Robust Face Recognition Method Based on Kernel Regularized Relevance Weighted Discriminant Analysis and Deterministic Approach

Di Wu

Abstract
A novel feature dimensionality reduction strategy based on kernel regularized relevance weighted discriminant analysis is proposed in this paper with some interesting characteristics. First, the proposed method has shown its effectiveness in dealing with a small sample size problem when using the regularized linear discriminant analysis (RLDA) technique and Kernel theory. Second, a new computation method is proposed to solve the complicated and inefficient computation procedure problem in the traditional RLDA technique while using the cross-validation method. The experimental results indicate that the proposed algorithm shows better performance than the other methods.

Keywords  Face recognition · Linear discriminant analysis (LDA) · Kernel relevance weighted discriminant analysis (KRWDA) · Regularized linear discriminant analysis (RLDA) · Deterministic approach

1 Introduction

In the past decades, the field of biometrics has shown its usefulness and has come to play a major role in our daily life, especially in such applications as surveillance [1], human–computer interface, and security identification. Biometrics recognition
is an automatic method of recognizing people by means of comparing feature vectors derived from their physiological and behavioral traits. The common features used include the face, voice, fingerprints, palm, iris, ear, and so on. We can obtain facial images using inexpensive fixed cameras without the user’s active participation. Hence, face recognition systems are more easily accepted by people and have been receiving more significant research attention in recent years. As a result, numerous algorithms have been developed in this area, including a detailed survey found in [2].

The linear subspace analysis method has been widely used in face recognition research for linear dimensionality reduction and feature extraction [3]. Linear discriminant analysis (LDA) is a well-known technique, which seeks to project the input data into a low-dimensional space through the optimal projection directions while maximizing the between-class scatter and minimizing the within-class scatter. In order to limit the initial LDA algorithm, many variant-improved strategies have been proposed, including the kernel discriminant analysis (kDa), LDA/QR technology, relevance weighted LDA (RW-LDA), regularized LDA (RLDA) [4, 5], etc. These algorithms all improve the recognition performance and handle the Small Sample Size (SSS) problem in some ways.

Recent studies [6] have shown that a hybrid modified LDA algorithm, which can work in both linear and nonlinear (kernel theory) dimensionality reduction algorithms [7, 8], can help find a better one in both algorithms. Hence, in this paper, we combine the linear and nonlinear LDA algorithms in a principled way. We use RLDA and kernel relevance weighted discriminant analysis (KRWDA) for the linear and nonlinear LDA algorithms, respectively. Moreover, for the traditional RLDA, we use cross-validation technique, which we denote as CV-RLDA. As the computation is complex and the parameter changes between \([a, b]\), it may not be able to estimate its value. Finally, the regularization parameter can be estimated using the CV-LDA method for a particular classifier wherein the special value cannot be generalized.

In this paper, we propose a new feature dimensionality reduction algorithm for a robust face recognition method based on kernel regularized relevance weighted discriminant analysis (KRRWDA). First, we develop a new method called KRRWDA together with the RLDA technique to solve the SSS problem. The regularized parameter heavily influences the performance of the RLDA algorithm. In this paper, we attempt to find a deterministic way to optimize the regularized parameter.

The main contributions of this paper can be summarized as follows:

1. We combine the RLDA technique with the KRWDA algorithm to produce a new algorithm, which employs not only the merits of the kernel trick but also the RLDA technique.
2. A deterministic way of finding the optimal regularized parameter is given in this paper in order to avoid the use of the cross-validation method and improve the computation efficiency.

The remainder of this paper is organized into sections. The related studies are discussed in Sect. 2. In Sect. 4, we provide the detailed description of the proposed KRRWDA algorithm. In Sect. 5, the experimental results on various public
facial databases under different environment are given. Finally, in Sect. 6, the concluding remarks and recommendations for future works are presented.

2 Related Works

In this section, we briefly provide a survey of studies on LDA algorithm to solve Small Sample Size (SSS) problems. The LDA technique finds an orientation $W$ that transforms high dimensional feature vectors belong to different classes to a lower dimensional space that the projected feature vectors of a class on this lower dimensional space are well separated from the feature vectors of other space. If the dimensional reduction is from $d$ dimensional space to $h$ dimensional space, then the size of the orientation matrix $W$ it has $h$ column vectors. In other words, by the eigenvalue decomposition (EVD) of $S_w^{-1}S_B$, for a $c$ class problem, the value of $h$ will be $\min(c - 1, d)$. If $d$ is very large compared to the training vector $n$, the within-class matrix $S_w$ is singular and the computation of its inverse becomes impossible. This drawback is considered to be the main problem of the LDA and is commonly known as the Small Sample Size (SSS) problems.

In order to solve this thickest problem, we approximate the computation of matrix $W$ and avoid the inverse computation of the within-scatter matrix $S_w$. Therefore, various improved techniques are generated to overcome the SSS problem.

