A Robust Face Recognition Algorithm Based on Kernel Regularized
Relevance-Weighted Discriminant Analysis

Di Wu 1, 2

1 College of Electrical and Information Engineering,
2 Hunan Provincial Key Laboratory of Wind Generator and Its Control, Hunan Institute of Engineering, Xiangtan, China.
[e-mail: wudi6152007@163.com]

Abstract - In this paper, we propose an effective feature dimensionality-reduction method, called Kernel Regularized Relevance Weighted Discriminant Analysis (KRRWDA), for robust face recognition, with several interesting characteristics. First, it can effectively deal with the small sample size (SSS) problem by using Regularized Linear Discriminant Analysis (RLDA) technique, which is a dimensionality reduction method in kernel theory. Second, in traditional RLDA technique, the regularization parameter is computed using a complicated cross-validation procedure of low computational efficiency, which we improve in this paper by a novel deterministic way to significantly reduce the computation cost. Experimental results on various public face databases demonstrate that the proposed algorithm provides a better feature dimension-reduction performance and achieves higher recognition rates compared with several state-of-the-art algorithms.

Keywords - face recognition; Linear Discriminant Analysis (LDA); Kernel Relevance Weighted Discriminant Analysis (KRWDA); Regularized Linear Discriminant Analysis (RLDA); deterministic approach.

I. INTRODUCTION

During the past few decades, biometrics have played a very useful role in many fields such as surveillance[1], human-computer interface and security identification. Biometrics recognition is an automatic method of recognizing peoples by means of comparing feature vectors derived from their physiological or behavioral characteristics. The common characteristics used here include face, voice, fingerprints, finger geometry, hand veins, palm, iris, retina, ear and so on. Facial images can be easily acquired by a few inexpensive fixed cameras without the user’s active participation. Hence, face recognition systems are more easily accepted by peoples and have been receiving more researchers significant attentions. As a result, numerous algorithms have been developed in this area and a detail survey can be found in [2].

Linear subspace analysis method have been wide used in face recognition research for linear dimensionality reduction and feature extraction[3]. Linear Discriminant Analysis (LDA) is the most well-known technique which seek to project the input data into a low dimensional space through the optimal projection directions while maximize the between class scatter and minimize the within class scatter.

However, a critical issue using LDA based methods for face recognition is the Small Sample Size (SSS) problem which is encountered in practice since the number of pixels available is large but the total number of training samples is less than the dimension of the feature space[4]. This phenomenon implies that all scatter metrics are singular and thus the traditional LDA algorithm fails to use.

To overcome this limitation many variants of original LDA methods have been proposed, such as Kernel Discriminant Analysis (KDA), LDA/QR technology, relevance weighted LDA (RW-LDA), Regularized LDA (RLDA) [5] etc. Among them, they all improve the recognition performance and handle the SSS problem in some ways.

In this paper, we propose an effective feature dimensionality reduction algorithm called kernel regularized relevance weighted discriminant analysis (KRRWDA) for robust face recognition. At first, we combine RLDA technique with KRWDA algorithm to develop a new method called KRRWDA and apply it to face recognition. More over, the performance of RLDA technique depends on the choice of the regularized parameter . In traditional RLDA method[6], the parameter is choose rather heuristically such as applying Cross-Validation procedure on the training data which denote as CV-LDA [7], but the cross-validation procedure used in the CV-LDA is biased towards the classifier used. In order to address these drawback of CV-LDA technique, this paper explore a deterministic way for finding the optimal regularized parameter .

Compared to the traditional algorithms, the proposed KRRWDA algorithm has the following characteristics: First, the proposed new algorithm can handle the SSS problem by using the kernel trick to map input data into a implicit feature space which produce nonlinear discriminant features that describe complex variations of face images and they can work on more realistic situations. Second, RLDA technique is one of the pioneering methods for solving SSS problem, in this paper we combine RLDA technique with KRWDA algorithm to produced a new algorithm which not only employ the merits of the kernel trick but also the RLDA technique. At last, we explore a deterministic way for finding the regularization parameter which avoids the use of the cross-validation procedure for parameter estimation and its can improves the computational efficiency.

