

Article

# Learning Kernel-Based Robust Disturbance Dictionary for Face Recognition

Biwei Ding and Hua Ji \*

School of Information Science and Engineering, Shandong Normal University, Jinan 250358, China; dingbiwei@stu.sdn.edu.cn

\* Correspondence: jihua@sdnu.edu.cn; Tel.: +86-133-3515-0180

Received: 24 January 2019; Accepted: 17 March 2019; Published: 20 March 2019



**Abstract:** In this paper, a kernel-based robust disturbance dictionary (KRDD) is proposed for face recognition that solves the problem in modern dictionary learning in which significant components of signal representation cannot be entirely covered. KRDD can effectively extract the principal components of the kernel by dimensionality reduction. KRDD not only performs well with occluded face data, but is also good at suppressing intraclass variation. KRDD learns the robust disturbance dictionaries by extracting and generating the diversity of comprehensive training samples generated by facial changes. In particular, a basic dictionary, a real disturbance dictionary, and a simulated disturbance dictionary are acquired to represent data from distinct subjects to fully represent commonality and disturbance. Two of the disturbance dictionaries are modeled by learning few kernel principal components of the disturbance changes, and then the corresponding dictionaries are obtained by kernel discriminant analysis (KDA) projection modeling. Finally, extended sparse representation classifier (SRC) is used for classification. In the experimental results, KRDD performance displays great advantages in recognition rate and computation time compared with many of the most advanced dictionary learning methods for face recognition.

**Keywords:** face recognition; dictionary learning; kernel discriminant analysis (KDA); sparse representation classifier (SRC)

---

## 1. Introduction

Face recognition is a significant research area in image processing, pattern recognition and computer vision [1] with a very wide scope, especially in the context of information security [2,3]. Face recognition is tremendously challenging due to the influence of various factors such as environment, illumination, expression, and posture when the face image is acquired. In addition, some facial disturbance components, such as wearing hats and glasses, need to be further addressed and improved. Therefore, the accuracy and speed of computer face detection and recognition by the computer are the key elements in current face recognition technology.

In recent years, several major successful face recognition algorithms have been widely applied. Munawar et al. [4] proposed a data-driven unconstrained face recognition method. This method can simultaneously characterize learning and joint registration. Parkhi et al. [5] proposes a method and program for deep face recognition that traverses the complexity of deep networks by assembling large data sets. Munawar et al. [6] proposes a simplified binary single representation method that does not use images. This method proposes a classification strategy for multiple image sets classifying. Xie et al. [7] proposed a multi-column network that uses images as input to obtain feature descriptors. This network implemented a new visual and content-based neural network architecture. Lacey et al. [8] applied a mixed-effects regression model to quantify various aspects of the subject and used longitudinal analysis to study the persistence of automatic face recognition. Duan et al. [9] proposed a

binary unsupervised feature learning method for face recognition using context-aware localization by shifting numbers, and its experimental results are superior to other facial descriptors.

Currently, feature extraction and classifier selection are the two most important components in face recognition research. Feature extraction-based algorithms that have been successfully used for face recognition [10–13]. However, the feature vectors extracted by these methods have a high dimensionality probably leading to causing dimensionality disasters. In order to deal with the problem of high dimensionality, the linear discriminant analysis (LDA) [14] and the principal component analysis (PCA) [15] of the dimensionality reduction method on projecting high-dimensional features onto low-dimensional subspaces are proposed. Nevertheless, due to the face images in the real world are mostly non-linear and the faces have disturbance items, these conventional methods cannot handle the face occlusion problem well. Classifier-based algorithms, such as support vector machine (SVM) [16], AdaBoost [17], nearest subspace (NS) [18] and nearest neighbor (NN) [19], have been successfully adopted to face recognition. However, since the significant differences between testing and training samples, these methods are always subject to universal problems [20].

## 2. Related Works

Wright et al. proposed a classification method based on sparse representation classifier (SRC) [21], which has been experimentally proven to have better performance than classifier-based algorithms for partially occluded samples. Due to their satisfactory performance, various extensions of the SRC have been proposed [22–25]. Deng et al. [26] also put forward the superposition of SRC (SSRC). However, ESRC and SSRC are slow in classification due to the large number of atoms. The work of robust kernel representation with statistical local features (SLF-RKR) [27] and regularized robust coding (RRC) [28] is based on extended SRC, which improves the performance of occluded samples, but they also reduce the performance of non-occluded samples due to overfitting problems. Liu et al. [24] developed a face recognition method that uses a multiscale retinex(MSR) to obtain a similarity matrix and construct a weighted kernel sparse representation. Lin et al. [23] developed a new robust, discriminative and comprehensive dictionary that can learn commonality, class specialty and interference, which effectively strengthens the classification effect of the dictionary. Ou et al. [25] proposed a linear representation classifier based on superposition through the centroid and internal association of the class.

Kernel discriminant analysis (KDA) [29] is a kernel-based face recognition method that projects data into nondiscriminant subspaces. KDA has seen many improvements in recent years. Ren et al. [30] developed a multiple orientation and scale transforms (MOST) algorithm based on Gabor kernel transform. Wu et al. [31] proposed a multicore discriminative dictionary learning method that combines multifeature kernel learning with dictionary learning techniques to improve the recognition rate. Mustafa et al. [32] proposed a face recognition system that combines k-NearestNeighbor (kNN), KDA and SVM. Fan et al. [33] presented a virtual dictionary that can automatically generate new training samples in the training set to construct a model, which solves the problem of small training samples. It is noted that KDA performs better in handling nonoccluded samples, but for an occluded face image, the performance of KDA is worse than SRC.

Huang et al. [34] proposed a kernel extend dictionary(KED) to solve the problem that KDA cannot suppress occlusion variations well. KED has similarities with KRDD, but there are also essential differences. First of all, KED also uses KDA and SRC algorithms, which makes KED and KRDD models have similarities. However, the KRDD model has one more layer structure than KED, so it is more complicated. Secondly, the KED model has only the extended dictionary besides the basic dictionary, and the KRDD model has the real disturbance dictionary and the simulated disturbance dictionary in addition to the basic dictionary. Finally, and the most importantly, KRDD has a series of processed alternative training samples in addition to original training samples. This allows KRDD to process a larger amount of noisy data and make the layering of the data clearer.

Since KDA and SRC have their advantages in different aspects, in this paper, a kernel-based robust disturbance dictionary (KRDD) is proposed for face recognition, which is an effective approach to combine KDA, SRC, and KPCA. It is noteworthy that the KRDD is not a simple addition of these algorithms, as it constructs two species disturbance dictionaries to effectively cope with variations in the occlusion samples, which makes the disturbance dictionary more robust. The image in the gallery sample is not the same as the image in the training sample. KRDD has better adaptability as it can handle various situations according to the actual conditions.

The contributions of this paper are summarized below:

1. A kernel-based robust disturbance dictionary (KRDD) is proposed. The KRDD consists of two types of disturbance dictionaries that can effectively represent more types of occlusion variations.
2. KRDD effectively combines SRC, KDA and KPCA to exploit the respective advantages of these three algorithms. In the case of a small feature dimension, both nonoccluded samples and occluded samples can be processed.
3. A diverse sampling, which includes obtaining real face changes and acquiring virtual samples, is utilized for kernel-based robust disturbance dictionary learning.
4. The dictionary that the KRDD learned includes a basic dictionary, a real disturbance dictionary and a simulated disturbance dictionary to comprehensively represent the data.
5. The experimental results of KRDD are superior to those of other advanced methods.

In the Section 3, SRC, ESRC, and KDA are introduced briefly. In Section 4, the algorithm is stated in detail. In Section 5, several different experiments using 4 face data sets are conducted. KRDD is compared with deep learning on face recognition in Section 6. Finally, the proposed method is summarized in Section 7.

### 3. Fundamental Techniques

Since SRC and ESRC have an efficient processing capability for image recognition, and the KDA algorithm is also a very important factor to the proposed algorithm in this paper, they are briefly introduced here.