The most classical method to solve the SSS problem is to perform the PCA algorithm first when applying the LDA algorithm as the preprocessing step, with the aim of reducing the dimensionality [9]. In [10, 11], a QR decomposition called LDA/QR method is used to solve the eigenvalue problem. In [12], a pseudoinverse LDA (PLDA) method is presented to estimate the inverse of the within-scatter $S_w$ and then compute the orientation matrix $W$ to tackle the eigenvalue problem. In [13], a direct LDA (DLDA) is presented, and the dimensionality step can be divided into two stages: the first stage transforms all the training samples into the range space of $S_B$ by the computation of transformation matrix, and the second stage transforms the dimensionality of this transformed samples by regulating matrix.

Moreover, a past study [14] reported that the classical LDA algorithm optimization is not crucial in achieving classification accuracy. Moreover, sample projection can cause unnecessary overlap between neighboring classes, because the projection preserves the distance of well-separated classes. In order to tackle this thickest problem, that study [15] extended the classification criterion by introducing a new weighting function in the estimation process of $S_B$. Recently, in [16], an enhanced LDA algorithm called relevance weighted LDA (RWLDA) is presented, which uses the interclass relationships as the relevance weights to estimate the within-scatter matrix and replaces the unweight LDA [17]. The kernel discriminant analysis (kDa) has been proposed and proven to be as effective as the KPCA [18]. In [19], Liu presented a novel kernel discriminant strategy that not only generates the nonlinear features but can also deal with the troublesome singularity problem of the within-scatter matrix $S_w$. In [20], we introduced a new kernel method called kernel relevance weighted discriminant analysis (KRWDA), which is based on the RWLDA technique.
In [21], the researcher presented a novel reduction strategy called null LDA (NLDA), which can be computed into two stages: first, the training samples are projected into the null space of $S_W$ and, second, the orientation matrix $W$ is obtained by maximizing $|W^T S_B W|$. In [22], a novel technique called orthogonal LDA (OLDA) is presented, in which we diagonalize the scatter matrix simultaneously in order to obtain the orientation matrix $W$. Our experiment results show that the presented method has the same performance with the NLDA under a mild condition. A detailed overview of the SSS problem regarding LDA is given in [23, 24].

3 The KRWDA and RLDA Techniques

We first provide a brief overview of the KRWDA algorithm in Sect. 3.1. Then, the notation of the RLDA method is given in Sect. 3.2.

3.1 An Overview of the KRWDA Algorithm

In this paper, we denote some notations as follows: suppose the training set consists of $N$ training images $\{x_i\}_{i=1}^N$; in the training set, each image is defined as a vector and the length and size of the facial image are $d(w \times h)$ and $w \times h$, respectively [25]. We also assumed that the training set has $C$ class and every facial image is one of the $C$ classes $\{X_i\}_{i=1}^C$, each having $n_i$ images; thus, we arrive at $N = \sum_{i=1}^C n_i$.

For the classical LDA, the purpose is to minimize the within-class scatter $S_W$ while maximizing the between-class scatter $S_B$ in the lower feature space [26] as presented in Eqs. (1) and (2)

$$S_B = \sum_{i=1}^C P_i (m_i - m)(m_i - m)^T,$$

$$S_W = \sum_{i=1}^C P_i \sum_{j=1}^{n_i} (x_{ij} - m_i)(x_{ij} - m_i)^T,$$

where $m_i$ stands for the mean of the class $i$ with prior probability $P_i = \frac{n_i}{N}$, $m$ is the total mean, and $x_{ij}$ is the $j$th face sample from class $i$. Finally, using the solution of the criterion, we can obtain the optimal transformation $W$. The criterion is given by [27]

$$J(W) = \frac{|W^T S_B W|}{|W^T S_W W|}. \tag{3}$$

However, in the traditional LDA criterion, in the lower-dimensional space, the classification ability is impaired because the classify criterion may not be optimal in terms of minimizing the classification error rate. In order to solve this thickest problem, Loog
et al. [28] added a weighting function into the discriminant criterion. In their new criterion, the within-class scatter and the between-class scatter are defined respectively as

\[ \hat{S}_W = \sum_{i=1}^{C-1} \sum_{j=i+1}^{C} w(d_{ij}) P_i P_j (m_i - m)(x_i - m)^T, \]

\[ \hat{S}_W = \sum_{i=1}^{C} P_i r_i \sum_{j=1}^{n_i} (x_{ij} - m_i)(x_{ij} - m_i)^T. \]

The value of the weighting function \( w(d_{ij}) \) depends on the Euclidean distance between the means of class \( i \) and class \( j \) [29] given by

\[ w(d_{ij}) = \|m_i - m_j\|^{-2h}. \]

The weights \( r_i \) are based on the relevance, and is given by [30]

\[ r_i = \sum_{j \neq i} \frac{1}{w(d_{ij})}. \]

As a result, the RWLDA and the discriminant criterion can be transformed as (8) by using the new defined between-scatter matrices \( \hat{S}_b \) and within-scatter matrices \( \hat{S}_W \) [30].

\[ \hat{J}(W) = \frac{|W^T \hat{S}_b W|}{|W^T \hat{S}_W W|} \]