The remainder of this paper is organized as follows. Related work is discussed in Section 2. In Section 3, we will briefly give the notation of KRWDA algorithm and RLDA method. A detailed description of the proposed KRRWDA
algorithm is presented in Section 4. The experimental results on various public face databases under different environment are given in Section 5. Finally, the concluding remarks and future work are provided in Section 6.

II. RELATED WORK

In this Section, we will briefly give a survey about LDA algorithm to solve SSS problems in Section 2.1. The motivation of this work is given in Section 2.2.

A. LDA Algorithms to Solve SSS Problems

In a SSS problem, the within-class matrix \( S_W \) becomes singular and its inverse computation becomes impossible, in order to solve this trickiest problem, generally inverse computation of \( S_W \) is avoid or approximated for the computation of orientation matrix \( W \). So there are various techniques are proposed in order to overcome this SSS problem.

The most popular algorithm is to perform PCA first as preprocessing step aiming to reduce dimensionality prior to applying LDA[9]. Recently, reference[10] have presented LDA/QR by using QR decomposition in order to solve the eigenvalue problem. Reference [11] propose a novel way to solve this problem by estimating the inverse of \( S_W \) by its pseudo inverse and then the conventional eigenvalue problem can be solved to compute the orientation matrix \( W \). This technology is called pseudo-inverse LDA (PLDA), [12].

In the Direct LDA (DLDA) [13] method, the dimensionality is reduced into two stage, at first, a transformation matrix is computed to transform all training images into the range space of \( S_B \), then the dimensionality of this transformed samples is further transformed by some regulating matrices.

Moreover, reference [14]. have show that the class separability criterion that classical LDA optimize is not necessarily representative of classification accuracy and the resulting projection will preserve the distances of already well-separated classes while causing unnecessarily overlap of neighboring classes, to tackle this problem, reference[15]. have proposed an extended criterion by introducing a weighting function in the estimation of \( S_B \). More recently, reference [16] have introduced the inter class relationships as relevance weights to estimate the \( S_W \), they have presented an LDA enhancements algorithm call relevance weighted LDA(RWLDA) by replacing the unweighted LDA through the weighted scatter matrices in the classical LDA method[17]. In the same spirit as the kernel extensions of PCA, kernel extension of LDA which called kernel discriminant analysis (KDA) has also been developed and found to be more effective than PCA[18].

Liu et al.[19] have introduced an interesting kernel based method which can not only produce nonlinear discriminant features but also avoid the singularity problem of the \( S_W \). Wu have introduce a new kernel method based on RWLDA technique which called kernel relevance weighted discriminant analysis (KRWDA) [20].

For the Null LDA (NLDA) technique [21], the orientation \( W \) is computed in two stages, in the first stage, the data is projected on the null space of \( S_W \), and in the second stage it finds \( W \) that maximizes \( W^T S_B W \). In orthogonal LDA (OLDA) technique, the orientation matrix \( W \) is obtained by simultaneously diagonalizing scatter matrices, it has shown that OLDA method in equivalent to NLDA under a mild condition. A detailed explanation regarding LDA in given in [22] and an overview regarding SSS based on LDA technique is given in [23]. There are other techniques which can solve the SSS problem and applied in various fields of research. In this paper, we focus on kernel relevance weighted discriminant analysis (KRWDA) algorithm and regularize LDA (RLDA) technique. This techniques overcome SSS problem by using kernel theory and a small perturbation to the within class scatter matrix.

B. Motivation

Recent studies [24] has suggested that a hybrid modified LDA algorithm, which makes use of both linear and nonlinear (kernel theory) dimensionality reduction algorithm could potentially offer the best of the two types of algorithms. Hence, in this paper we combine linear LDA algorithm and nonlinear LDA algorithm in a principled way. For the linear LDA algorithm, we using regularized LDA(RLDA) while the nonlinear LDA algorithm we adopt kernel relevance weighted discriminant analysis (KRWDA).

Moreover, for traditional RLDA the regularized parameter we often computed by cross-validation technique which we always noted as CV-RLDA, since its computation complexity is extremely large and the parameter over a finite range \([a, b]\) so it may not estimate its optimum value. Last, the CV-LDA estimates the regularization parameter for a particular classifier, thus, the estimated value is specific to the classifier and cannot be generalized to other classifier.

