Fast and Robust Face Recognition via Coding Residual Map Learning based Adaptive Masking

Meng Yang*, ZhizhaoFeng*, Simon C. K. Shiu, Lei Zhang[1]

Dept. of Computing, The Hong Kong Polytechnic University, Hong Kong, China

Abstract: *Robust face recognition (FR) is an active topic in computer vision and biometrics, while face occlusion is one of the most challenging problems for robust FR. Recently, the representation (or coding) based FR schemes with sparse coding coefficients and coding residual have demonstrated good robustness to face occlusion; however, the high complexity of l_1-minimization makes them less useful in practical applications. In this paper we propose a novel coding residual map learning scheme for fast and robust FR based on the fact that occluded pixels usually have higher coding residuals when representing an occluded face image over the non-occluded training samples. A dictionary is learned to code the training samples, and the distribution of coding residuals is computed. Consequently, a residual map is learned to detect the occlusions by adaptive thresholding. Finally the face image is identified by masking the detected occlusion pixels from face representation. Experiments on benchmark databases show that the proposed scheme has much lower time complexity but comparable FR accuracy with other popular approaches.*

Keywords: *Robust face recognition, coding residual map, adaptive masking*

*The first two authors contribute equally to this work.

[1] Corresponding author. Email: cslzhang@comp.polyu.edu.hk.

1 Introduction

Face recognition (FR) has been an active research topic in the fields of computer vision and biometrics for more than three decades [1-7, 26-28]. Early FR methods analyze the geometric features of facial images, such as the location of nose, eyes, and mouth, [1-2]. However, these methods are very sensitive to the changes in illumination and facial expression. To solve this problem, the appearance based FR methods extract some holistic features from the original face image vectors for matching. Many subspace learning based holistic feature extraction methods have been developed, including Eigenfaces [29], Fisherfaces [36], Local Preserving Projection (LPP) [4], 2D-PCA [31], Independent component analysis (ICA) [30], Sparsity Preserving Projections [5], etc.

The subspace learning based FR methods are simple to implement, fast, and work well for face images without occlusion. However, in many practical FR applications, the face images are occluded (e.g., with sunglasses and scarf), and the conventional subspace based methods [4-6, 29-31, 36] cannot deal with occlusion well. Representation based face classification methods have been recently proposed [7, 9-11, 24-25] for robust FR. One typical example is the sparse representation based classification (SRC) scheme [7], which hypothesizes that the multiple training images of a subject could well reconstruct other samples from the same subject. In SRC, the query image (or its holistic feature vector) is sparsely coded over the training images (or their holistic feature vectors), and the identity of the query sample is assigned to the class which yields the minimum coding residual. In the coding process of SRC, the l_1-norm sparsity is imposed on the coding coefficients to increase discrimination. Furthermore, to increase the robustness to occlusions, in SRC the l_1-norm is also used to characterize the coding residual, which makes SRC more time consuming.

Apart from SRC, other methods which are robust to face occlusions have also been developed, such as Eigenimages [37-38], probabilistic local approach [39], auto-associative network [40], and occlusion estimation via partial filters [41]. In [40] an auto-associative network was used to detect outliers and the occluded face was then completed by replacing the occluded pixels by the outputs of the auto-associative network. Compared to these methods [37-41], SRC could handle more general types of occlusions, including real disguise, continuous occlusion, pixel-wise corruption with unknown location and unknown intensity, etc.

FR methods based on local features (e.g., local binary pattern [42], line edge map [44], parabola edge map [45], and oriented edge magnitude feature [46]) have also been proposed. For example, Gao and Leung [44] proposed to represent face images by line edge maps and recognize face images by minimizing the line segment Hausdorff distance. In order to encode not only local structure but also global structure of a face image, Chen and Gao [43] proposed the Stringface based on face edge/line features, and transformed FR as a string-to-string matching problem. Although edge feature such as Stringface [43] is robust to some types of face variations (e.g., lighting changes) to some extent, it cannot handle facial occlusion with random block and random pixel corruption, where accurate edge detection is very difficult to achieve [47].

SRC based FR has been attracting much attention from researchers due to its promising results to recognize occluded face images. Many related works have been developed, such as the l_1-graph for image classification [8], kernel based SRC [9], Gabor feature based SRC [10, 48], robust sparse coding (RSC) [24], robust alignment with sparse and low rank decomposition [25], joint dimension reduction and dictionary learning [49], face and ear multimodal biometric system [50], etc. In particular, the RSC method [24] has shown excellent results in FR with various occlusions. From the viewpoint of maximal likelihood estimation, the l_1-norm or l_2-norm characterization of the representation residual is only optimal when the residual follows Laplacian or Gaussian distribution, thus Yang *et al.* [24] used a robust regression function to measure the representation residual. Although the RSC method has high recognition accuracy, it is also very time-consuming, like SRC.

Very recently, Zhang *et al*. [12] verified that it is not the sparse representation (i.e., the l_1-norm sparsity imposed on the coding coefficients) but the collaborative representation (i.e., using the training samples from all classes to collaboratively represent the query face image) that plays the key role in SRC for face classification. Zhang *et al*. then proposed to use the l_2-norm to regularize the representation coefficients, and the so-called collaborative representation based classification with regularized least square (CRC_RLS) method achieves similar accuracy to SRC but with significantly less computational cost [12].

In CRC_RLS, the coding residual is modeled by l_2-norm. Although it achieves reasonable performance in FR with real disguise (e.g., sunglass and scarf) on the AR database, it cannot robustly deal with other types of face occlusions (e.g., random corruption, random occlusion). If we use l_1-norm to model the representation residual for robustness to occlusion, this will increase much the computational cost. It is therefore very important for us to develop a robust but fast FR scheme to handle face occlusion. This is the motivation of our work.

When a query face image is represented as a linear combination of non-occluded face images, the representation residual actually can be used to detect many face occlusion pixels because those occluded pixels tend to have larger reconstruction errors than the normal non-occluded pixels. Therefore, if we can detect and remove these pixels from the face representation, more robust FR results can be obtained. However, different face features, such as eyes, nose, mouth and cheek, will have different representation accuracy and thus we cannot use a single threshold to detect the face occlusion pixels. In this paper, we propose to learn a face coding residual distribution map from the training samples by coding them over a dictionary, which is also learned from the training samples. Then for a given query image, it is coded over the learned dictionary, and is then adaptively masked based on its coding residual and the pre-learned coding residual map. By removing the detected occlusions, the query face image can then be robustly represented and classified. Although the idea of our occlusion detection method is quite simple, our experimental results showed that the proposed scheme can achieve competitive results with other well-known and robust FR methods in terms of both speed and accuracy.