Using the classical kernel trick, we respectively define a weighted between-class scatter and a weighted within-class scatter in the implicit feature space \( F \)

\[ \hat{S}_b^\phi = \sum_{i=1}^{C-1} \sum_{j=i+1}^{C} w(d_{ij}) P_i P_j (m_i^\phi - m_j^\phi)(m_i^\phi - m_j^\phi)^T, \]

\[ \hat{S}_W^\phi = \sum_{i=1}^{C} P_i r_i \sum_{j=1}^{n_i} (\phi(x_{ij}) - m_i^\phi)(\phi(x_{ij}) - m_i^\phi)^T. \]

Then, we can define a novel discriminant classify criterion in the implicit space \( F \) based on the new defined \( \hat{S}_b^\phi \) and \( \hat{S}_W^\phi \) presented in Eqs. (9) and (10) above. The new equation is given by

\[ \hat{J}_\phi(W) = \frac{|W^T \hat{S}_b^\phi W|}{|W^T \hat{S}_W^\phi W|} \]
At the same time, using the kernel trick, similar to the traditional kDa, the Fisher criterion defined in (11) can also be transformed as

$$\hat{J}_f(A) = \frac{A^T \hat{K}_b A}{A^T \hat{K}_W A},$$

(12)

Therefore, the solution of Eq. (12) $A = [\alpha_1, \ldots, \alpha_m]$ regarding the optimization problem is formed by the $m$ leading eigenvectors of matrix $(K_W^{-1} \hat{K}_b)$ corresponding to the nonzero eigenvalues.

### 3.2 The Notation of the RLDA Algorithm

In this section, we denote the notation about RLDA [31]. In the definition of the RLDA method, the tedious singularity problem of $S_W$ can be solved by a regularization operation. Often, the regularization parameter $\alpha$ is a small positive constant, and we add it to the diagonal elements of matrix $S_W$. The improved discriminant criterion is given by

$$J(W) = \frac{|W^T S_b W|}{|W^T (S_W + \alpha I)W|}.$$  

(13)

Then, the orientation matrix is computed by EVD of $(S_W + \alpha I)^{-1} S_b$, and the proper choice of the regularization parameter $\alpha$ is crucial for the performance of the RLDA strategy.

### 4 The Proposed KRRWDA Algorithm

In this section, we define the proposed KRRWDA algorithm to achieve an improved performance.

First, we incorporate the RLDA technique into the KRWDA to overcome the SSS problem and overemphasize the well-separated classes. Second, to estimate the parameter $\alpha$, we attempt a deterministic approach in the proposed new method, which is derived from the RLDA technique.

To make it possible, we must redefine the within-class scatter and between-class scatter previously given in Eqs. (4) and (5). The new equations are respectively expressed as

$$S_{b_{pro}} = \hat{S}_b = \sum_{i=1}^{C-1} \sum_{j=i+1}^{C} w(d_{ij}) P_i P_j (m_i - m)(m^\phi_i - m)^T,$$

(14)
The notations \( S_{bpro} \) and \( S_{Wpro} \) represent the between-class scatter and within-class scatter, respectively, which are used in the KRRWDA method proposed in this paper. Thus, the Fisher criterion proposed in Eq. (8) is replaced by

\[
J(W) = \frac{|W^T S_{bpro} W|}{|W^T S_{Wpro} W|} = \frac{|W^T \hat{S}_b W|}{|W^T (\hat{S}_W + \alpha I) W|}.
\]  

(16)

Similar to the mathematic scheme we used in the KRWDA algorithm, in this section, we also use the kernel trick to deal with the SSS problem in an implicit feature space \( F \). Thus, we can obtain

\[
\hat{S}_{bpro}^\phi = \hat{S}_b = \sum_{i=1}^{C-1} \sum_{j=i+1}^C w(d_{ij}) P_i P_j (m_i^\phi - m_j^\phi)(m_i^\phi - m_j^\phi)^T \tag{17}
\]

\[
\hat{S}_{Wpro}^\phi = \hat{S}_W + \alpha = \left( \sum_{i=1}^C P_i r_i \sum_{j=1}^{n_i} (\phi(x_{ij}) - m_i^\phi)(\phi(x_{ij}) - m_i^\phi)^T + \alpha I \right) \tag{18}
\]

Based on \( \hat{S}_{bpro}^\phi \) and \( \hat{S}_{Wpro}^\phi \), we define a new Fisher discriminant criterion in \( F \), which is given by

\[
J(W) = \frac{|W^T \hat{S}_{bpro}^\phi W|}{|W^T \hat{S}_{Wpro}^\phi W|} = \frac{|W^T \hat{S}_b^\phi W|}{|W^T (\hat{S}_W^\phi + \alpha I) W|}. \tag{19}
\]

As shown in Sect. 3, the Fisher discriminant criterion in (24) can be transformed into

\[
J(A) = \frac{|A^T \hat{K}_{bpro}^\phi A|}{|A^T \hat{K}_{Wpro}^\phi A|} = \frac{|A^T \hat{K}_b^\phi A|}{|A^T (\hat{K}_W^\phi + \alpha I) A|}. \tag{20}
\]