#### 3.1. Sparse Representation Classifier

Given a set of gallery samples with  $c$  classes of subjects,  $X = [x_1, x_2, \dots, x_c] \in \mathbb{R}^{d \times n}$ , where  $d$  is the feature dimension,  $n = \sum_{i=1}^c n_i$  represents the number of all training samples,  $x_i$  represents the  $x_i$  class label set to  $c_i$ , the sparse representation coefficient  $\beta \in \mathbb{R}^{n \times 1}$  for a given test sample  $y \in \mathbb{R}^{d \times 1}$  can be computed by calculating the following equation:

$$\min_{\beta} \|y - X\beta\|_2^2 + \lambda \|\beta\|_1. \quad (1)$$

where  $\lambda$  is a scalar constant.

Set  $\delta_c(\beta) \in \mathbb{R}^{n \times 1}$  is a vector of zero except for items related to class  $c$ .  $y$  is assigned to the subject which minimizes the residual by SRC as follows:

$$r_c(y) = \|y - X\delta_c(\beta)\|_2. \quad (2)$$

SRC uses the minimized residual  $C$  related to each subject to perform classification.

$$identity(y) = \arg \min_c r_c(y). \quad (3)$$

$y$  is represented collaboratively by the SRC through all classes of samples and affected by the conditions of the sparse coefficients. If  $y$  comes from a subject, then it is probably that a few samples of this subject can be used to indicate  $y$ . However, in numerous actual face recognition activities, the testing sample  $y$

may be fractionally occluded or disturbed. Under such circumstances, the sparse coefficient can be computed by the following formula:

$$\min_{\beta, \tilde{\beta}} \|y - [X, I] \begin{bmatrix} \beta \\ \tilde{\beta} \end{bmatrix}\|_2^2 + \lambda \left\| \begin{bmatrix} \beta \\ \tilde{\beta} \end{bmatrix} \right\|_1. \tag{4}$$

where the identity matrix  $I \in \mathbb{R}^{d \times d}$  is introduced.

### 3.2. Extended Sparse Representation Classifier (ESRC)

Let  $\tilde{x}_i$  be the  $i$ th sample in the training data used to obtain the intraclass variations,  $\tilde{\mu}_i$  be a natural sample relevant to  $\tilde{x}_i$  or the mean value of the class, then  $\tilde{E}$  can be obtained from the following equation:

$$\tilde{E} = [\tilde{x}_1 - \tilde{\mu}_1, \tilde{x}_2 - \tilde{\mu}_2, \dots, \tilde{x}_n - \tilde{\mu}_n]. \tag{5}$$

In ESRC, the gallery set  $X$  and intra-class variant bases  $\tilde{E}$  can be sparsely coded to obtain the testing sample  $y$ . Consequently, the coding coefficient can be acquired by the following formula:

$$\min_{\beta, \tilde{\beta}} \|y - [X, \tilde{E}] \begin{bmatrix} \beta \\ \tilde{\beta} \end{bmatrix}\|_2^2 + \lambda \left\| \begin{bmatrix} \beta \\ \tilde{\beta} \end{bmatrix} \right\|_1. \tag{6}$$

where  $\beta$  is the coding coefficient of  $y$  over  $X$ , and  $\tilde{\beta}$  is the coding coefficient of  $y$  over  $\tilde{E}$ .

The minimized residual of every subject is indicated as follows:

$$r_c(y) = \|y - [X - \tilde{E}] \begin{bmatrix} \delta_c(\beta) \\ \tilde{\beta} \end{bmatrix}\|_2. \tag{7}$$

### 3.3. Kernel Discriminant Analysis

The following formula of KDA is given according to Baudat et al. [29], where  $z_i$  and  $z_j$  denote the training samples. First, the feature space  $F$  is an issue of nonlinear mapping:  $\phi : \mathbb{R}^d \rightarrow F$ . According to a proper  $\phi$ , an inner product  $\langle \cdot, \cdot \rangle$  can be obtained on  $F$  as follows:

$$\langle \phi(z), \phi(z') \rangle = k(z, z') \tag{8}$$

where  $\tilde{Z} = [\tilde{z}_1, \tilde{z}_2, \dots, \tilde{z}_n] \in \mathbb{R}^{d \times \hat{n}}$  be  $\hat{n}$  samples are used to calculate the KDA projection vector. On the basis of Baudat et al. [29], the projection vector of KDA is learned as follows:  $p^\phi = \sum_{i=1}^{\hat{n}} a_i \phi(\tilde{z}_i)$ . Then, the samples  $z$  were projected to the vector. The formula is as follows:

$$\langle p^\phi, \phi(z) \rangle = \sum_{i=1}^{\hat{n}} a_i \langle \phi(\tilde{z}_i), \phi(z) \rangle = a^T K(:, z) \tag{9}$$

where  $K(:, z) = [k(\tilde{z}_1, z), k(\tilde{z}_2, z), \dots, k(\tilde{z}_{\hat{n}}, z)]$ .

## 4. Model of Robust, Comprehensive Dictionary Learning

KDA is a desirable method to inhibit intraclass compilation and maximize the disparity between different subjects. However, when dealing with facial occlusion problems, SRC performs better than KDA. KPCA can extract several effective kernel principal components of occlusion variables on the basis of dimensionality reduction. To improve the performance of dictionary learning in face recognition, a new kernel-based robust disturbance dictionary (KRDD) model for face recognition is proposed by combining KDA, SRC and KPCA. The KRDD model learns a basic dictionary  $D$ , a simulated disturbance dictionary  $D_s$ , and a real disturbance dictionary  $D_r$ .  $D$  represents the data commonality of all the subjects, while the disturbance dictionaries can represent data from disturbance

components of noise, occlusion, etc. In reality, face images have various kinds of disturbances, such as noise, expression, occlusion and illumination variations. The basic dictionary cannot recover the identity of the face image well, and the relevant representation coefficients are destroyed. They play a vital role in restoring face images.

Figure 1 shows the basic structure of the proposed KRDD. First, KDA is trained from the training samples and get  $D$  from the gallery set of the KDA projection. Secondly, the discrepancy between the disturbance sample and the normal sample are utilized to learn  $r$  principal components and to further obtain the actual perturbation model, and the real disturbance dictionary is obtained from the  $r$  principal components of the KDA projection. Third, a series of processing steps are performed on the alternative samples. KPCA [35] are also used to learn  $r$  principal components to form a simulated disturbance model, and a KDA projection of the model is utilized to obtain a simulated disturbance dictionary. Finally, the testing sample is classified as the subject that minimizes the residual.

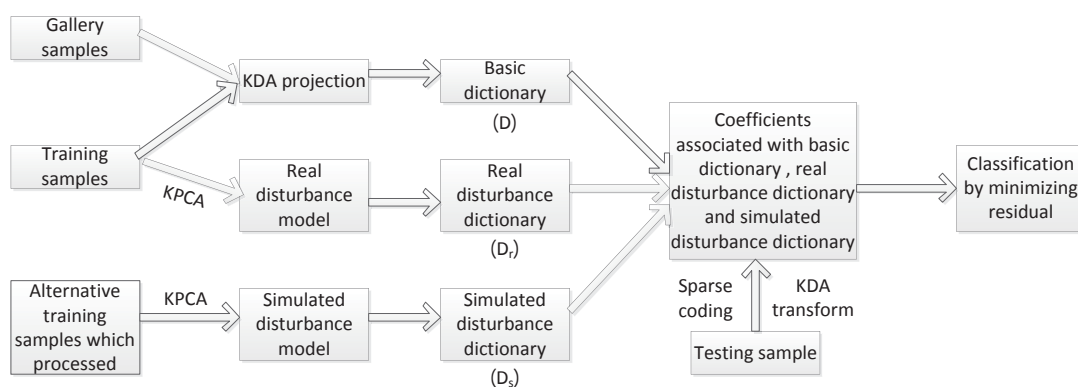


Figure 1. The structure of the proposed kernel-based robust disturbance dictionary (KRDD).