The rest of this paper is organized as follows. Section 2 briefly introduces the main procedures of SRC and CRC. The proposed error map learning method is presented in detail in Section 3. Section 4 introduces the spatially adaptive masking. Section 5 conducts extensive experiments to test the proposed method and compare it with other well-known methods. Finally, Section 6 concludes this paper.

2 SRC and CRC

In [7], Wright *et al.* proposed a sparse representation based classification (SRC) scheme for face recognition. Let $\boldsymbol{A} = [\boldsymbol{A}_1, \boldsymbol{A}_2, \ldots, \boldsymbol{A}_c]$ be the set of training samples from all the c classes, where $\boldsymbol{A}_i$ is the subset of the training samples from class i. Denote by $\boldsymbol{y}$ a query sample. In SRC, $\boldsymbol{y}$ is sparsely coded over $\boldsymbol{A}$ via l_1-minimization

$$\hat{\boldsymbol{\alpha}} = \arg\min_{\boldsymbol{\alpha}} \left\{ \|\boldsymbol{y} - \boldsymbol{A}\boldsymbol{\alpha}\|_2^2 + \gamma \|\boldsymbol{\alpha}\|_1 \right\} \tag{1}$$

where γ is a scalar constant. The classification is done by

$$\text{identity}(\boldsymbol{y}) = \arg\min_i \{e_i\} \tag{2}$$

where $e_i = \|\boldsymbol{y} - \boldsymbol{A}_i\hat{\boldsymbol{\alpha}}_i\|_2^2$, $\hat{\boldsymbol{\alpha}} = [\hat{\boldsymbol{\alpha}}_1; \ldots; \hat{\boldsymbol{\alpha}}_c]$ and $\hat{\boldsymbol{\alpha}}_i$ is the coefficient vector associated with class i.

To make SRC robust to face occlusion, an identity matrix $\boldsymbol{I}$ is introduced as a dictionary to code the occluded pixels [7]:

$$\hat{\boldsymbol{\alpha}} = \arg\min_{\boldsymbol{\alpha},\boldsymbol{\beta}} \|[\boldsymbol{\alpha}; \boldsymbol{\beta}]\|_1 \quad \text{s.t.} \quad \boldsymbol{y} = [\boldsymbol{A}, \boldsymbol{I}] \cdot [\boldsymbol{\alpha}; \boldsymbol{\beta}] \tag{3}$$

Actually Eq. (3) is essentially equivalent to:

$$\hat{\boldsymbol{\alpha}} = \arg\min_{\boldsymbol{\alpha}} \{\|\boldsymbol{y} - \boldsymbol{A}\boldsymbol{\alpha}\|_1 + \boldsymbol{\gamma} \|\boldsymbol{\alpha}\|_1\} \tag{4}$$

Since both the coding coefficient and coding residual are modeled by l_1-norm, the complexity of SRC is very high.

Though the role of sparse representation in robust FR is much emphasized in [7], Zhang *et al.*[12] verified that it is the collaborative representation mechanism (i.e., representing the query image collaboratively using all training samples) in SRC that affects much FR. The l_1-norm sparsity on the coding coefficients is not that crucial to FR. Zhang *et al.* proposed to simply use the regularized least square to code the query image, and the so-called CRC_RLS scheme represents $\boldsymbol{y}$ as [12]

$$\hat{\boldsymbol{\alpha}} = \arg\min_{\boldsymbol{\alpha}} \left\{ \|\boldsymbol{y} - \boldsymbol{A}\boldsymbol{\alpha}\|_2^2 + \gamma \|\boldsymbol{\alpha}\|_2^2 \right\} \tag{5}$$

The classification in CRC_RLS is similar to that in SRC.

The complexity of CRC_RLS is significantly lower than SRC because $\hat{\boldsymbol{\alpha}}$ can be simply calculated as $\hat{\boldsymbol{\alpha}} = \boldsymbol{P} \cdot \boldsymbol{y}$, where $\boldsymbol{P} = \left(\boldsymbol{A}^T \boldsymbol{A} + \gamma \cdot \boldsymbol{I}\right)^{-1} \boldsymbol{A}^T$ can be pre-calculated and it is independent of $\boldsymbol{y}$. However, CRC_RLS has no special settings to deal with occlusion and it is less robust to occluded faces.

3 Coding residual map learning

3.1 Dictionary learning

In representation based FR, the recognition is performed by coding the query image over a dictionary. One straightforward way is to use the original training samples as the dictionary, such as in SRC [7] and CRC_RLS [12]. Since the training samples are generally non-occluded face images, the occluded pixels in a query image usually cannot be well reconstructed by the non-occluded samples and thus they will have big reconstruction errors. Intuitively, we can use the coding residual to detect the occluded pixels (for example, by setting a detection threshold), and then mask the detected occlusion pixels from the face coding to achieve robust FR.

However, the face has various features, e.g., eyes, nose, mouth and cheek, which will have different variances of coding residuals. Therefore, it is hard to use a global threshold to effectively detect the occlusions in different facial areas. In order to make the occlusion detection more accurate, knowing the coding residual variances of different facial features is important such that spatially adaptive occlusion detection can be carried out.

To achieve the above objective, we can learn a dictionary from the training samples, and use this dictionary to code the training samples. The variances of coding residuals can then be computed to build the coding residual map. By coding a query image over this dictionary and with the learned coding residual map, the occlusion pixels can then be adaptively detected. In addition, compared with using the original training samples as the naïve dictionary for face representation, dictionary learning can also bring advantages such as removing noise and trivial structures for more accurate face representation, as well as making the representation more discriminative.

Many dictionary learning methods have been proposed for image processing [13-14, 32-33] and pattern recognition [15-16, 34, 49] in the past. In [16], a Fisher discrimination dictionary learning

(FDDL) method was proposed for sparse representation based image recognition. Inspired by FDDL and considering that the sparsity on the coding coefficients is not that critical for FR [12], we propose a simpler dictionary learning model. Denote by $\boldsymbol{D}$= [$\boldsymbol{D}_1$, $\boldsymbol{D}_2$, …, $\boldsymbol{D}_c$] the dictionary to be learned, where $\boldsymbol{D}_i$ is the class-specified sub-dictionary associated with class i. The dictionary $\boldsymbol{D}$ is learned from the training dataset $\boldsymbol{A}$. In general, we require that each column of the dictionary $\boldsymbol{D}_i$ is a unit vector, and the number of atoms in $\boldsymbol{D}_i$ is no bigger than the number of training samples in $\boldsymbol{A}_i$.