In the following text, we discuss the computation of the regularized parameter \( \alpha \) in a novel deterministic way. First, let us denote the two functions as

\[
f_{bpro} = A^T \hat{K}_{bpro}^\phi A = A^T \hat{K}_b^\phi A, \tag{21}
\]

\[
f_{Wpro} = A^T \hat{K}_{Wpro}^\phi A = A^T (\hat{K}_W^\phi + \alpha I) A - b = 0. \tag{22}
\]
where $b$ is larger than zero and is any constant Fisher criterion aimed to make the difference between classes larger; the difference within classes must be smaller to find the maximum of $f_{bpro}$ under the constraint. Let us further define a function as

$$F = f_{bpro} - \lambda f_{Wpro}$$  \hspace{1cm} (23)

where $\lambda$ is the Lagrange’s multiplier. Then, by setting its derivative to zero, we can obtain

$$\frac{\partial F}{\partial A} = \frac{\partial (f_{bpro} - \lambda f_{Wpro})}{\partial A} = 2K_b^\phi A - \lambda (2K_b^\phi A + 2\alpha A) = 0$$  \hspace{1cm} (24)

Equation (24) can be transformed as

$$\left( \frac{1}{\lambda} \hat{K}_b^\phi - \hat{K}_W^\phi \right) = \alpha A$$  \hspace{1cm} (25)

Substituting the value of $\alpha A$ from Eq. (25) into (22), we can obtain

$$f_{Wpro} = A^T \hat{K}_W^\phi A = A^T (\hat{K}_W^\phi + \alpha A)A - b$$

$$= A^T \hat{K}_W^\phi A + A^T \left( \frac{1}{\lambda} \hat{K}_b^\phi - \hat{K}_W^\phi \right)A - b,$$

$$= 0$$

where

$$A^T \hat{K}_b^\phi A = \lambda b.$$  \hspace{1cm} (27)

we obtain

$$\lambda = \frac{A^T \hat{K}_b^\phi A}{A^T (\hat{K}_W^\phi + \alpha I)A}$$  \hspace{1cm} (28)

$$\lambda_{\text{max}} = \max \left( \frac{A^T \hat{K}_b^\phi A}{A^T (\hat{K}_W^\phi + \alpha I)A} \right) \approx \max \left( \frac{A^T \hat{K}_b^\phi A}{A^T \hat{K}_b^\phi A} \right) \approx \lambda_{\text{arg est}} \text{ eigen value of } (\hat{K}_W^\phi)^{-1} \hat{K}_b^\phi$$  \hspace{1cm} (29)

In order to obtain the proposed Fisher discriminant criterion, we set $\lambda$ so that it is equal to the maximum of $J(A)$. We must also have eigenvector $A$ to correspond to the maximum eigenvalue of $(\hat{K}_W^\phi)^{-1} \hat{K}_b^\phi$. Thereby, the evaluation of $\alpha$ can be carried out from Eq. (25) by performing the EVD of $\left( \frac{1}{\lambda} \hat{K}_b^\phi - \hat{K}_W^\phi \right)$ where $\lambda = \lambda_{\text{max}}$. After evaluating the optimal parameter $\alpha$, orientation vector $s$ can be obtained by performing the EVD of $\left( \hat{K}_W^\phi + \alpha I \right)^{-1} \hat{K}_b^\phi$ from

$$\left( \hat{K}_W^\phi + \alpha I \right)^{-1} \hat{K}_b^\phi A - rA$$  \hspace{1cm} (30)
5 Results and Discussion

In order to assess the performance of the proposed novel KRRWDA strategy, we test the method on different well-known facial data databases. First, we introduce several related competing algorithms and some experiment settings in Sect. 5.1. In Sect. 5.2, the determination of the optimal parameters in the KRRWDA algorithm is presented. In Sect. 5.3, using two popular illumination variation facial database (Multi-PIE and FRGC), we test the robustness of the given KRRWDA method against different illumination changes. In Sect. 5.4, we evaluate the attitude and facial expressions found in the FERET and LFW facial databases. In Sect. 5.5, the computation complexity is presented.

5.1 The Experiments’ Detailed Settings

In order to test the performance of the proposed KRRWDA algorithm, we choose several classical and popular face recognition algorithms for the performance comparison on different aspects, such as the baseline Eigenface method, Fisherface, OTF-based CFA [32], sparse representation-based classification (SRC) [33], and some state-of-the-art local FE methods, including Block-FLD (B-FLD) [34], Cascade LDA (C-LDA) [35], and Block-based Bag of Words (BBOW) [36]. The RLDA technique and KRWDA algorithms are also compared. For the proposed KRRWDA algorithm, we evaluate the KRRWDA(CV) method (using the cross-validation method for computing the regularized parameter $\alpha$) and the KRRWDA(DE) method (using the proposed deterministic method for computing the regularized parameter $\alpha$).

