.14	85.83	92.50
SRC [4]	74.44	86.56	87.50	92.08	93.50
KSRC [9]	75.56	87.19	88.93	92.92	95.50
CRC [12]	75.83	88.44	88.57	92.50	94.50
KCRC	76.11	90.94	91.79	94.50	95.00

(a) When there are not enough gallery samples per subject (say, less than 4), the sparsity assumption does not hold and SRC, KSRC perform worse than CRC and KCRC. This further implies that it is truly the collaborative representation but not the sparsity property that improves face recognition accuracy.

(b) When there are enough gallery samples per subject (say, more than 3), the sparsity assumption intends to approximately hold and the performance of SRC and KSRC are comparable to CRC and KCRC.

(c) With pose variations existed, a test sample may not be easy to be linearly coded by the galleries, which can explain why KCRC and KSRC have better performance than their linear counterpart, respectively.

4.2. Results on the FERET Face Database

The number of subjects in the AT&T face database is limited. To further validate the proposed method on practical applications, we have conducted experiments on the FERET database, one of the most commonly used large-scale face databases. We have used the standard FERET protocol to conduct our experiments. The gallery set Fa consists of $1,196$ images of $1,196$ subjects. There are four probe sets: Fb (different expressions with gallery, $1,195$ images of $1,196$ subjects), Fc (different illumination conditions with gallery, 194 images of 194 subjects), Dup I (images taken later in time, 722 images of 243 subjects), Dup II (images taken at least 18 months after the corresponding gallery, 234 images of 75 subjects). All face images are properly aligned, cropped and resized to 128×128 with the centers of the eyes fixed at (29,34) and (99,34). No further preprocessing is performed. Figure 2 shows samples of the same person from the five sets.

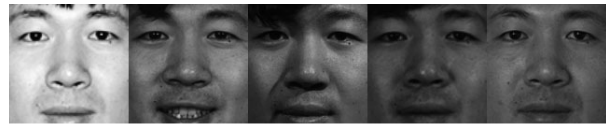

Fig. 2. Sample face images in the FERET face database. (a) Fa (b) Fb (c) Fc (d) Dup I (e) Dup II.

Table 2. Recognition rates (%) on the FERET Face Database.

Methods	Fb	Fc	Dup I	Dup II
LRC [19]	74.53	69.56	53.20	49.98
SRC [4]	82.43	82.99	60.54	59.82
KSRC [9]	83.12	84.10	60.23	59.82
CRC [12]	80.67	77.84	60.95	58.80
KCRC	81.92	82.25	61.63	60.98

All experiments are conducted with images downsampled to an order of 24×24, and we set $\gamma = 0.1$ for the RBF kernel, $m = 450$. The recognition rates of all the algorithms on the four subsets are illustrated in Table 2, from which we can see that KCRC consistantly performs better than CRC on all subsets and is comparable to KSRC in average.

4.3. Efficiency

Efficiency is important for real-time face recognition applications. As illustrated in Algorithm 1, for each probe, the proposed method mainly requires computing a kernel function with each gallery sample and then conducts CRC. We compare the efficiency of KCRC with other competing algorithms experimentally by measuring the CPU time on the FERET database. The results are illustrated in Table 3, from which we could conclude that CRC is about 4 times faster than KCRC, while KCRC is 3.6 and 17.5 times faster than SRC and KSRC.

Table 3. Average CPU time (in ms) per probe on the FERET face database.

LRC [19]	SRC [4]	KSRC [9]	CRC [12]	KCRC
14.7	215.8	1056.4	15.2	60.5

5. CONCLUSION

We propose a novel nonlinear classifier, KCRC, for face recognition. It is an extension to the classical CRC algorithm and is more suitable for pattern classification in nonlinear space by incorporating the kernel trick into the CRC framework. Experimental results on the AT&T and the FERET face databases demonstrate the priority of KCRC to CRC and several state-of-the-art methods in terms of accuracy and efficiency.