III. KRWDA AND RLDA TECHNIQUE

In this Section, at first, we will briefly give a overview of the KRWDA algorithm in Section 3.1. The notation of the RLDA method is given in Section3.2.

A. An Overview of the KRWDA Algorithm

In this paper, we set some notation as follows: a set of \( N \) training images \( \{x_i\}_{i=1}^{N} \) is available, each image is defined as a vector as a vector which the length is \( d(w \times h) \) and \( w \times h \) is the size of the face image[25]. It is also assumed that the training set have \( C \) class and each face image...
belongs to one of the \( C \) classes \( \{X_i\}_{i=1}^C \) and each class have \( n_i \) images so that \( N = \sum_{i=1}^{C} n_i \).

For classical LDA, we aimed to minimize the within class scatter \( S_w \) and to maximize the between class scatter \( S_b \) in the lower feature space[26]:

\[
S_b = \sum_{i=1}^{C} P_i (m_i - \mu)(m_i - \mu)^T \tag{1}
\]

\[
S_w = \sum_{i=1}^{C} P_i \sum_{j=1}^{n_i} (x_{ij} - m_i)(x_{ij} - m_i)^T \tag{2}
\]

where \( m_i \) denotes the mean of the class \( i \) with prior probability \( P_i = \frac{n_i}{N} \) and \( m \) is the total mean; \( x_{ij} \) is the \( j \)-th face sample from class \( i \). Finally, the optimal transformation \( W \) is the solution of the following criterion[27]:

\[
J(W) = \frac{|W^T S_w W|}{|W^T S_b W|} \tag{3}
\]

However, in traditional LDA criterion, it is not optimal with respect to minimizing the classification error rate in the lower dimensional space, as a result, the classification ability will be impaired. To overcome this problem, Loog et al.[28] have introduced a weighting function to the discriminant criterion that the within class scatter and the between class scatter have been defined as follows:

\[
\hat{S}_b = \sum_{i=1}^{C-1} \sum_{j=i+1}^{C} w(d_{ij}) P_i P_j (m_i - \mu)(m_i - \mu)^T \tag{4}
\]

\[
\hat{S}_w = \sum_{i=1}^{C} P_i \sum_{j=1}^{n_i} w(d_{ij}) (x_{ij} - m_i)(x_{ij} - m_i)^T \tag{5}
\]

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

\[
w(d_{ij}) = \left| m_i - m_j \right|^{-2\theta} \tag{6}
\]

where \( r_i \) are the relevance based weights defined as[30]:

\[
r_i = \sum_{j=1}^{C} \frac{1}{w(d_{ij})} \tag{7}
\]

As conclusion, using weighted scatter matrices \( \hat{S}_b \) and \( \hat{S}_w \), the classical discriminant criterion in (3) is weighted and the resulting is referred as relevance weighted LDA (RWLDA) and the criterion can be transformed as:

\[
\hat{J}(W) = \frac{|W^T \hat{S}_w W|}{|W^T \hat{S}_b W|} \tag{8}
\]

Using kernel trick, we can solve the problem of LDA in an implicit feature space \( F \). To make that, we define a weighted between-class scatter and a weighted within-class scatter in feature space \( F \) as follows:

\[
\hat{S}_b = \sum_{i=1}^{C} \sum_{j=1}^{C} w(d_{ij}) P_i P_j (m_i^\theta - m_j^\theta)(m_i^\theta - m_j^\theta)^T \tag{9}
\]

\[
\hat{S}_w = \sum_{i=1}^{C} P_i \sum_{j=1}^{n_i} w(d_{ij}) (\phi(x_{ij}) - m_i^\theta)(\phi(x_{ij}) - m_i^\theta)^T \tag{10}
\]

Then, based on definition of \( \hat{S}_b \) and \( \hat{S}_w \) in equation (9) and (10), we can define a new Fisher criterion in \( F \) as:

\[
\hat{J}(W) = \frac{|W^T \hat{S}_w W|}{|W^T \hat{S}_b W|} \tag{11}
\]