In order to extract the facial features in the face databases and ensure that the facial region only contains the face, each facial image is normalized. At the beginning, the eyes are manually annotated at the center. Then, we align the centers of the eyes to the designated locations after the rotate and scale transformations and compute the distances between the two eyes. Finally, we crop the image size so that it becomes $64 \times 64$ pixels. Finally, the standard histogram is utilized for the photometric normalization.

In the experiment, when extracting the facial feature for both training set and test set, the nearest neighbor classifier is employed for the final classification, and the cosine similarity measure is used for the measurement. In the training set, every set is formed with $t$ images, and the rest of the images of the database are used for the test set. The experiments are performed 20 times, and the average recognition rates and the standard deviations are used for the final results.

In this paper, we mainly solve the most challenging issues in face recognition called the SSS problem, which is one of the most challenging and thickest issues in face recognition. The value of $t$ is set to 2–5 for all facial databases to evaluate the effectiveness of different feature extraction methods in solving the SSS problem.
5.2 Determining the Related Parameters

In the KRRWDA technique, two important parameters (regularized parameter $\alpha$ and kernel functions), which affect the accuracy of face recognition are determined. On the one hand, the regularized parameter should be carefully set in order to reduce the high variance caused by the SSS problem, which can influence the estimation of the covariance matrix. On the other hand, the kernel function is crucial for the kernel trick. Hence, in this sub-section, we compare the performances of the polynomial kernel function, Gaussian RBF kernel function, and the fractional polynomial kernel function.

First, we determine the regularized parameter $\alpha$ used in the KRRWDA technique. Table 1 shows the values of $\alpha$ using the different facial databases.

Second, for the kernel trick, different kernel functions have different performances; thus, in this part, we compare the performance between the above-mentioned kernel functions in the AR facial database using the KRRWDA(CV) algorithm.

The AR database consists of 126 subjects (70 men and 56 women) and has more than 4000 facial images. In the database, each image incorporates different facial expressions, occlusion modes, and illuminations changes. In this experiment, a subset belonging to the AR database contains 120 subjects, and each subject has 14 images. Each image shows facial expression and a state of illumination. See Fig. 1 for some examples.

In this experiment, we only focus on the performance influence of the kernel functions and on the computation of the parameter $\alpha$ using the cross-validation (CV) technique. From Table 2, we report the results of the performance.

From the recognition results shown in Table 2, we find that the best recognition result is obtained using the Gaussian RBF kernel function $\sigma^2 = 10^5$. Thus, in the rest of this paper, we use the Gaussian RBF kernel function $\sigma^2 = 10^5$.

### Table 1 The regularized parameter

| Database   | AR      | Multi-PIE | FRGC     | FERET    | LFW     |
|------------|---------|-----------|----------|----------|---------|
| $\alpha$   | 5715.6  | $6.54 \times 10^9$ | $1.11 \times 10^{11}$ | $1.24 \times 10^9$ | $2.29 \times 10^{10}$ |

![Sample facial images from the AR database](image_url)
5.3 Robustness to Illumination Variations

In face recognition, the most classical fundamental challenges are facial appearance variations because of illumination changes. Thus, in this part, we assess the robustness of the proposed KRRWDA method versus illumination variations on two famous facial databases: Multi-PIE and FRGC.

The Multi-PIE database has 337 subjects with more than 750,000 facial images and different variations in poses and illuminations; all images are captured in four sessions. Figure 2 shows the 20 facial images belonging to one subject in this database. The FRGC database is composed of three categories of facial images for each subject: controlled images, uncontrolled images, and three-dimensional facial images. Here, we decide to use a subset containing 6000 images of 300 subjects. In Fig. 3, the facial images of one subject of the FRGC database are shown. In Tables 3 and 4, we can see the average recognition performance obtained by the different algorithms on the Multi-PIE and FRGC databases, respectively.

From the results shown in these two tables, we can see that the best recognition accuracies are obtained by the proposed KRRWDA(DE) methods among the other competing methods. Compared with the KRRWDA(CV) algorithm, the proposed KRRWDA(DE) algorithm improves the recognition performance by about 4%–5%. Comparing the results, we also find that the SRC algorithm is better than

| t | d=2 | d=3 | d=4 | \( \delta^2 = 10^2 \) | \( \delta^2 = 10^7 \) | \( \delta^2 = 10^9 \) | d = 0.2 | d = 0.4 | d = 0.6 |
|---|-----|-----|-----|----------------|-----------------|----------------|-------|-------|-------|
| 2 | 81.28 | 79.85 | 77.09 | 83.31 | 82.06 | 82.53 | 81.34 | 83.37 | 83.28 |
| 3 | 91.96 | 90.21 | 88.78 | 91.92 | 91.30 | 91.14 | 91.25 | 92.47 | 91.71 |
| 4 | 96.54 | 95.29 | 93.16 | 96.20 | 95.08 | 96.75 | 95.37 | 95.45 | 95.16 |
| 5 | 97.05 | 96.05 | 96.90 | 98.60 | 97.07 | 97.40 | 96.95 | 96.70 | 96.25 |

Fig. 2 The facial images of one subject in the Multi-PIE database

Fig. 3 The facial images of one subject in the FRGC database
the B-FLD method, suggesting that the SRC method is robust in dealing with illumination variations.