We denote by $\boldsymbol{X}_i$ the coding coefficient matrix of $\boldsymbol{A}_i$ over $\boldsymbol{D}$, and $\boldsymbol{X}_i = [\boldsymbol{X}_i^1; \ldots; \boldsymbol{X}_i^j; \ldots; \boldsymbol{X}_i^c]$, where $\boldsymbol{X}_i^j$ is the coding matrix of $\boldsymbol{A}_i$ over the sub-dictionary $\boldsymbol{D}_j$. We propose to learn the dictionary $\boldsymbol{D}$ by optimizing the following objective function:

$$J_{(\boldsymbol{D},\boldsymbol{X})} = \arg\min_{(\boldsymbol{D},\boldsymbol{X})} \sum\nolimits_{i=1}^{c} \left\{ R_i\left(\boldsymbol{D}\right) + \lambda_1 \left\|\boldsymbol{X}_i\right\|_F^2 + \lambda_2 \left\|\boldsymbol{X}_i - \bar{\boldsymbol{X}}_i\right\|_F^2 \right\} \tag{6}$$

where

$$R_i(\boldsymbol{D}) = \left\|\boldsymbol{A}_i - \boldsymbol{D}\boldsymbol{X}_i\right\|_F^2 + \left\|\boldsymbol{A}_i - \boldsymbol{D}_i\boldsymbol{X}_i^i\right\|_F^2 + \sum\nolimits_{\substack{j=1 \\ j\neq i}}^{c} \left\|\boldsymbol{D}_j\boldsymbol{X}_i^j\right\|_F^2 \tag{7}$$

and $\bar{\boldsymbol{X}}_i$ is the column mean matrix of $\boldsymbol{X}_i$, i.e., every column of $\bar{\boldsymbol{X}}_i$ is the mean vector $\boldsymbol{m}_i$ of all the columns in $\boldsymbol{X}_i$. The parameters λ_1 and λ_2 are positive scalar numbers to balance the F-norm terms in Eq. (6).

From Eq. (7), one can see that the term $R_i(\boldsymbol{D})$ ensures that the training samples from class i (i.e., $\boldsymbol{A}_i$) can be well reconstructed by the learned sub-dictionary $\boldsymbol{D}_i$, while they have small representation coefficients on the other sub-dictionaries $\boldsymbol{D}_j$, $j\neq i$. Therefore, the learned whole dictionary $\boldsymbol{D}$ will be discriminative in terms of reconstruction. On the other hand, the term $\left\|\boldsymbol{X}_i - \bar{\boldsymbol{X}}_i\right\|_F^2$ in Eq. (6) will make the representation of samples from the same class close to each other, reducing the intra-class variations. Finally, we use the F-norm, instead of the l_1-norm, to regularize the coding coefficients $\boldsymbol{X}$ in Eq. (6), and this significantly reduces the complexity of optimizing Eq. (6).

The objective function in Eq. (6) is a joint optimization problem of $\boldsymbol{D}$ and $\boldsymbol{X}$, and it is convex to $\boldsymbol{D}$ or $\boldsymbol{X}$ when the other is fixed. Like in many multi-variable optimization problems, we could solve

Eq.(6) by optimizing ***D*** and ***X*** alternatively from some initialization. Since in each step, the optimization is convex and all the terms involved are of *F*-norm, the optimization can be easily solved. The learned dictionary ***D*** will be different by using different settings of parameters λ_1 and λ_2. Our experimental results show that the final FR rates are not sensitive to λ_1 and λ_2 in a wide range. In our experiments, we set them as λ_1=0.001 and λ_2=0.002 for all the databases by experience.

3.2 Residual map learning

Once the dictionary ***D*** is learned from the training dataset ***A***, it can be used to code a given query sample ***y*** by

$$\hat{\boldsymbol{\alpha}} = \arg\min_{\boldsymbol{\alpha}} \left\{ \left\| \boldsymbol{y} - \boldsymbol{D}\boldsymbol{\alpha} \right\|_2^2 + \gamma \left\| \boldsymbol{\alpha} \right\|_2^2 \right\} \tag{8}$$

The solution $\hat{\boldsymbol{\alpha}}$ can be easily calculated as $\hat{\boldsymbol{\alpha}} = \boldsymbol{P} \cdot \boldsymbol{y}$, where the projection matrix $\boldsymbol{P} = \left(\boldsymbol{D}^T \boldsymbol{D} + \gamma \cdot \boldsymbol{I} \right)^{-1} \boldsymbol{D}^T$ can be pre-computed. We can then calculate the coding residual $\boldsymbol{e}_y = \boldsymbol{y} - \boldsymbol{D}\hat{\boldsymbol{\alpha}}$. When there are occlusions in the query image ***y***, its coding residual at the occluded locations will probably exceed the “normal range”. Therefore, if we know the “normal range”, more specifically the standard deviation, of each element of $\boldsymbol{e}_y$, we can adaptively detect the occlusions in ***y***.

Obviously, the deviation of coding residual will vary with different facial areas. In most cases, the areas such as eyes and mouth will have bigger residual than the areas such as cheek because they have more edge structures which are more difficult to reconstruct. This coding residual deviation map can be learned by coding the training samples in ***A*** over the dictionary ***D***. There is

$$\hat{\boldsymbol{\Lambda}} = \arg\min_{\boldsymbol{\Lambda}} \left\{ \left\| \boldsymbol{A} - \boldsymbol{D}\boldsymbol{\Lambda} \right\|_F^2 + \gamma \left\| \boldsymbol{\Lambda} \right\|_F^2 \right\} \tag{9}$$

Clearly, $\hat{\boldsymbol{\Lambda}} = \boldsymbol{P} \cdot \boldsymbol{A}$, and the coding residual matrix is $\boldsymbol{E} = \boldsymbol{A} - \boldsymbol{D}\hat{\boldsymbol{\Lambda}}$.

Each row of ***E***, denoted by $\boldsymbol{e}_k$, contains the coding residuals of all face samples at the same location *k*, and thus its standard deviation can be used to define the normal range of the coding residual at this location. Denote by

$$\sigma_k = std(\boldsymbol{e}_k) \tag{10}$$

the standard deviation of $\boldsymbol{e}_k$, and then all the σ_k together will build a coding residual map, which indicates the normal range of coding residual at each location. Certainly, from Eq. (9) we know that the residual map depends on the parameter γ, and Fig. 1 shows several residual maps calculated on the AR database [17] with different values of γ. It can be seen that the different face structures will have different residual deviation, while the residual map varies with γ. Therefore, a suitable γ must be determined for robust FR.

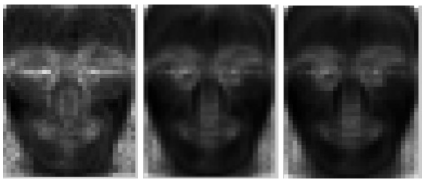

Fig. 1. Examples (by the AR database) of the learned coding residual map with different settings of γ. From left to right, γ=0.1, 1, 2, respectively.