#### 4.1. Learning the Real Disturbance Model in Kernel Space

Although KDA can suppress intraclass variables, it does not suppress occlusion variables well. For the purpose of dealing with this problem, the real disturbance model is obtained using KPCA. The principle of KPCA is to effectively represent possible changes by a few principal components, which are mainly composed of several principal components. In this way, the first  $n$  samples are utilized to obtain the real disturbance model. The covariance matrix of disturbance changes is acquired from the following equation:

$$\tilde{S}^\phi = \sum_{i=1}^m (\phi(X_i) - \phi(\tilde{\gamma}_i))(\phi(X_i) - \phi(\tilde{\gamma}_i))^T = \sum_{i=1}^m \Phi(i)\Phi^T(i) \quad (10)$$

where  $\Phi(i) = \phi(X_i) - \phi(\tilde{\gamma}_i)$ ,  $X_i$  is the disturbance sample, and  $\tilde{\gamma}_i$  is the corresponding natural sample or class mean.

Then the optimal projection  $\tilde{q}^\phi$  that maximizes the covariance after projection could be calculated from the following equation:

$$\max_{q^\phi} (q^\phi)^T \tilde{S}^\phi q^\phi \quad s.t. \langle q^\phi, q^\phi \rangle = 1 \quad (11)$$

where the principal component of  $\tilde{S}^\phi$  is  $q^\phi$ , and it can be seen from the following equation:

$$\tilde{S}^\phi q^\phi = \lambda q^\phi \quad (12)$$

According to Baudat et al. [29], the solution is  $q^\phi = \sum_{i=1}^m \tilde{a}_i \Phi(i)$ , where  $\tilde{a} = [\tilde{a}_1, \dots, \tilde{a}_m]^T$  satisfies the following formula:

$$\begin{aligned} \tilde{K}_{i,j} &= \langle \Phi(i), \Phi(j) \rangle \\ &= \langle \phi(X_i) - \phi(\tilde{\gamma}_i), \phi(X_j) - \phi(\tilde{\gamma}_j) \rangle \\ &= \langle \phi(X_i), \phi(X_j) \rangle - \langle \phi(X_i), \phi(\tilde{\gamma}_j) \rangle \\ &\quad - \langle \phi(\tilde{\gamma}_i), \phi(X_j) \rangle + \langle \phi(\tilde{\gamma}_i), \phi(\tilde{\gamma}_j) \rangle \end{aligned} \tag{13}$$

Then, the top  $r$  kernel principal components  $\tilde{q}^\phi = [\tilde{q}_1^\phi, \tilde{q}_2^\phi, \dots, \tilde{q}_r^\phi]$  constitute the real disturbance model.

#### 4.2. Learning the Simulated Disturbance Model in Kernel Space

Huang et al. [34] proposed an occlusion model to solve the problem that KDA cannot suppress occlusion variations well. However, such an occlusion model uses only real disturbance and does not take more simulated disturbance conditions into consideration. In this regard, the simulated disturbance dictionary in the new kernel-based robust disturbance model is generated. To make the learning dictionary comprehensive and robust to lighting conditions, disguises, facial gestures and facial expressions, a series of processing steps are performed on  $Z$  to obtain alternative training samples  $Z_i, i = 1, 2, 3$ .

##### 4.2.1. Speckle Noise Addition

$Z_1$  can be obtained by adding noise items to the original face image using speckle noise. The specific implementation method is the MATLAB function “imnoise”. Figure 2 shows the comparison of two samples.

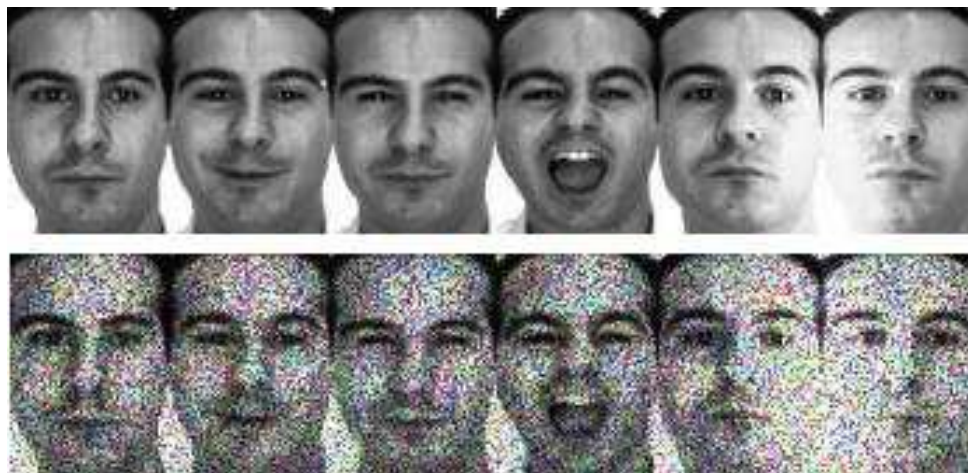


Figure 2. Original Training Sample (Upper) & Samples after Speckle Noise Processing (Lower).

##### 4.2.2. Mirror Conversion

The mirror conversion of the original training sample is used as  $Z_2$ . The mirror image  $t^*$  of the original training image  $t$  can be obtained by the following formula:

$$t^*(k,l) = t(k,h-l+1), (k = 1, \dots, p; l = 1, \dots, h) \tag{14}$$

where  $p$  and  $h$  denote the rows and columns of the image matrix respectively,  $t(k,l)$  represent the pixels located in the  $k$ -th row of  $t$ , and  $t^*(k,l)$  is the pixels located in the  $l$ -th column of  $t^*$ . Figure 3 shows the comparison of two samples.



**Figure 3.** Original Training Sample (left) & Samples after Mirror Processing (right).

#### 4.2.3. Random Square Block Processing

Random square block processing is performed on the original training sample as  $Z_3$ . The position of the square is not subject to human control and it is completely random. Figure 4 shows the comparison of two samples.  $Z_1, Z_2, Z_3$  are then reorganized to obtain the alternative training samples from the above methods. The specific method is to divide each training sample into three groups randomly, from each of the three alternative training samples, to extract a group and reconstitute the new  $\tilde{Z}$ .



**Figure 4.** Original Training Sample (Upper) & Samples after Random Square Block Processing (Lower).

#### 4.3. Concise Steps of KRDD

The steps of KRDD are concisely generalized below:

1. KDA projection was performed on the training sample  $X$  to get  $p^\phi = [p_1^\phi, p_2^\phi, \dots, p_{C-1}^\phi]$ , where  $C$  is the quantity of subjects.
2. The gallery samples can be outside the training samples. The gallery samples  $\tilde{X} = [\tilde{x}_1, \tilde{x}_2, \dots, \tilde{x}_n]$  were projected to obtain the basic dictionary  $D$  by KDA through Formula (9):

$$D = (p^\phi)^T \phi(\tilde{X}) \tag{15}$$

3. KPCA projection was performed on the training sample  $X$  to obtain  $q^\phi = [q_1^\phi, q_2^\phi, \dots, q_r^\phi]$ , then the  $r$  principal component of disturbance variations in the kernel space was used as the kernel real disturbance model.
4. KDA was used to project the kernel principal component  $q^\phi = \sum_{i=1}^m \tilde{a}_i \Phi(i)$  with the following equation without having to explicitly know the nonlinear mapping by (9).