5.4 Robustness to Variations in Poses and Facial Expressions

In this section, we evaluate the effect of pose and facial changes on the performance of the given novel method by using the FERET and the LFE databases.

To evaluate the performance of the face recognition algorithms, the FERET database can be used as a standard facial database. As shown in Fig. 4, a subset of the FERET database is used, which includes 1400 images of 200 subjects. Some challenges, such as variations in facial expression and poses, are included in this subset. Furthermore, we have also performed an experiment on the LFW facial database, which is a more realistic facial database captured in unconstrained environments and is often used to evaluate face recognition algorithms in real scenarios. The LFW database contains the images of 5749 different individuals collected from the web. LFW-a is a version of the LFW after face alignment, and a subset with
150 subjects is chosen. The samples of one subject on the LFW database used in this experiment are shown in Fig. 5. Tables 5 and 6 show the experimental results using the FERET and LFW databases, respectively. As can be seen, the proposed KRRWDA(DE) recognition method obtains comparable or better recognition rates than other competing algorithms. Particularly, the performance of KRRWDA(DE) increases significantly when more training samples are used. The recognition performance of CFA-OTF is lower than that obtained by KRRWDA(DE), because the use of the whole face region makes CFA-OTF sensitive to pose variations. In contrast, KRRWDA(DE) alleviates this problem by using kernel trick and RLDA technique. In addition, BBOW obtains lower recognition rate than KRRWDA(DE) in the LFW facial database, indicating that when the training set contains the changes of posture and facial expression, BBOW cannot effectively capture the internal difference information.

### Table 5
The performances obtained by the different algorithms on the FERET database

| Algorithm     | $t=2$       | $t=3$       | $t=4$       | $t=5$       |
|---------------|-------------|-------------|-------------|-------------|
| Eigenface     | 53.27 ± 3.0 | 60.12 ± 2.9 | 65.50 ± 2.8 | 70.22 ± 2.1 |
| Fisherface    | 66.63 ± 1.8 | 67.79 ± 1.7 | 76.23 ± 1.4 | 77.54 ± 1.3 |
| CFA-OTF       | 58.96 ± 1.7 | 65.53 ± 1.5 | 74.18 ± 1.1 | 78.97 ± 1.4 |
| SRC           | 66.21 ± 2.1 | 67.14 ± 2.3 | 71.16 ± 1.2 | 73.36 ± 1.9 |
| B-FLD         | 67.57 ± 1.8 | 69.95 ± 1.4 | 73.28 ± 1.7 | 80.95 ± 1.6 |
| C-LDA         | 68.83 ± 0.9 | 70.17 ± 2.3 | 75.36 ± 2.4 | 83.27 ± 2.3 |
| BBOW          | 74.15 ± 0.8 | 77.42 ± 1.2 | 86.00 ± 1.3 | 92.34 ± 1.5 |
| KRRWDA (CV)   | 75.10 ± 1.9 | 81.14 ± 1.8 | 90.25 ± 1.1 | 92.11 ± 1.4 |
| CNN           | 80.60 ± 1.4 | 84.72 ± 1.3 | 94.26 ± 1.2 | 95.85 ± 1.1 |
| KRRWDA (DE)   | 83.16 ± 1.5 | 86.14 ± 2.4 | 95.56 ± 1.1 | 96.84 ± 1.3 |

**Fig. 4** The facial images in the FERET database

**Fig. 5** The facial images in the LFW database
Compared with the recognition performance obtained by using other facial databases, the proposed KRRWDA(DE) algorithm obtains lower recognition accuracies on the LFW database because of two reasons: (1) some facial images contain the surrounding background, which decreases the discriminability of features extracted by our proposed algorithm, and (2) when dealing with large pose variations, the training samples and test samples may mismatch at some instances.

### 5.5 The Computational Complexity of the Proposed Algorithm

In this section, we compare and analyze the Eigenface, Fisherface, CFA-OTF, and KRRWDA(CV) methods using the computational time of the proposed KRRWDA(DE) algorithm with extraction algorithm. The workstation has two inter Xeon E5620 CPUs on the MATLAB platform, and all the computational times are reported based on this. Table 7 shows the computational time spent on the training set and test stages and all these are based on the algorithm of the CAS-PEAL facial database.

As shown in Table 7, the computational time of the proposed KRRWDA(DE) is higher than those of the other algorithms except the KRRWDA(CV) technique. Hence, the computational complexity of the proposed algorithm will not