We determine γ by checking which value of γ can make the coding in Eq. (9) optimal. It can be empirically found that the coding residuals in $\boldsymbol{E}$ and the coding coefficients in $\boldsymbol{\Lambda}$ are nearly Gaussian distributed. We assume that the residuals in $\boldsymbol{E}$ and the coefficients in $\boldsymbol{\Lambda}$ follow i.i.d. Gaussian distributions, respectively. Based on the *maximum a posterior* (MAP) principle, the desired coefficients $\boldsymbol{\Lambda}$ should make the probability $P(\boldsymbol{\Lambda}|\boldsymbol{A})$ maximized. According to the Bayesian formula and after some straightforward derivations, the parameter γ which could lead to the MAP solution of $\boldsymbol{\Lambda}$ should satisfy

$$\gamma_{opt} = \|\boldsymbol{E}\|_F^2 \big/ \|\boldsymbol{\Lambda}\|_F^2 \tag{11}$$

In implementation, we use a set of different values of γ to code $\boldsymbol{A}$ by Eq. (9), and the corresponding coefficient $\boldsymbol{\Lambda}$ and residual $\boldsymbol{E}$ can be obtained. Then we could check if the used γ is close enough to the associated γ_{opt} in Eq. (11). The γ which is close to its associated γ_{opt} the most is then used to compute the residual map elements σ_k.

4 Recognition with adaptive masking

4.1 Detecting the occlusion pixels

Once the residual map is learned, we can use it to detect the occluded outlier pixels in the query image $\boldsymbol{y}$ based on its coding residual $\boldsymbol{e}_y$. Intuitively, at location k, if $|\boldsymbol{e}_y(k)|$ is much bigger than σ_k in the residual map, it is highly possible that pixel k in $\boldsymbol{y}$ is occluded. It is empirically found that the coding residual at location k, i.e., $\boldsymbol{e}_k$, approximately follows zero-mean Gaussian distribution, while the shape of the Gaussian distribution is controlled by σ_k. Fig. 2 plots the histograms of $\boldsymbol{e}_k$ at three different types of areas, eye, nose and cheek, respectively. We can see that the distributions are Gaussian like, and the eye and nose regions have much higher standard deviation values than the smooth cheek area.

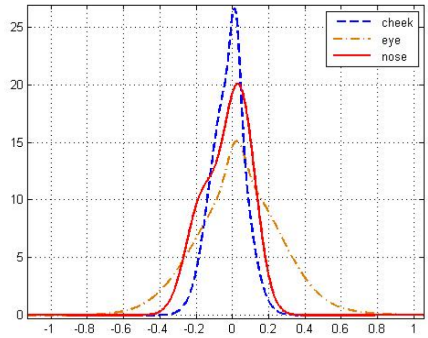

Fig. 2. Histograms of the coding residuals at regions of eye (in brown), nose (in red) and cheek (in blue), respectively.

It is known that most of the values of a Gaussian distribution will fall into the interval bounded by several times of its standard deviation. Therefore, if a pixel at location k is occluded, the coding residual at this location will often exceed the normal range, and $|\boldsymbol{e}_y(k)|>c\sigma_k$ is very likely to happen, where c is a constant. In this paper, we use the following simple rule

$$\textit{pixel } k \textit{ is occluded} \;\; \text{if } |\boldsymbol{e}_y(k)|>c\sigma_k \tag{12}$$

to detect the occlusions. This occlusion detection rule is rough, however, it makes the detection efficient and it is able to remove a large portion of occluded pixels and improve the recognition rate significantly, as will be seen in our experimental results.

4.2 Masking and coding

After detecting the occluded pixels in $\boldsymbol{y}$, we can partition the query image $\boldsymbol{y}$ into two parts: $\boldsymbol{y}$=[$\boldsymbol{y}_{nc}$;$\boldsymbol{y}_{oc}$], where $\boldsymbol{y}_{nc}$ denotes the non-occluded part and $\boldsymbol{y}_{nc}$ denotes the occluded part. Since each pixel in $\boldsymbol{y}$ has a corresponding row in the learned dictionary $\boldsymbol{D}$, we can accordingly partition the dictionary $\boldsymbol{D}$ into two parts, i.e., $\boldsymbol{D}$=[$\boldsymbol{D}_{nc}$; $\boldsymbol{D}_{oc}$], where $\boldsymbol{D}_{nc}$ is the sub-dictionary for $\boldsymbol{y}_{nc}$ and $\boldsymbol{D}_{oc}$ is for $\boldsymbol{y}_{oc}$.

Since the occlusion in the query image will reduce the FR accuracy, obviously we can exclude $\boldsymbol{y}_{oc}$ from coding and use only $\boldsymbol{y}_{nc}$ to recognize the identity of $\boldsymbol{y}$. Therefore, the coding after masking is performed as:

$$\hat{\boldsymbol{\alpha}}_{nc} = \arg\min_{\boldsymbol{\alpha}} \left\{ \left\| \boldsymbol{y}_{nc} - \boldsymbol{D}_{nc}\boldsymbol{\alpha} \right\|_2^2 + \gamma \left\| \boldsymbol{\alpha} \right\|_2^2 \right\} \tag{13}$$

The solution of Eq. (13) is $\hat{\boldsymbol{\alpha}}_{nc} = \boldsymbol{P}_{nc} \cdot \boldsymbol{y}_{nc}$ with $\boldsymbol{P}_{nc} = \left(\boldsymbol{D}_{nc}^T \boldsymbol{D}_{nc} + \gamma \cdot \boldsymbol{I} \right)^{-1} \boldsymbol{D}_{nc}^T$. Since $\boldsymbol{P}_{nc}$ depends on the input query sample $\boldsymbol{y}$ and it cannot be pre-computed, the calculation of $\hat{\alpha}_{nc}$ is the most time-consuming step of our proposed scheme. The detailed complexity analysis will be made in Section 5.5, and the running time comparison will demonstrate that the proposed scheme is much faster than state-of-the-art robust FR methods with comparable FR accuracy.

4.3 Classification

After $\hat{\boldsymbol{\alpha}}_{nc}$ is obtained by solving Eq. (13), the coding coefficient and class-specific coding residual can be used to determine the identity of query image $\boldsymbol{y}$. The class-specific coding residual can be calculated as $e_i = \left\| \boldsymbol{y}_{nc} - \boldsymbol{D}_{nc\_i} \hat{\boldsymbol{\alpha}}_{nc\_i} \right\|_2$, where $\boldsymbol{D}_{nc\_i}$ and $\hat{\boldsymbol{\alpha}}_{nc\_i}$ are the sub-dictionary and sub-coding vector associated with class i, respectively. Recall that in the dictionary learning stage in Section 3.1, we have also learned the mean coding vector $\boldsymbol{m}_i$ of each class. Denote by $\boldsymbol{m}_{nc\_i}$ the corresponding mean coding vector to the non-occluded face vector $\boldsymbol{y}_{nc}$. The distance between $\hat{\boldsymbol{\alpha}}_{nc\_i}$ and $\boldsymbol{m}_{nc\_i}$ can also help classifying $\boldsymbol{y}$. Let $g_i = \left\| \hat{\boldsymbol{\alpha}}_{nc} - \boldsymbol{m}_{nc\_i} \right\|_2$. Finally, we could fuse e_i and g_i for decision making:

$$f_i = e_i + w \cdot g_i \tag{14}$$

where weight w is a constant. The identity of the query image is then determined by: identity($\boldsymbol{y}$) = $\text{argmin}_i\{f_i\}$.