$$\begin{aligned}
 (p^\phi)^T q^\phi &= \langle p^\phi, q^\phi \rangle = \left\langle \sum_{i=1}^m a_i \phi(X_i), \sum_{j=1}^m \tilde{a}_j \Phi(j) \right\rangle \\
 &= \left\langle \sum_{i=1}^m a_i \phi(X_i), \sum_{j=1}^m \tilde{a}_j (\phi(X_j) - \phi(\tilde{\gamma}_j)) \right\rangle \\
 &= a^T K \tilde{a} - a^T K^\gamma \tilde{a}
 \end{aligned}
 \tag{16}$$

where  $a$  and  $\tilde{a}$  are the coefficient vectors of KDA and the real disturbance model, respectively.  $K_{i,j} = \langle \phi(X_i), \phi(X_j) \rangle$ ,  $K_{i,j}^\gamma = \langle \phi(\tilde{z}_i), \phi(\tilde{\gamma}_j) \rangle$ .  $p^\phi$  was projected by KDA to obtain the kernel real disturbance dictionary:

$$D_r = (p^\phi)^T q^\phi \tag{17}$$

5. Three special schemes were used on  $Z$  to obtain alternative training samples  $\tilde{Z}$ , the sample in  $Z$  can be different from the sample in  $X$ .
6. The  $r$  principal components of disturbance changes were trained in kernel space as the simulated disturbance model, i.e.,  $l^\phi = [l_1^\phi, l_2^\phi, \dots, l_r^\phi]$ , from the set of  $\tilde{Z}$ , as the kernel simulated disturbance model.
7.  $l^\phi$  was projected by KDA to obtain the kernel simulated disturbance dictionary via (17) as follows:

$$D_s = (p^\phi)^T l^\phi \tag{18}$$

8. A testing sample  $y$  was projected by KDA:

$$y_{KDA} = (p^\phi)^T \phi(y) \tag{19}$$

9. The sparse representation of  $y_{KDA}$  was found according to the basic dictionary  $D$ , the real disturbance dictionary  $D_r$  and the simulated disturbance dictionary  $D_s$ :

$$\min_{\beta, \beta_r, \beta_s} \|y_{KDA} - [D, D_r, D_s] \begin{bmatrix} \beta \\ \beta_r \\ \beta_s \end{bmatrix}\|_2^2 + \lambda \left\| \begin{bmatrix} \beta \\ \beta_r \\ \beta_s \end{bmatrix} \right\|_1.
 \tag{20}$$

10.  $y$  was classified to the subject that minimizes the residual as follows:

$$r_c(y_{KDA}) = \|y_{KDA} - [D, D_r, D_s] \begin{bmatrix} \delta_c(\beta) \\ \beta_r \\ \beta_s \end{bmatrix}\|_2.
 \tag{21}$$

### 5. Experimental Results

In the experiment, several large data sets are applied to verify the performance of KRDD algorithm. The programming environment is MATLAB R2012b, 2.50 GHz CPU and 8G RAM. In terms of data sets, the training samples and testing samples are augmented with alternative training samples and gallery samples to evaluate performance. The alternative training samples are processed sample. As the face recognition system in reality usually obtains a single sample of each subject, Deng et al. [36,37] designed an equidistant circular embedding method to solve the single-sample face recognition for



each subject. The training set is applied to establish a model first, then the gallery samples and the testing samples are matched to each other through this recognition model. All methods were tested using LBP features.

### 5.1. Experiment on AR Dataset

The AR dataset contains over 4000 pictures of 126 objects. Figure 5 shows a diagram of some examples from the AR face database. The dimension of each face image is  $165 \times 120$ . Each person had 26 photographs taken in two separate sessions, except 17 people had fewer than 26 images. Here, the nonoccluded image training set from session 1 and all images of the 17 people are selected for training. There are two cases in the gallery set. One gallery sample is equal to the training samples, which means each theme has multiple samples, and the other gallery set has only one sample per subject, which is a collection of neutral samples for each subject in session 1. The alternative training samples are generated by mirroring, with added noise and random square block occlusion of the original training samples. Finally, all images in session 2 are chosen as the testing samples, the test samples are divided into three subsets: expression, lighting, occlusion. The proposed KRDD method learned from the real occlusion variations in the 17 subjects and the stimulated disturbance from the alternative training samples.



**Figure 5.** Example Images of one Subject on the AR Dataset.

Table 1 shows the recognition results and classification time of the KRDD algorithm and the six comparison methods using the AR dataset. Each person in the gallery set used in this experiment has multiple gallery samples. From Table 1, it can be noted that KRDD achieved the highest recognition rate. The KED is the second-best algorithm. In addition, the KRDD recognition rate is 1.3% higher than that of the KED in the occlusion subset. All these results show that the KRDD performs better than KED in both disturbance and nondisturbance samples. The fastest classification method is MOST because it is not based on the sparse method. KRDD is much faster than the other methods. Table 2 differs from Table 1 because each person in the gallery set uses a single gallery sample. Based on the experimental results, it is obvious that the performance of KRDD is virtually unchanged, and the classification speed is faster, which proves the robustness of KRDD. From Tables 1 and 2 we can conclude that the recognition rate of KRDD in the three subsets is higher than those of KED. Especially in the occlusion subset, KRDD has more obvious improvement. This is because the KRDD model has two different types of disturbance dictionary, which have one more disturbance dictionary than the KED model has, to cope with more noisy data.

**Table 1.** Recognition accuracy and classification time for advanced methods on AR Database with multiple gallery samples per subject.

Method	Expression	Lighting	Occlusion	Classification Time (s)
SRC [21]	86.4	90.3	45.4	144
ESRC [22]	90.8	92.4	58.1	392
RRC [28]	92.2	94.6	60.7	954
SLF-RKR [27]	96.8	97.7	85.5	7427
MOST [30]	97.1	96.6	81.9	<b>64</b>
KED [34]	98.4	98.2	94.2	81
KRDD	<b>99.2</b>	<b>99.0</b>	<b>95.5</b>	78

**Table 2.** Recognition Accuracy and Classification Time for Advanced Methods on AR Database with A Single Gallery Sample Per Subject.

Method	Expression	Lighting	Occlusion	Classification Time (s)
SRC [21]	62.4	91.4	20.8	73
ESRC [22]	87.5	91.6	68.1	171
RRC [28]	72.6	66.7	42.5	295
SLF-RKR [27]	33.8	26.9	21.4	1682
MOST [30]	96.8	96.3	87.9	<b>44</b>
KED [34]	98.6	98.2	94.0	52
KRDD	<b>99.3</b>	<b>99.0</b>	<b>95.2</b>	49

### 5.2. Experiment on LFWA Dataset

The second personal face database chosen is the largest database—the aligned labeled face in the wild (LFWA), which has unconstrained changes in posture, lighting, expression, misalignment, occlusion, etc. The LFWA face data set has a total of 13,233 images of 5749 people. Figure 6 shows a face image of a subject in the LFWA database. A subset of 5500 images is selected for this experiment. The training sample consists of 9 samples per person (or more than 9 samples per person) and the remainder are used as a testing sample. Similarly, the alternative training samples are generated using the same methods used for the AR dataset. PCA is utilized to reduce the feature dimensions to 300, 400, and 500 to compare these advanced methods.

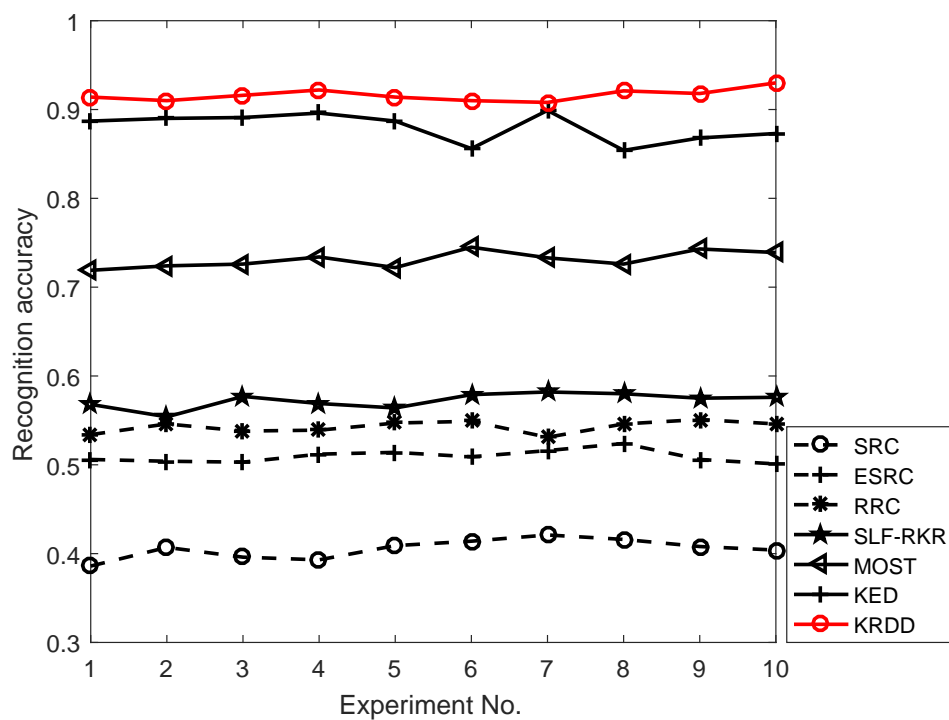
**Figure 6.** Example Images of one Subject on the LFWA Dataset.