### Table 6 The performances obtained by the different algorithms on the LFW database

| Algorithm    | $t = 2$     | $t = 3$     | $t = 4$     | $t = 5$     |
|--------------|-------------|-------------|-------------|-------------|
| Eigenface    | 24.15 ± 3.2 | 28.10 ± 3.2 | 32.23 ± 3.5 | 37.00 ± 3.7 |
| Fisherface   | 27.89 ± 2.8 | 33.42 ± 1.3 | 38.42 ± 2.4 | 44.25 ± 2.5 |
| CFA-OTF      | 25.27 ± 3.5 | 30.17 ± 0.8 | 32.17 ± 1.6 | 35.24 ± 3.5 |
| SRC          | 30.25 ± 2.5 | 35.24 ± 2.3 | 39.98 ± 1.2 | 45.15 ± 1.9 |
| B-FLD        | 32.53 ± 2.3 | 36.78 ± 2.4 | 40.12 ± 1.9 | 45.24 ± 1.5 |
| C-LDA        | 31.10 ± 2.2 | 35.41 ± 2.1 | 38.82 ± 1.5 | 44.99 ± 1.8 |
| BBOW         | 31.27 ± 1.9 | 33.41 ± 1.9 | 41.27 ± 1.9 | 48.21 ± 1.7 |
| KRRWDA (CV)  | 37.17 ± 1.1 | 42.17 ± 2.1 | 47.05 ± 1.8 | 50.72 ± 1.3 |
| CNN          | 38.20 ± 1.6 | 43.10 ± 1.7 | 48.58 ± 1.4 | 52.20 ± 1.4 |
| KRRWDA (DE)  | 43.16 ± 1.5 | 46.14 ± 2.4 | 55.56 ± 1.1 | 56.84 ± 1.3 |

### Table 7 The comparison of the computational times (in seconds) used by the competing algorithms

| Algorithm    | Training time | Recognition time |
|--------------|---------------|------------------|
| Eigenface    | 51.41         | 70.61            |
| Fisherface   | 83.74         | 20.62            |
| CFA-OTF      | 522.74        | 32.88            |
| KRRWDA (CV)  | 3134.47       | 123.80           |
| KRRWDA (DE)  | 2202.78       | 82.64            |
| CNN          | 2487.14       | 102.75           |
constrain its application to real-world tasks because the training stage is often performed offline.

6 Conclusion

In order to solve the problem on the computation efficiency of the regularized parameter $\alpha$, a computation method is proposed in this paper. With the aim of testing the performance of this novel feature extraction technique, we have evaluated the KRRWDA(DE) method in different cases, such as variations in illumination, facial expressions, and poses, while dealing with the SSPP problem. The experimental results have proven that, at present, in solving the SSS problem, KRRWDA(DE) outperforms the most advanced feature extraction algorithms when tested using public facial databases.

As mentioned in our experiments, the proposed method cannot handle face recognition with very large pose variations well. However, the recent study has demonstrated that the Deep Learning (DP) method (DP) can help improve the recognition performance. Hence, how to design an effective face recognition based on DP and space method under the condition of multiple pose variations is an interesting direction of our future work. In addition, we are also interested in exploring the effective computation method of the regularized parameter $\alpha$.

Acknowledgments The author would like to thank the anonymous reviewers for their constructive comments. This work was supported by the National Natural Science Foundation of China under Grant Nos. 61263031 and 61841103, by the National Natural Science Foundation of Hunan Province of China under Grant Nos. 2018JJ4039, 2019JJ50106, and by the Key Project of Education Department of Hunan Province of China under Grant No. 17A048, 18B385.

References

1. Seal, A., Bhattacharjee, D., & Nasipuri, M. (2016). Human face recognition using random forest based fusion of à-trous wavelet transform coefficients from thermal and visible images. AEU International Journal of Electronics and Communications, 70(8), 1041–1049.
2. Kar-Ann Toh. (2012). Multimodal biometrics: an overview and some recent developments. In Biometrics engineering research center report, pp. 381–400.
3. Heikkila, M., & Pietikainen, M. (2009). Description of interest regions with local binary patterns. Pattern Recognition, 42(3), 425–436.
4. Sharma, A., & Paliwal, K. K. (2015). A deterministic approach to regularized linear discriminant analysis. Neurocomputing, 151, 207–214.
5. Zhong, F., & Zhang, J. (2013). Face recognition with enhanced local directional patterns. Neurocomputing, 119, 375–384.
6. Ramirez-Rivera, A., Rojas-Castillo, J., Oksam Chae, O., et al. (2013). Local directional number pattern for face analysis:Face and Expression recognition. IEEE Transaction Image Processing, 22(5), 1740–1752.
7. Cottrell, G. W. (2006). New life for neural networks. Science, 313(5786), 454–455.
8. Hinton, G. E., Osindero, S., & Teh, Y.-W. (2006). A fast learning algorithm for deep belief nets. Neural Computation, 18(7), 1527–1554.
9. Yan, Y., Wang, H., & Suter, D. (2014). Multi-subregion based correlation filter bank for robust face recognition. *Pattern Recognition, 47*, 3487–3501.

10. Fujin, Zhong, & Jiashu, Zhang. (2012). Face recognition with improved local directional patterns. *Journal of Optoelectronics, 23*(8), 1570–1575.

11. Li, Li, Ge, Hongwei, & Gao, Jianqiang. (2016). Maximum–minimum–median average MSD-based approach for face recognition. *AEU International Journal of Electronics and Communications, 70*(7), 920–927.

12. Huang, P., & Gao, G. (2015). Local similarity preserving projections for face recognition. *AEU: International Journal of Electronics and Communications, 69*(11), 1724–1732.