5 Experimental results

In this section, we perform extensive experiments on benchmark face databases to demonstrate the performance of our proposed algorithm. We first discuss the parameter selection in Section 5.1; in Section 5.2 we test the proposed algorithm on two databases (AR [17] and MPIE [18]) without disguise and occlusion; in Sections 5.3 and 5.4, we test the proposed method on three databases, AR, Extended Yale B and MPIE, with disguise and occlusion; finally, we will discuss the computational efficiency of the proposed method in Section 5.5.

5.1 Parameter selection

There are five parameters in our algorithm: λ_1 and λ_2 in Eq. (6), γ in Eq. (9) and Eq. (13), the constant c in Eq. (12) and the weight w in Eq. (14). Among these parameters, λ_1 and λ_2 are related to dictionary learning, and w is related to the distance measurement. Based on our experimental experience, we simply fix λ_1=0.001, λ_2=0.002, and w=0.1 on all datasets. As discussed in Section 3.2, the value of γ is automatically determined by Eq. (11) in the training phase.

The value of c is critical in detecting the occlusion points. If c is too small, too many points will be regarded as outlier; if c is too large, fewer outliers will be detected. Fig. 3 shows some detection examples. If no specific instruction, in our following experiments, c is set as 2, 1 and 1 in the AR, Extended Yale B and MPIE databases, respectively.

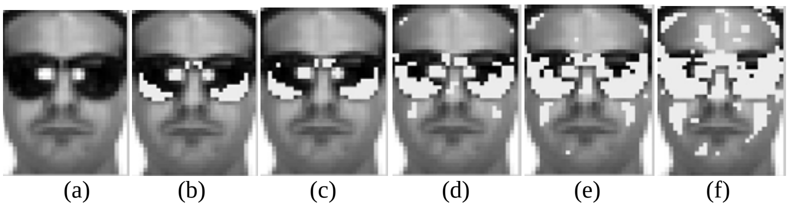

Fig. 3. Example of occlusion point detection. (a) is the original test image. From (b) to (f): the occlusion detection results by letting c=12,10,6,4,2, respectively.

5.2 Recognition without occlusion

Although our algorithm is mainly designed to handle occlusion, it can also be applied to recognize normal face images. In clean face images without occlusion or disguise, there are still some pixels which may lead to recognition error, and they can be viewed as outliers. We can detect these pixels by using our proposed method and exclude them from recognition. In this section, we evaluate our algorithm in the popular databases AR and MPIE.

a) AR database: The setting in our experiments is the same as in [7]. A subset in AR [17] that contains 50 males and 50 females with only illumination and expression variances is used. For each individual, the seven images from Section 1 are used as training samples while the other seven images from Section 2 are used for testing. The original face image is downsampled to 36×22 in our method. The best results of competing methods, including the nearest neighbor (NN) classifier, SRC [7], CRC_RLS [12], RSC [24] and the proposed method, are presented in Table 1. In addition, to more clearly show the advantage brought by the learned residual map, we also present the result of CRC_RLS coupled with an outlier detector by global thresholding. That is, we first remove the pixels whose representation residuals are larger than a predefined global threshold, and then classify the face image using the remaining part by CRC_RLS. We call this global thresholding based scheme CRC_RLS_GT. In the following experiments, we report the best results of CRC_RLS_GT by choosing a suitable threshold.

CRC_RLS, CRC_RLS_GT, and the proposed method have similar reconstruction strategy, while the proposed method achieves higher recognition rate. This demonstrates that the proposed outlier pixel detection method can remove some insignificant (even negative) pixels in the face images, and hence improve the recognition rate. Using a global threshold to remove the outlier pixels is not helpful, and CRC_RLS_GT has the same accuracy as CRC_RLS. Compared with NN and SRC, the proposed method also achieves about 23.8% and 1.8% higher recognition rate. The accuracy of the proposed algorithm is only slightly lower than RSC, whose complexity is much higher than our method (please refer to Section 5.5. for the running time comparison).

Table 1. Recognition rates on the AR database by different methods.

NN	SRC	CRC_RLS	CRC_RLS_GT	RSC	Proposed
71.3%	93.3%	93.7%	93.7%	96.0%	95.1%

b) Multi PIE database: In this experiment, all the 249 subjects in Session 1 in CMU Multi-PIE [18] are used. We use 7 frontal images with extreme illuminations {0, 1, 7, 13, 14, 16, 18} and neutral expression of each subject as the training set. For testing set, 4 images of illuminations taken with smile expressions from each person in the same session are used. The face images are directly down-sampled to 25×20. Table 2 shows the recognition rates of different methods. Again, the proposed method achieves the second best recognition rate, just after the RSC scheme.

Table 2. Recognition rates on the MPIE database by different methods.

NN	SRC	CRC_RLS	CRC_RLS_GT	RSC	Proposed
86.4%	93.9%	94.1%	94.1%	97.8%	94.4%

5.3 Recognition with real disguise

As in [7], a subset from the AR database, consisting of 50 male and 50 female subjects, is used here. 800 images (about 8 samples per subject) of non-occluded frontal views with various facial expressions in both sessions are used for training, while the samples with sun glasses and scarves (1 sample per subject) in both sessions are used for testing. Fig. 4 shows some example query images with disguise. The images are directly down-sampled to 42×30 with normalization.

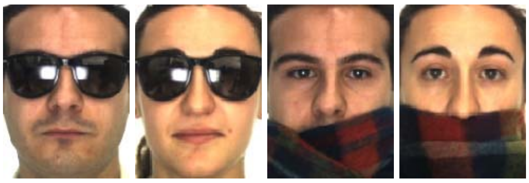

Fig. 4. Testing samples with sunglasses and scarves in the AR database.

The occlusion pixels detected by the proposed method are illustrated in Fig. 5. It can be seen that the proposed algorithm detects many outlier points, and also makes some wrong judgments. Fortunately, occluded parts are detected as outliers more often than non-occluded parts. For smaller ratio of occlusion (e.g., sunglasses disguise), the detection result is more accurate (see Figs. 5(a) and 5(c)); but when the occlusion ratio is large (e.g., the scarf disguise), the detection result is less accurate (see Figs. 5(b) and 5(d)). It should be stressed that our goal is to introduce a robust and efficient FR scheme instead of occlusion detection.