6. REFERENCES

[1] W. Zhao, R. Chellappa, A. Rosenfeld, and P.J. Phillips, "Face recognition: A literature survey," *ACM Computing Surveys*, pp. 399–458, 2003.

[2] M. A. Turkand A. P. Pentland, "Eigenfaces for recognition," *J. Cognitive Neuroscience*, vol. 3, no. 1, pp. 71–86, 1991.

[3] P. Belhumeur, J. Hespanha, and D. Kriegman, "Eigenfaces vs. fisherfaces: Recognition using class specific linear projection," *IEEE Trans. on Patt. Analysis and Machine Intell.*, vol. 19, no. 7, pp. 711–720, 1997.

[4] J. Wright, A. Yang, A. Ganesh, S. Sastry, and Y. Ma, "Robust face recognition via sparse representation," *IEEE Trans. on Patt. Analysis and Machine Intell.*, vol. 31, no. 2, pp. 210–227, 2009.

[5] A. Yang, S. Sastry, A. Ganesh, and Y. Ma, "Fast $l1$-minimization algorithms and an application in robust face recognition: A review," in *ICIP*, 2010, pp. 1849–1852.

[6] A. Wagner, J. Wright, A. Ganesh, Z. Zhou, H. Mobahi, and Y. Ma, "Towards a practical face recognition system: Robust alignment and illumination by sparse representation," *IEEE Trans. on Patt. Analysis and Machine Intell.*, vol. 34, no. 2, pp. 372–386, 2012.

[7] C. Kang, S. Liao, S. Xiang, and C. Pan, "Kernel sparse representation with local patterns for face recognition," in *ICIP*, 2011, pp. 3009–3012.

[8] S. Gao, I. Tsang, and L. Chia, "Kernel sparse representation for image classification and face recognition," in *ECCV*, 2010, pp. 1–14.

[9] L. Zhang, W. Zhou, P. Chang, J. Liu, Z. Yan, T. Wang, and F. Li, "Kernel sparse represenation-based classifier," *IEEE Trans. on Signal Process.*, vol. 60, no. 4, pp. 1684–1695, 2012.

[10] J. Wright, Y. Ma, J. Mairal, G. Sapiro, T. Huang, and S. Yan, "Sparse representation for computer vision and pattern recognition," *Proceedings of IEEE, Special Issue on Applications of Compressive Sensing & Sparse Representation.*, vol. 98, no. 6, pp. 1031–1044, 2010.

[11] Q. Shi, A. Eriksson, A. Hengel, and C. Shen, "Is face recognition really a compressive sensing problem?," in *CVPR*, 2011, pp. 553–560.

[12] L. Zhang, M. Yang, and X. C. Feng, "Sparse representation or collaborative representation which helps face recognition," in *ICCV*, 2011, pp. 471–478.

[13] M. Yang, "Kernel eigenfaces vs. kernel fisherfaces: Face recognition using kernel methods," in *Int. Conf. Automatic Face and Gesture Recognition*, 2002, pp. 215–220.

[14] V. N. Vapnik, *Statistical Learning Theory*, New York: Wiley-Interscience, 1998.

[15] B. Scholkopf, A.J. Smola, and K.-R. Muller, "Nonlinear component analysis as a kernel eigenvalue problem," *Neural Computing*, vol. 10, no. 5, pp. 1299–1319, 1998.

[16] G.A.F. Seber, *Linear Regression Analysis*, Wiley-Interscience, 2003.

[17] F. Samaria and A. Harter, "Parameterisation of a stochastic model for human face identification," in *Proc. Second IEEE Workshop Applications of Computer Vision*, 1994.

[18] P. J. Phillips, H. Moon, P. Rizvi, and P. Rauss, "The feret evaluation method for face recognition algorithms," *IEEE Trans. on Patt. Analysis and Machine Intell.*, vol. 22, no. 10, pp. 1090–1104, 2000.

[19] I. Naseem, R. Togneri, and M. Bennamoun, "Linear regression for face recognition," *IEEE Trans. on Patt. Analysis and Machine Intell.*, vol. 32, no. 11, pp. 2106–2112, 2010.