13. Feng, W., & XiaoBo, J. (2015). A positive recognition algorithm for fake plate vehicles based on fusion strategy. *Journal of Optoelectronics. Laser, 26*(11), 2209–2216.

14. Wu, D. (2015). Audio-visual person recognition based on rank level fusion and gaussian mixture model. *International Journal of Control and Automation, 8*(4), 313–332.

15. Wu, D. (2015). Multimodel biometrics Fusion based on FAR and FRR using Triangular Norm. *International Journal of Computational Intelligence Systems, 8*(4), 779–786.

16. Liang, Yi Cong, Ding, Xiao Qing, Liu, Chang Song, & Xue etc, Jing Hao. (2016). Combining multiple biometric traits with an order preserving score level fusion algorithm. *Neurocomputing*, 171, 252–261.

17. Ladyzynski, P., & Grzegorzewski, P. (2015). Vague preferences in recommender systems. *Expert Systems with Applications, 42*(24), 9402–9411.

18. Pohan, N., Ross, A., Leec, W., & Kittler, J. (2013). A user-specific and selective multimodal biometric fusion strategy by ranking subjects. *Pattern Recognition, 46*, 3341–3357.

19. Hang, L., Zhang, D., Zhu, H., et al. (2010). Online finger-knuckle-print verification for personal authentication. *Pattern Recognition, 43*(7), 2560–2571.

20. Sung, E., & Yau, W.-Y. (2010). Face recognition in global harmonic subspace. *IEEE Transactions on Information Forensics and Security, 5*(3), 416–424.

21. L. Wolf, T. Hassner and I. Maoz, Face recognition in unconstrained videos with matched background similarity, in: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2011, pp. 529–534.

22. Huynh, T., Min, R., & Dugelay, J. L. (2013). An efficient LBP-based descriptor for facial depth images applied to gender recognition using RGB-D face data. In *Proceedings of computer vision-ACCV 2012 workshops* (pp. 133–145). Springer, Berlin

23. Choi, K., Toh, K.-A., & Byun, H. (2011). Realtime training on mobile devices for face recognition applications. *Pattern Recognition, 44*(2), 386–400.

24. Chakraborty, S., Balasubramanian, V., & Panchanathan, S. (2013). Generalized batch mode active learning for face-based biometric recognition. *Pattern Recognitient., 46*(2), 497–508.

25. Rim, D., Hasan, M. K., Puech, F., & Pal, C. J. (2015). Learning from weakly labeled faces and video in the wild. *Pattern Recognition, 48*(3), 759–771.

26. Juefei-Xu, F., & Savvides, M. (2015). Pokerface: partial order keeping and energy repressing method for extreme face illumination normalization. In *2015 IEEE seventh international conference on biometrics: theory, applications and systems (BTAS)*, pp. 1–8.

27. Juefei-Xu, F., & Savvides, M. (2015). Single face image super-resolution via solo dictionary learning. In *IEEE international conference on image processing (ICIP)*.

28. Juefei-Xu, F., & Savvides, M. (2015). Pareto-optimal discriminant analysis. In *IEEE international conference on image processing (ICIP)*.

29. Juefei-Xu, F., & Savvides, M. (2015). Encoding and decoding local binary patterns for harsh face illumination normalization. In *IEEE international conference on image processing (ICIP)*.

30. Juefei-Xu, F., Pal, D. K., Savvides, M. (2015). NIR-VIS heterogeneous face recognition via cross-spectral joint dictionary learning and reconstruction. In *2015 IEEE conference on computer vision and pattern recognition workshops (CVPRW)*.

31. Juefei-Xu, F., Pal, D. K., Singh, K., & Savvides, M. (2015). A preliminary investigation on the sensitivity of COTS face recognition systems to forensic analyst-style face processing for occlusions. In *2015 IEEE conference on computer vision and pattern recognition workshops (CVPRW)*.

32. Seshadri, K., Juefei-Xu, F., Pal, D. K., & Savvides, M. (2015). Driver cell phone usage detection on strategic highway research program (SHRP2) face view videos. In *2015 IEEE conference on computer vision and pattern recognition workshops (CVPRW)*.
33. Juefei-Xu, F., Pal, D. K., & Savvides, M. (2014). Hallucinating the full face from the periorbital region via dimensionally weighted K-SVD. In *2014 IEEE conference on computer vision and pattern recognition workshops (CVPRW)*.

34. G.B. Huang, E. Learned-Miller, Labeled faces in the wild: updates and new reporting procedures, in: *UMass Amherst Technical Report UM-CS-2014-003*, 2014.

35. Zhang, J., Shan, S., Kan, M., & Chen, X. (2014). Coarse-to-fine auto-encoder networks (CFAN) for real-time face alignment. In *Proc. ECCV*, 2014, pp. 1–16.

36. Chen, D., Cao, X., Wen, F., & Sun, J. (2013). Blessing of dimensionality: High-dimensional feature and its efficient compression for face verification. In *Proceedings of CVPR*, pp. 3025–3032.

**Publisher’s Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.