The recognition rates by competing methods are listed in Table 4. Although the occlusion detection is not accurate enough, the proposed scheme can still obtain much better results than all the competing methods except for RSC. By detecting outlier pixels with a global threshold, the CRC_RLS_GT method could improve the performance in some case (e.g., FR with sunglasses), while the proposed adaptive thresholding method can achieve 15% higher accuracy than CRC_RLS_GT in the sunglass case. Again, although the RSC method has the highest accuracy, we would like to emphasize that it has much higher complexity than the proposed method.

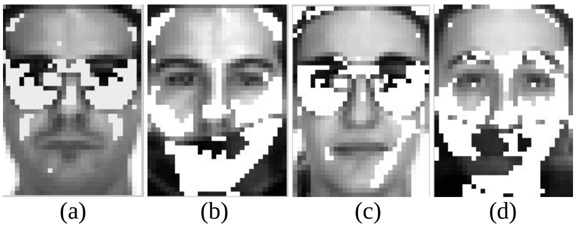

Fig. 5. The occlusion detection results of some testing samples in the AR database.

Table 3. Recognition results by different methods on the AR database with sunglasses and scarves disguise.

Algorithm	Sun-	Scarves
NN	70.0%	12.0%
SRC [7]	87.0%	59.5%
CRC_RLS [12]	68.5%	90.5%
CRC_RLS_GT	77.5%	90.5%
RSC [24]	99.0%	97.0%
Proposed	93.0%	90.5%

5.4 Recognition with random block occlusion

In this section, we test the robustness of our algorithm to random block occlusion. As in [7], Subsets 1 and 2 of the Extended Yale B database [6] are used for training and Subset 3 for testing. Each testing sample will be inserted an unrelated image as block occlusion, and the blocking ratio is from 10% to 50% as illustrated in Fig 6. The images are cropped and down-sampled to 48×42.

All training and testing samples are normalized to reduce the effect of illumination variance. The occlusion detection results of the example images in Fig. 6 are shown in Fig. 7. The recognition rates by the competing methods are listed in Table 4. We can see that when the block occlusion ratio is low, all methods can achieve good recognition accuracy; when the block occlusion ratio increases, the accuracy of NN, CRC_RLS and SRC will decrease rapidly, while RSC and the proposed method can still have good results. Meanwhile, the proposed method performs much better than the global thresholding based CRC_RLC_GT (e.g., over 5% and 7% improvements in 40% and 50% block occlusion, respectively). When the occlusion ratio is 50%, the proposed method surpasses SRC more than 10%, while being only about 6% lower than RSC.

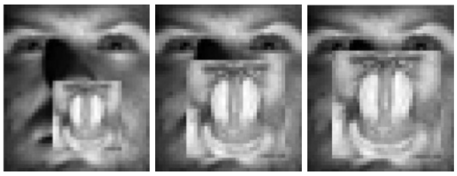

Fig. 6. Examples of random block occlusion in the Extended Yale B database. From left to right: occlusion ratio is 20%, 40%, 50%, respectively.

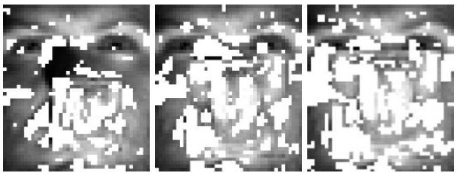

Fig. 7. Occlusion detection results of the samples in Fig. 6.

Table 4. Recognition results by different methods on the Extended Yale B database with various random occlusion ratios.

Occlusion ratio	10%	20%	30%	40%	50%
NN	90.1%	85.2%	74.2%	63.8%	48.1%
SRC [7]	100%	99.8%	98.5%	90.3%	65.3%
CRC_RLS [12]	99.8%	93.6%	82.6%	70.0%	52.3%
CRC_RLS_GT	100%	99.8%	96.9%	88.1%	70.9%
RSC [24]	100%	100%	99.8%	96.9%	83.9%
Proposed	100%	99.8%	98.5%	93.6%	77.9%

5.5 Recognition with random pixel corruption

In this section, we evaluate the robustness of the proposed method to random pixel corruption. As in [7], we still use Subsets 1 and 2 (717 images, normal-to-moderate lighting conditions) of the Extended Yale B database [6] for training, and use Subset 3 (453 images, more extreme lighting conditions) for testing. The images were resized to 96×84 pixels. For each testing image, we replaced a certain percentage of its pixels by uniformly distributed random values within [0, 255]. The corrupted pixels were randomly chosen for each test image and the locations are unknown to the algorithm. Some corrupted face images are shown in Fig. 8, from which we can observe that it is difficult to recognize the corrupted face images even for humans.

The recognition rates of all competing methods under different corruption ratios (from 0% to 70%) are listed in Table 5. Here we set c as 6 in all tests. The proposed method, SRC and RSC all achieve 100% FR rates when the corruption ratio is from 0% to 40%. When the corruption ratio is 50% or 60%, the performance of the proposed method is still very close to RSC and SRC. Meanwhile, the proposed method clearly outperforms NN, CRC_RLS, and CRC_RLS_GT.

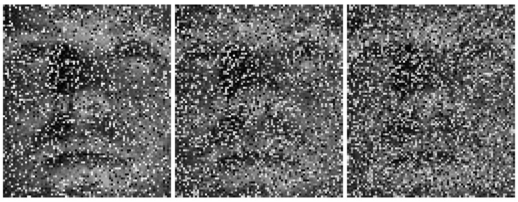

Fig. 8. Example face images with random pixel corruption in the Extended Yale B database. From left to right: corruption ratio is 40%, 50%, 60%, respectively.

Table 5. Recognition results by different methods on the Extended Yale B database with various random pixel corruption ratios.

Corruption	0%	10%	20%	30%	40%	50%	60%	70%
NN	94.0%	96.2%	96.5%	94.7%	82.3%	65.6%	40.2%	25.8%
SRC [7]	100%	100%	100%	100%	100%	100%	99.3%	90.7%
CRC_RLS[12]	100%	100%	100%	99.8%	98.9%	96.4%	79.9%	45.7%
CRC_RLS_GT	100%	100%	100%	100%	100%	98.5%	88.3%	58.3%
RSC[24]	100%	100%	100%	100%	100%	100%	100%	99.3%
Proposed	100%	100%	100%	100%	100%	98.9%	91.4%	65.3%

5.6 Complexity analysis

From the experimental results in Sections 5.2~5.5, we see that the proposed method's recognition accuracy is only slightly lower than RSC but much higher than SRC (in most cases) and CRC_RLS, which are among the state-of-the-art methods. Let's then compare the time complexity and running time between our method and SRC, CRC_RLS and RSC. (Note that the complexity of CRC_RLS_GT is the same as the proposed method.)