From Table 3 it is noted that KRDD achieves the highest recognition rate, and the classification time is almost the same as MOST. The recognition rate of KED is the second-best after KRDD. When the feature dimension is at 500, KRDD is 1.7% higher than KED in the recognition rate. MOST is also a KDA-based approach, so it handles suppressing intraclass variations better than other methods. To observe the performance of KRDD more intuitively, the experiment is repeated 10 times with a feature dimension of 600. As shown in Figure 7 among KDA-based methods, KRDD has the best

performance. Compared with SRC, ESRC, RRC, SLF-RKR, MOST and KED, KRDD was improved by an average of 50.5%, 40%, 36.7%, 35.0%, 19.2%, and 2.7%, respectively.

**Table 3.** The Recognition Rates and Computing Time for Training Dictionaries and Classifying A Testing Sample on the LFWA Database.

Method	300	400	500	Classification Time (s)
SRC [21]	38.7	39.6	40.3	152
ESRC [22]	49.6	51.3	50.8	376
RRC [28]	50.1	52.7	53.4	829
SLF-RKR [27]	53.6	54.7	56.6	6844
MOST [30]	70.4	71.3	71.8	62
KED [34]	88.5	89.2	89.8	74
KRDD	90.1	90.8	91.5	70



**Figure 7.** Face Recognition Results for Advanced Methods and the Proposed Methods on the LFWA Dataset.

We also compare the proposed method with deep learning-based algorithms that require a large amount of additional training data. Such as, ConvNet-RBM [38], GaussianFace [39], DeepID3 [40]. There are also methods that do not require additional training data, such as VMRS [41], Sub-SML [42], and high-dim LBP [43].

Table 4 lists the average recognition rates of advanced algorithms in the LFWA data set. The proposed method achieves a recognition rate of 92.46%, which is better than some advanced methods. Although the proposed method has a lower recognition rate than the deep learning based method, KRDD does not need to collect a large amount of additional training data.

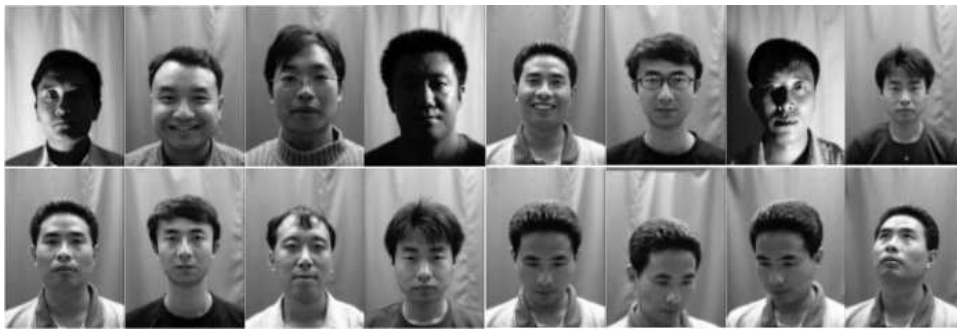
**Table 4.** Mean recognition accuracy for Advanced Methods on LFWA Dataset.

Method	Mean Accuracy
VMRS [41]	92.05
Sub-SML [42]	90.75
high-dim LBP [43]	93.18
KRDD	92.46
ConvNet-RBM [38]	91.75
GaussianFace [39]	98.52
DeepID3 [40]	<b>99.53</b>

### 5.3. Experiment on CAS-PEAL Dataset

#### 5.3.1. Experimental Results without Subset

The CAS-PEAL dataset is a wide-ranging Chinese face database. The progress of face recognition is facilitated by providing face images from different sources with differences in gestures, expressions, lights, and accessories. The CAS-PEAL dataset includes 99,594 faces images of a total of 1040 people, 595 men and 445 women. In this experiment, 10,311 face images from 1040 people are selected. Figure 8 shows a facial image of a subject in the CAS-PEAL database.

**Figure 8.** Example Images of one Subject on the CAS-PEAL Database.

Several classes are randomly selected from the lighting, expression, and subset to constitute a training set, and the alternative training samples are obtained in the same way. One image of per subject is used to compile a gallery set. The test set comprises the remaining face images. For the sake of fairness, all methods use the same LBP features.

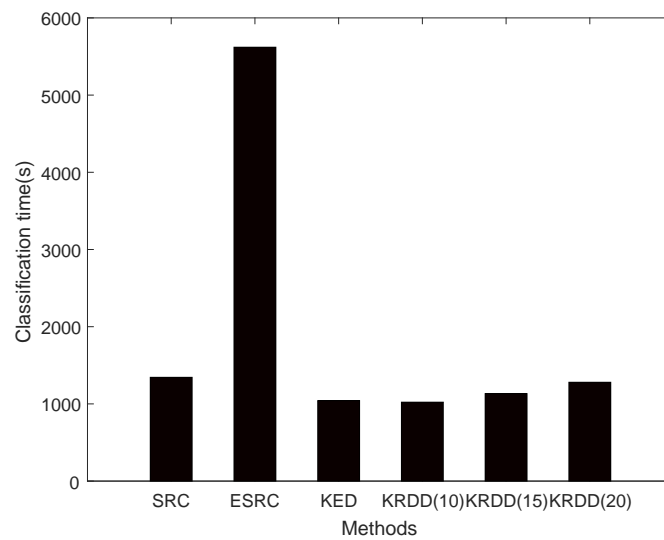
Since KRDD combines three algorithms: SRC, KDA, and PCA, it is not a simple stack of the three algorithms. To prove that KRDD is fundamentally different from SRC and KDA, KRDD is compared with PCA, SRC, ESRC, KDA, KDA + SRC, KDA + ESRC and KED. All methods were tested using LBP features. Again, PCA is applied to reduce the feature dimensions to 600, 700, and 800 to compare these advanced methods. In this experiment, KRDD learned from the first 10, 15 and 20 components of the real disturbance variable and the simulated disturbance variable in kernel space.

It can be seen in Table 5 that the recognition rate of the simple association of KDA and SRC is indeed nearly 20% higher than the rate of KDA and SRC alone. However, KRDD is not a simple addition of KDA and SRC. The recognition rate of KRDD is approximately 5.8% higher than that of KDA plus SRC and KDA plus ESRC. KRDD also achieves higher recognition accuracy than that of KED. Among all the methods, KRDD achieved the best results. Figure 9 shows the classification time for several algorithms. Although ESRC has a higher recognition accuracy than SRC, it takes a much longer time because it expands the quantity of atoms in the dictionary.  $D$  in KRDD is based on low dimensions and only adds a small number of atoms, which is faster than SRC. It is remarkable that KRDD has a higher recognition rate and takes less time than KED for the same number of selected occlusion components. Additionally, a positive correlation could be observed between selected

occlusion components and the recognition rate of KRDD. Nevertheless, selecting more occlusion components results in longer time computation times.

**Table 5.** Comparison of Related Methods on the CAS-PEAL Database.

Method	600	700	800
PCA [15]	70.6	71.4	71.7
SRC [21]	71.5	72.6	73.5
ESRC [22]	86.3	86.8	87.8
KDA [29]	75.1	75.7	76.4
KDA + SRC	89.7	89.9	90.1
KDA + ESRC	89.9	90.2	91.0
KED	95.0	95.2	96.0
KRDD(10)	95.5	95.6	96.3
KRDD(15)	95.6	95.9	96.7
KRDD(20)	95.8	96.3	97.1



**Figure 9.** The Classification Time for Different Algorithms on the CAS-PEAL Database.

### 5.3.2. Experimental Results with Subset

For the purpose of verifying the degree of interference of each change on the recognition rate, the CAS-PEAL dataset in the subset created by Deng et al. [34] is applied in this experiment. This dataset is the same as the one used in the previous section, and the training samples and the gallery samples are created in a similar way. The difference is that the data set of Deng et al. [34] contains 9 testing sets and 9 corresponding subsets: accessory, lighting, expression, time, background, distance, glasses, sunglasses and hat. Six subsets are selected for experimental comparison. To ensure the recognition rate of the results are consistent, the experiment is repeated 10 times for every algorithm.