The computational cost in our method mainly comes from solving Eq.(13). The solution of Eq. (13) can be written as

$$\boldsymbol{D}_{nc}^{T} \cdot \boldsymbol{y}_{nc} = \left(\boldsymbol{D}_{nc}^{T} \boldsymbol{D}_{nc} + \gamma \cdot \boldsymbol{I} \right) \hat{\boldsymbol{\alpha}}_{nc} \tag{15}$$

Eq. (15) can be viewed as a linear equation system, where the coefficient $\hat{\boldsymbol{\alpha}}_{nc}$ is to be computed. Here we use the Conjugate Gradient Method (CGM) [19-20] to solve this linear equation system, which is much more efficient than solving the inverse problem in Eq. (15). Suppose that $\boldsymbol{D}_{nc}$ is an $N \times M$ matrix. In FR problems, usually we have $N>M$. The solution of Eq. (15) is a vector of length M. The computational complexity of CGM is $O(M)$ for each iteration, so the complexity of solving Eq. (15) is $O(KM)$ if there are totally K iterations. Since we do not need a very accurate solution to Eq. (15), we set K=30 as the maximum number of iterations in our experiments. The complexity for calculating $\boldsymbol{D}_{nc}^{T}\boldsymbol{D}_{nc}$ is $O(NM^2)$. Thus the total time complexity of our method is $O(NM^2+KM)$.

The following Tables 6~8 show the running time of CRC_RLS [12], RSC [24], SRC with l_1_*ls* [21] and SRC with fast homotopy [22] (for other fast l_1-norm minimization methods please refer to [23]), and our proposed algorithm in three experiments, which are conducted under MATLAB programming environment on a PC with Intel R Core 2 1.86 GHz CPU and 2.99 GB RAM. The settings are the same as those in previous sections, i.e., AR with sunglass disguise, Extended Yale B with 50% occlusion, and MPIE without occlusion. All samples are directly down-sampled from the original face images. The reported running time is the average time consumed by each testing sample.

Table 6. Recognition rates and running time on the AR database with sunglass disguise.

Algorithm	Recognition Rate	Running Time
SRC(l_1_*ls*)	87.0%	34.50s
SRC(homotopy)	86.4%	0.120s
CRC_RLS	68.5%	0.017s
RSC	99.0%	46.91s
Proposed	93.0%	0.141s

Table 7. Recognition rates and running time on the Extended YaleB database with 50% block occlusion.

Algorithm	Recognition Rate	Running Time
SRC(l_1_*ls*)	65.3%	53.34s
SRC(homotopy)	63.5%	0.115s
CRC_RLS	52.3%	0.034s
RSC	83.9%	73.36s
Proposed	77.9%	0.278s

Table 8. Recognition rates and running time on the MPIE database without occlusion.

Algorithm	Recognition Rate	Running Time
SRC(l_1_*ls*)	93.9%	59.25s
SRC(homotopy)	92.0%	0.560s
CRC_RLS	94.1%	0.417s
RSC	97.8%	120s
Proposed	94.4%	0.858s

From Tables 6~8, we can make the following conclusions. First, in all the experiments, RSC always achieves the best recognition rates, while the proposed method always has the second best recognition rates. However, the speed of our method is hundreds of times faster than RSC. Second, CRC_RLS is the fastest algorithm among all the competing methods. It has similar recognition rate to our proposed method for FR without occlusion (refer to Table 8), but has much worse recognition rates than ours for FR with occlusion. Third, SRC implemented by homotopy techniques has similar running time to our method, whereas its recognition rates are lower than our method, especially for FR with occlusion. Finally, SRC implemented by l_1_ls is very slow without improving much the recognition rates compared to SRC implemented by homotopy. In summary, the proposed method achieves a very good balance between robustness and efficiency. In practical FR applications, the database can be of large scale, and our method could lead to desirable recognition accuracy with acceptable time consumption.

6 Conclusion

In this paper, we proposed a simple yet robust and efficient FR scheme by coding the query sample over a dictionary learned from the training samples. To make the FR robust to occlusions and disguise, a coding residual map was first learned from the training samples, and then it is used to detect adaptively the outlier points in the query sample. The detected outliers were then excluded from the coding of the query sample to improve the robustness of FR to occluded samples. Our extensive experimental results on benchmark face databases show that the proposed scheme is very competitive with state-of-the-art methods in terms of accuracy, and it is much faster than the methods such as SRC [7] and RSC [24]. Overall, the proposed method has both robust FR performance and efficiency. It is a very good candidate for real-time face recognition applications.

7 References

1. I. Cox, I, J. Ghosn, and P. Yianil, Feature-base face recognition using mixture distance. In ICCV, 1996.
2. L. Wiskott, J.M. Fellous, N. Kuiger, C. Malsburg. Face recognition by elastic bunch graph matching. IEEE Trans. Pattern Analysis and Machine Intelligence, 19(7): 775-779, 1997.
3. J. H. Wang, J. You, Q. Li, and Y. Xu. Orthogonal discriminant vector for face recognition across pose. Pattern Recognition, 45(12): 4069-4079, 2012.
4. X. He and P. Niyogi, Locality Preserving Projections. In NIPS, 2003.
5. L. S. Qiao, S. C. Chen, and X. Y. Tan. Sparsity preserving projections with applications to face recognition. Pattern Recognition, 42: 331-341, 2010.
6. K. Lee, J. Ho, and D. Kriegman. Acquiring linear subspaces for face recognition under variable lighting. IEEE Trans. Pattern Analysis and Machine Intelligence, 27(5): 684–698, 2005.
7. J. Wright, A. Yang, A. Ganesh, S. Sastry, and Y. Ma. Robust face recognition via sparse representation. IEEE Trans. Pattern Analysis and Machine Intelligence, 31(2): 210–227, 2009.
8. B. Cheng, J. Yang, S. Yan, Y. Fu, and T. Huang. Learning with l_1-graph for image analysis. IEEE Trans. Image Processing, 19(4): 858-866, 2010.
9. S. Gao, I. Tsang, L. Chia, Kernel sparse representation for image classification and face recognition. In ECCV 2010.
10. M. Yang and L. Zhang. Gabor feature based sparse representation for face recognition with Gabor occlusion dictionary. In ECCV2010.
11. J. Wright, Y. Ma, J. Mairal, G. Sapiro, T. Huang, and S. Yan. Sparse representation for computer vision and pattern recognition. Proceedings of IEEE, 98(6): 1031-1044, 2010.
12. L. Zhang, M. Yang, X. Feng. Sparse representation or collaborative representation: Which helps face recognition? In ICCV, 2011.
13. M. Aharon, M. Elad, and A.M. Bruckstein. The K-SVD: an algorithm for designing of overcomplete dictionaries for sparse representation. IEEE Trans. Signal Processing, 54(11): 4311-4322, 2006.
14. A.M. Bruckstein, M. Elad, Dictionaries for Sparse Representation Modeling. Proceedings of the IEEE , 98(6): 1045-1057,2010
15. H. R. Wang, C. F. Yuan, W. M. Hu, and C. Y. Sun. Supervised class-specific dictionary learning for sparse modeling in action recognition. Pattern Recognition, 45(11): 3902-3911, 2012.