Figures 10 and 11 show the recognition rates of KRDD and the other advanced algorithms on different subsets. By observing the changes in recognition rates over six subsets, illumination variations are the most challenging for these algorithms. Among all the algorithms, it is significant that ESRC, MOST, KED and KRDD are more robust than the other algorithms. ESRC is relatively more stable than SRC because ESRC benefits from an intraclass variant dictionary. Although the recognition rates of both MOST and KED are satisfying, KRDD still achieves the highest recognition rate in each subset. This proves that KRDD is not only effective in dealing with nonoccluded samples, but also it will not be over fit when processing occluded samples since it is sparse.

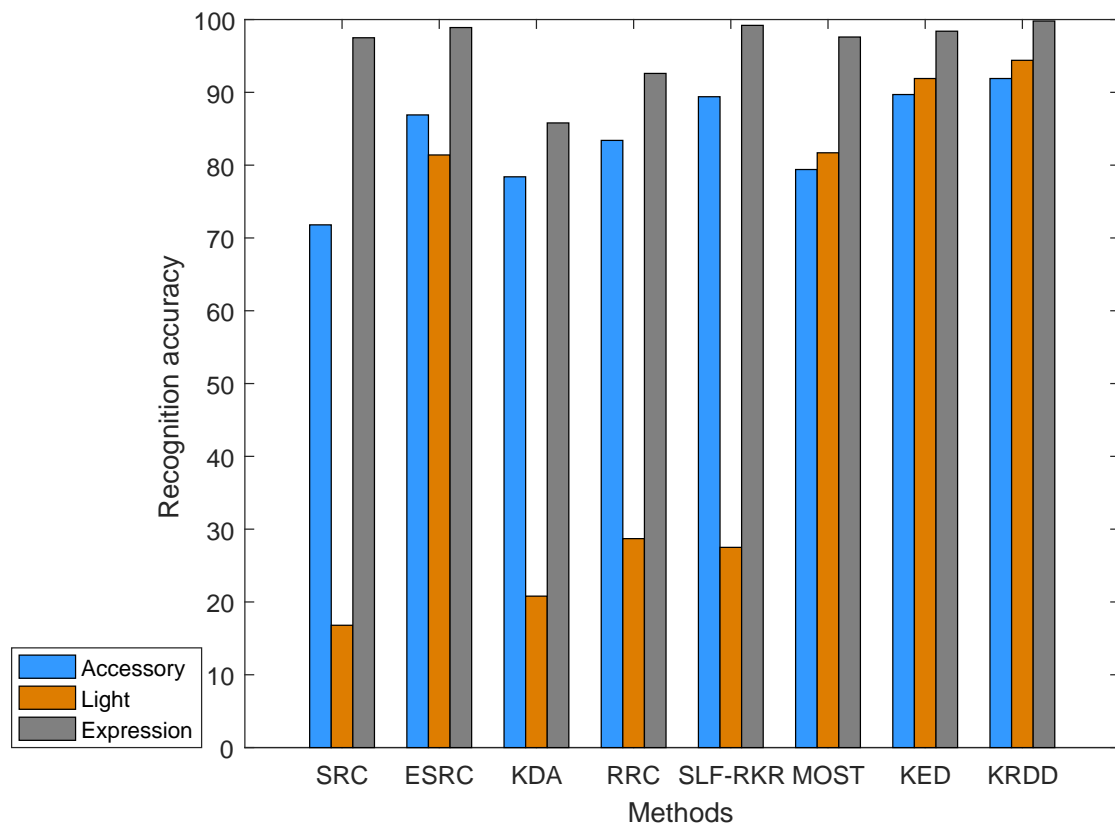


Figure 10. The Recognition Accuracy for Different Algorithms on the CAS-PEAL Database.

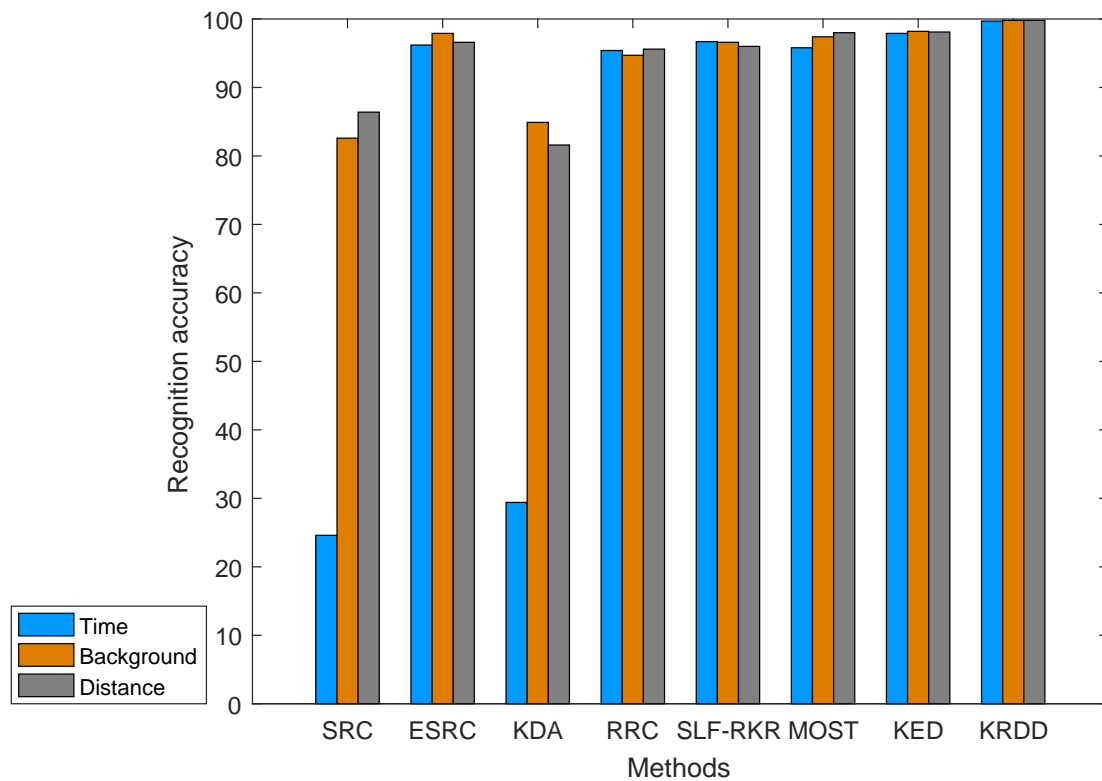


Figure 11. The Recognition Accuracy for Different Algorithms on the CAS-PEAL Database.

### 5.4. Experiment on FERET Dataset

The last experiment was executed on the largest face database FERET. The image is taken in a semicontrolled condition. There is a large difference among the images of the same subjects in the FERET dataset in terms of expressions, lighting, gesture, and time variation. In this experiment, an image of each class is selected in subset A of the dataset as the gallery set. The test set consists of four subsets of expression, lighting, and time variation. The remaining images constitute a training set and an alternative training set. Figure 12 shows some example images from the FERET data set.



Figure 12. Example Images of one Subject on the FERET Database.

Figure 13 displays the recognition accuracy of several advanced algorithms with the FERET database. It can be observed that two subsets separating by 2 present a huge challenge for these algorithms, because people’s facial features change slightly over time. In summary, KRDD achieved the highest recognition accuracy in all subsets. For the four subsets, KRDD displayed an increased average recognition rate by 33.5% compared with SRC, 24.6% compared with ESRC, 18.3% compared with RRC, 13.5% compared with SLF-RKR, 13.4% compared with MOST.

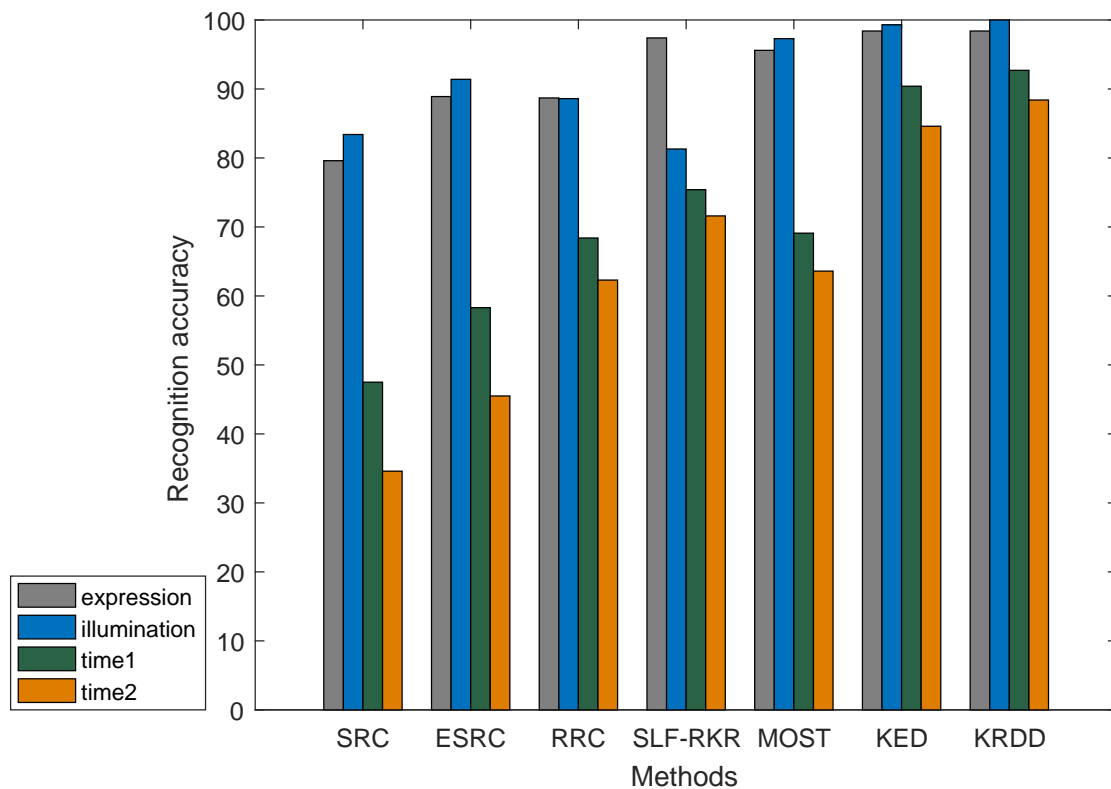


Figure 13. The Recognition Accuracy for Different Algorithms on the FERET Database.

### 6. Comparison between KRDD and Deep Learning

In recent years, deep learning has been widely used in pattern recognition to extract deep features of face images. Sun et al. [40] proposed two deep neural network architectures built by stacked convolution and pilot layers for face recognition (DeepID3). Ding et al. [44] proposed a new

end-to-end CNN model and a new loss function, which can share the convolutional layer to extract features and improve the robustness and discriminability of the model. Masi et al. [45] proposed an attitude-aware unconstrained model using several deep convolutional neural networks, which can solve the problem of posture changes.

Undoubtedly, the KRDD proposed from the original training sample and the alternative training sample using the convolutional neural network may achieve higher recognition accuracy. The model of KRDD with deep learning will be developed in the future. However, face recognition algorithms involving deep features require very complex models and a large number of parameters to adjust, especially the higher the accuracy of the model is, the worse the general robustness will be. They also require a large amount of outside training data that needs to be collected in the same circumstances as the data set used. In a fast-developing society, recognition accuracy is not the only important factor to be considered, since computing time is a factor that cannot be ignored. Therefore, the proposed KRDD model is a robust classifier based on hand-craft features. It has fewer parameters and is easier to be adjusted. Its computation time is much shorter than that of deep learning algorithms. KRDD will be further improved in the future.

## 7. Conclusions

In this paper, a novel kernel-based robust disturbance dictionary (KRDD) learning model is proposed. Three types of samples are extracted to generate a robust and comprehensive dictionary for KRDD, including training samples extracted from real disturbance variations, alternative training samples generated by several different schemes, and gallery samples. KRDD has learned to include the basic dictionary, the real disturbance dictionary and the simulated disturbance dictionary, which fully considers the integrity of the data. KRDD combines KDA, SRC and KPCA in an efficient way, which significantly enhances the recognition performance of occluded samples. Experiments on the AR, FLWA, CAS-PEAL and FERET face databases show that KRDD has a higher recognition rate than some of the most advanced methods, and it can also cope with more realistic situations.

**Author Contributions:** Data curation, B.D.; Formal analysis, B.D.; Funding acquisition, H.J.; Methodology, B.D.; Supervision, H.J.; Writing—original draft, B.D.; Writing—review & editing, H.J.

**Funding:** This paper was supported by the National Natural Science Foundation of China under Grant Nos. 61772322, 61572298, 61702310, 61873151.

**Conflicts of Interest:** The authors declared that they have no conflicts of interest to this work.

## References

1. Lei, Z.; Pietikäinen, M.; Li, S.Z. Learning discriminant face descriptor. *IEEE Trans. Pattern Anal. Mach. Intell.* **2014**, *36*, 289–302. [[CrossRef](#)] [[PubMed](#)]
2. Chai, Z.; Sun, Z.; Mendez-Vazquez, H.; He, R.; Tan, T. Gabor ordinal measures for face recognition. *IEEE Trans. Inf. Forensics Secur.* **2014**, *9*, 14–26. [[CrossRef](#)]
3. Best-Rowden, L.; Han, H.; Otto, C.; Klare, B.F.; Jain, A.K. Unconstrained face recognition: Identifying a person of interest from a media collection. *IEEE Trans. Inf. Forensics Secur.* **2014**, *9*, 2144–2157. [[CrossRef](#)]
4. Hayat, M.; Khan, S.H.; Werghe, N. Joint Registration and Representation Learning for Unconstrained Face Identification. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; pp. 1551–1560. [[CrossRef](#)]
5. Parkhi, O.M.; Vedaldi, A.; Zisserman, A. Deep face recognition. *BMVC* **2015**, *1*, 6.
6. Hayat, M.; Khan, S.H.; Bennamoun, M. Empowering simple binary classifiers for image set based face recognition. *Int. J. Comput. Vis.* **2017**, *123*, 479–4986. [[CrossRef](#)]
7. Xie, W.D.; Zisserman, A. Multicolumn Networks for Face Recognition. *Comput. Vis. Pattern Recognit.* **2018**, arXiv:1807.09192.
8. Best-Rowden, L.; Jain, A.K. Longitudinal Study of Automatic Face Recognition. *IEEE Trans. Pattern Anal. Mach. Intell.* **2018**, *40*, 148–162. [[CrossRef](#)]