16. M. Yang, L. Zhang, X. Feng, D. Zhang, Fisher Discrimination Dictionary Learning for Sparse Representation. In ICCV 2011.

17. A.M. Martinez and R. Benavente. The AR face database. CVC Technical Report No. 24, 1998.

18. R. Gross, I. Matthews. J. Cohn, T. Kanade, and S. Baker. Multi-PIE. Image and Vision Computing, 28:807–813, 2010.

19. T. Cover. Geometrical and statistical properties of systems of linear inequalities with applications in pattern recognition. IEEE Trans. on Electronic Computers, 14(3): 326-334, 1965.

20. R. Hestenes and E. Stiefel. Methods of conjugate gradients for solving linear systems. Journal of Research of the National Bureau of Standards 49(6), 1952

21. S.J. Kim, K. Koh, M. Lustig, S. Boyd, and D. Gorinevsky. A interior-point method for large-scale l_1-regularized least squares. IEEE Journal on Selected Topics in Signal Processing, 1(4): 606–617, 2007.

22. D. Malioutove, M. Cetin, and A. Willsky. Homotopy continuation for sparse signal representation, in ICASS 2005.

23. A. Yang, A. Ganesh, Z. Zhou, S. Sastry, and Y. Ma. Fast l_1-minimization algorithms and application in robust face recognition. UC Berkeley, Technique Report.

24. M. Yang, L. Zhang, J. Yang, D. Zhang. Robust sparse coding for face recognition. In CVPR, 2011.

25. Y. G. Peng, A. Ganesh, J. Wright, W. L. Xu, and Y. Ma. RASL: robust alignment by sparse and low-rank decomposition for linearly correlated images. In CVPR, 2010.

26. W. Zhao, R. Chellappa, P. J. Phillips, and A. Rosenfeld. Face recognition: A literature survey. ACM Computing Survey, 35(4): 399-458, 2003.

27. S. Z. Li and A. K. Jain. Handbook of Face Recognition (Second Edition). Springer, 2011.

28. R. Jafri and H. R. Abrania. A survey of face recognition techniques. Journal of Information Processing Systems, 5(2): 41-68, 2009.

29. M. Turk and A. Pentland. Eigenfaces for recognition. Journal of Cognitive Neuroscience, 3(1): 71-86, 1991.

30. P. Comon. Independent component analysis — a new concept? Signal Process, 36(3): 287-314.

31. J. Yang, D. Zhang, A. Frangi, and J. Yang. Two-dimensional PCA: A new approach to appearance-based face representation and recognition. IEEE Trans. Pattern Analysis and Machine Intelligence, 26(1): 131-137, 2004.

32. M. Elad and M. Aharon. Image denoising via sparse and redundant representations over learned dictionaries. IEEE Trans. Image Processing, 15(12): 3736-3745, 2006.

33. J. Marial, M. Elad, and G. Sapiro. Sparse representation for color image restoration. IEEE Trans. Image Processing, 17(1): 53-69, 2008.
34. J. Mairal, F. Bach, J. Ponce, G. Sapiro, andA. Zisserman, Supervised dictionary learning. In NIPS, 21: 1033–10402009.
35. W. H. Deng, J. N. Hu, J. Guo, W. D. Cai, and D. G. Feng. Robust, accurate and efficient face recognition from a single training image: A uniform pursuit approach. Pattern Recognition, 43(5): 1748-1762, 2010.
36. P. Belhumeur, J. Hespanda, and D. Kriegman, Eigenfaces versus Fisherfaces: Recognition Using Class Specific Linear Projection. IEEE Trans. Pattern Analysis and Machine Intelligence, 19(7): 711-720, 1997.
37. A. Leonardis and H. Bischof, Robust recognition using eigenimages, Computer Vision and Image Understanding, 78(1): 99-118, 2000.
38. S. Chen, T. Shan, and B.C. Lovell, "Robust face recognition in rotated eigenspaces," Proc. Int'l Conf. Image and Vision Computing New Zealand, 2007.
39. A.M. Martinez, Recognizing Imprecisely localized, partially occluded, and expression variant faces from a single sample per class, IEEE Trans. Pattern Analysis and Machine Intelligence, 24(6): 748-763, 2002.
40. T. Takahashi and T. Kurita, A robust classifier combined with an auto-associative network for completing partly occluded images, Neural Networks, 18: 958-966, 2005.
41. K. Hotta, Adaptive weighting of local classifiers by particle filters for robust tracking, Patter Recognition, 42: 619-628, 2009.
42. T. Ahonen, A. Hadid, and M. Pietikainen, Face recognition with local binary pattern: application to face recognition, IEEE Trans. Pattern Analysis and Machine Intelligence, 28(12):2037-2041, 2006.
43. W.P. Chen and Y.S. Gao, Recognizing partially occluded faces from a single sample per class using string-based matching, In ECCV 2010.
44. Y.S. Gao and M.K.H. Leung, Face recognition using line edge map, IEEE Trans. Pattern Analysis and Machine Intelligence, 24(6):764-779, 2002.
45. F. Deboeverie, P. Veelaert, K. Teelen, and W. Philips, Face recognition using parabola edge map, In ACIVS 2008.
46. N.-S. Vu and A. Caplier, Enhanced patterns of oriented edge magnitudes for face recognition and image matching, IEEE Trans. Image Processing, 21(3):1352-1365, 2012.
47. K.N. Le, A mathematical approach to edge detection in hyperbolic-distributed and Gaussian-distributed pixel-intensity images using hyperbolic and Gaussian masks, Digital Signal Processing, 21:162-181, 2011.

48. M. Yang, L. Zhang, Simon C.K. Shiu, and D. Zhang, Gabor Feature based Robust Representation and Classification for Face Recognition with Gabor Occlusion Dictionary, Pattern Recognition, 46(7):1865-1878, 2013.

49. Z.Z. Feng, M. Yang, L. Zhang, Y. Liu, and D. Zhang, Joint Discriminative Dimensionality Reduction and Dictionary Learning for Face Recognition, Pattern Recognition, 46(8):2134-2143, 2013.

50. Z.X. Huang, Y.G. Liu, C.G. Li, M.L. Yang, and L.P. Chen, A robust face and ear based multimodal biometric system using sparse representation, Pattern Recognition, 46(8): 2156-2168, 2013.