9. Duan, Y.; Lu, J.; Feng, J.; Zhou, J. Context-Aware Local Binary Feature Learning for Face Recognition. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, *40*, 1139–1153. [[CrossRef](#)]
10. Ahonen, T.; Hadid, A.; Pietikainen, M. Face description with local binary patterns: Application to face recognition. *IEEE Trans. Pattern Anal. Mach. Intell.* **2012**, *28*, 2037–2041. [[CrossRef](#)] [[PubMed](#)]
11. Dalal, N.; Triggs, B. Histograms of oriented gradients for human detection. In Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR), San Diego, CA, USA, 20–25 June 2005; pp. 886–893. [[CrossRef](#)]
12. Liu, C.; Wechsler, H. Gabor feature based classification using the enhanced Fisher linear discriminant model for face recognition. *IEEE Trans. Image Process.* **2002**, *11*, 467–476. [[CrossRef](#)]
13. Lowe, D. Distinctive image features from scale-invariant keypoints. *Int. J. Comput. Vis.* **2004**, *60*, 91–110. [[CrossRef](#)]
14. Belhumeur, P.N.; Hespanha, J.P.; Kriegman, D.J. Eigenfaces vs. fisherfaces: Recognition using class specific linear projection. *IEEE Trans. Pattern Anal. Mach. Intell.* **1997**, *19*, 711–720. [[CrossRef](#)]
15. Turk, M.A.; Pentland, A.P. Face recognition using eigenfaces. In Proceedings of the International Conference on Computer Vision and Pattern Recognition (CVPR), Maui, HI, USA, 3–6 June 1991; Volume 84, pp. 586–591. [[CrossRef](#)]
16. Vapnik, V.N. Statistical Learning Theory. *Ann. Inst. Stat. Math.* **2003**, *55*, 371–389. [[CrossRef](#)]
17. Yang, P.; Shan, S.; Gao, W.; Li, S.; Zhang, D. Face recognition using Ada-Boosted Gabor features. In Proceedings of the IEEE International Conference on Automatic Face and Gesture Recognition, Seoul, Korea, 19 May 2004; pp. 356–361. [[CrossRef](#)]
18. Ho, J.; Yang, M.; Lim, J.; Lee, K.; Kriegman, D. Clustering appearances of objects under varying illumination conditions. In Proceedings of the International Conference on Computer Vision and Pattern Recognition (CVPR), Madison, WI, USA, 18–20 June 2003; pp. 11–18. [[CrossRef](#)]
19. Cover, T.M.; Hart, P.E. Nearest neighbor pattern classification. *IEEE Trans. Inf. Theory* **1967**, *13*, 21–27. [[CrossRef](#)]
20. Zhang, W.; Shan, S.; Gao, W.; Chen, X. Local gabor binary pattern histogram sequence (LGBPHS): A novel non-statistical model for face representation and recognition. In Proceedings of the Tenth IEEE International Conference on Computer Vision (ICCV), Beijing, China, 17–21 October 2005; pp. 786–791. [[CrossRef](#)]
21. Wright, J.; Yang, A.Y.; Ganesh, A.; Sastry, S.S.; Ma, Y. Robust face recognition via sparse representation. *IEEE Trans. Pattern Anal. Mach. Intell.* **2009**, *31*, 210–227. [[CrossRef](#)]
22. Deng, W.; Hu, J.; Guo, J. Extended src: Undersampled face recognition via intraclass variant dictionary. *IEEE Trans. Pattern Anal. Mach. Intell.* **2012**, *34*, 1864–1870. [[CrossRef](#)]
23. Lin, G.J.; Yang, M.; Yang, J.; Shen, L.L.; Xie, W.C. Robust, Discriminative and Comprehensive Dictionary Learning for Face Recognition. *Pattern Recognit.* **2018**, *81*, 341–356. [[CrossRef](#)]
24. Liu, X.; Lu, L.Y.; Shen, Z.X.; Lu, K.X. A novel face recognition algorithm via weighted kernel sparse representation. *Future Gener. Comput. Syst.* **2016**, *80*, 653–663. [[CrossRef](#)]
25. Ou, W.; Luan, X.; Gou, J.; Zhou, Q. Face Recognition via Collaborative Representation: Its Discriminant Nature and Superposed Representation. *IEEE Trans. Pattern Anal. Mach. Intell.* **2018**, *40*, 2513–2521.
26. Deng, W.; Hu, J.; Guo, J. In defense of sparsity based face recognition. In Proceedings of the 2013 IEEE Conference on Computer Vision and Pattern Recognition, Portland, OR, USA, 23–28 June 2013; pp. 399–406. [[CrossRef](#)]
27. Yang, M.; Zhang, L.; Shiu, S.C.; Zhang, D. Robust kernel representation with statistical local features for face recognition. *IEEE Trans. Neural Netw. Learn. Syst.* **2013**, *24*, 900–912. [[CrossRef](#)]
28. Yang, M.; Zhang, L.; Yang, J.; Zhang, D. Regularized robust coding for face recognition. *IEEE Trans. Image Process.* **2013**, *22*, 1753–1766. [[CrossRef](#)]
29. Baudat, G.; Anouar, F. Generalized discriminant analysis using a kernel approach. *Neural Comput.* **2000**, *12*, 2385–2404. [[CrossRef](#)]
30. Ren, C.X.; Dai, D.Q.; Li, X.X.; Lai, Z.R. Band-reweighted Gabor kernel embedding for face image representation and recognition. *IEEE Trans. Image Process.* **2014**, *23*, 725–740. [[CrossRef](#)]
31. Wua, X.; Lia, Q.; Xua, L.L.; Chen, K.W.; Yao, L. Multi-feature kernel discriminant dictionary learning for face recognition. *Pattern Recognit.* **2017**, *66*, 404–411. [[CrossRef](#)]
32. Al-Dabagh, M.Z.N.; Alhabib, M.H.M. Face Recognition System Based on Kernel Discriminant Analysis, K-Nearest Neighbor and Support Vector Machine. *Int. J. Res. Eng.* **2018**, *5*, 335–338. [[CrossRef](#)]

33. Fan, Z.Z.; Zhang, D.; Wang, X.; Zhu, Q.; Wang, Y.F. Virtual dictionary based kernel sparse representation for face recognition. *Pattern Recognit.* **2018**, *76*, 1–13. [[CrossRef](#)]
34. Huang, K.K.; Dai, D.Q.; Ren, C.X.; Lai, Z.R. Learning Kernel Extended Dictionary for Face Recognition. *IEEE Trans. Neural Netw. Learn. Syst.* **2016**, *28*, 1082–1094. [[CrossRef](#)]
35. Smola, A. Nonlinear component analysis as a kernel eigenvalue problem. *Neural Comput.* **1998**, *10*, 1299–1319. [[CrossRef](#)]
36. Deng, W.; Hu, J.; Guo, J.; Cai, W.; Feng, D. Robust, accurate and efficient face recognition from a single training image: A uniform pursuit approach. *Pattern Recognit.* **2010**, *43*, 1748–1762. [[CrossRef](#)]
37. Deng, W.; Hu, J.; Zhou, X.; Guo, J. Equidistant prototypes embedding for single sample based face recognition with generic learning and incremental learning. *Pattern Recognit.* **2014**, *47*, 3738–3749. [[CrossRef](#)]
38. Sun, Y.; Wang, X.; Tang, X. Hybrid Deep Learning for Face Verification. In Proceedings of the International Conference on Computer Vision (ICCV), Sydney, NSW, Australia, 1–8 December 2013. [[CrossRef](#)]
39. Lu, C.; Tang, X. Surpassing Human-Level Face Verification Performance on LFW with GaussianFac. In Proceedings of the Computer Vision and Pattern Recognition (CVPR), Columbus, OH, USA, 23–28 June 2014.
40. Sun, Y.; Liang, D.; Wang, X.; Tang, X. DeepID3: Face Recognition with Very Deep Neural Networks. [Online]. Available online: <http://arxiv.org/abs/1502.00873> (accessed on 1 June 2015).
41. Barkan, O.; Weill, J.; Wolf, L.; Aronowitz, H. Fast High Dimensional Vector Multiplication Face Recognition. In Proceedings of the International Conference on Computer Vision (ICCV), Sydney, NSW, Australia, 1–8 December 2013. [[CrossRef](#)]
42. Cao, Q.; Ying, Y.; Li, P. Similarity Metric Learning for Face Recognition. In Proceedings of the International Conference on Computer Vision (ICCV), Sydney, NSW, Australia, 1–8 December 2013. [[CrossRef](#)]
43. Chen, D.; Cao, X.; Wen, F.; Sun, J. Blessing of dimensionality: Highdimensional feature and its efficient compression for face verification. In Proceedings of the 2013 IEEE Conference on Computer Vision and Pattern Recognition, Portland, OR, USA, 23–28 June 2013; pp. 3025–3032. [[CrossRef](#)]
44. Ding, C.; Tao, D. Trunk-Branch Ensemble Convolutional Neural Networks for Video-Based Face Recognition. *IEEE Trans. Pattern Anal. Mach. Intell.* **2018**, *40*, 1002–1014. [[CrossRef](#)] [[PubMed](#)]
45. Masi, I.; Chang, F.; Choi, J. Learning Pose-Aware Models for Pose-Invariant Face Recognition in the Wild. *IEEE Trans. Pattern Anal. Mach. Intell.* **2019**, *41*, 379–393. [[CrossRef](#)] [[PubMed](#)]



© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).