As same as traditional KDA, using kernel trick, Fisher criterion in (11) can be transformed as follows:

\[
\hat{J}(A) = \frac{|A^T \hat{K}_b^\theta A|}{|A^T \hat{K}_w^\theta A|} \tag{12}
\]

where

\[
\hat{K}_b = \sum_{i=1}^{C} \sum_{j=1}^{C} w(d_{ij}) P_i P_j (M_i - M_j)(M_i - M_j)^T \tag{13}
\]

\[
\hat{K}_w = \sum_{i=1}^{C} P_i \sum_{j=1}^{n_i} w(d_{ij}) (\xi_{ij} - M_i)(\xi_{ij} - M_i)^T \tag{14}
\]
\[ M_j = \left( \frac{1}{n_j} \sum_{k=1}^{n_j} k(x_{1j}, x_{d}), \ldots, \frac{1}{n_j} \sum_{k=1}^{n_j} k(x_{nj}, x_{d}), \ldots \right) \]

\[ \xi_k = (k(x_{1k}, x_k), \ldots, k(x_{nk}, x_k)) \]

\[ \Delta = [\alpha_1, \ldots, \alpha_m] \text{ of optimization problem is formed by the} \]

\[ \mathbf{m} \text{ leading eigenvectors of matrix } (K_b^\phi)^{-1} K_b^\phi \]

\[ m \text{ leading eigenvectors of matrix } (K_b^\phi)^{-1} K_b^\phi \]

\[ \text{To make that, we must redefine the within class scatter and between class scatter which given in equation (4) and (5):} \]

\[ S_{\text{bpro}}^\phi = S_b + \alpha \sum_{i=1}^{C-1} \sum_{j=i+1}^{C} w(d_y)^i P_j (m_i - m_j)(m_i - m_j)^T \]

\[ S_{\text{wpro}}^\phi = S_w + \alpha \sum_{i=1}^{C} \sum_{j=1}^{C} w(d_y)^i \sum_{P_j} (x_i - m_i)(x_i - m_i)^T + \alpha I \]

\[ J(W) = \left[ W^T S_{\text{bpro}}^\phi W \right]^{-1} = \left[ W^T (S_w + \alpha I) W \right]^{-1} \]

\[ J(W) = \left[ W^T S_{\text{bpro}}^\phi W \right]^{-1} = \left[ W^T (S_w + \alpha I) W \right]^{-1} \]

\[ J(A) = \frac{A^T K_{\text{bpro}}^\phi A}{A^T K_{\text{wpro}}^\phi A} = \frac{A^T K_{\text{bpro}}^\phi A}{A^T (K_b^\phi + \alpha I) A} \]

\[ J(A) = \frac{A^T K_{\text{bpro}}^\phi A}{A^T K_{\text{wpro}}^\phi A} = \frac{A^T (K_b^\phi + \alpha I) A}{A^T K_{\text{wpro}}^\phi A} \]
In the following text, we will focus on discussing the computation of the regularized parameter $\alpha$ which using a novel deterministic way. First, let us denote two functions:

$$f_{W_{pro}} = A^T \hat{K}_{W_{pro}} A = A^T \hat{K}_{b} A$$

(26)

$$f_{W_{pro}} = A^T \hat{K}_{W_{pro}} A = A^T (\hat{K}_{\phi} + \alpha I)A - b = 0$$

(27)

Where $b$ is larger than zero and is any constant, the Fisher criterion aimed to make the difference between classes larger and difference within classes smaller that equal to find the maximum of $f_{W_{pro}}$ under the constraint. Let us further define a function:

$$F = f_{W_{pro}} - \lambda f_{W_{pro}}$$

(28)

Where $\lambda$ is Lagrange’s multiplier, then by setting its derivative to zero, we can obtain:

$$\frac{\partial F}{\partial A} = \frac{\partial (f_{W_{pro}} - \lambda f_{W_{pro}})}{\partial A} = 2\hat{K}_{b} A - \lambda (2\hat{K}_{\phi} A + 2\alpha A) = 0$$

(29)

Also the equation (29) can be transformed as:

$$\frac{1}{\lambda} (\hat{K}_{\phi} - \hat{K}_{W}) = \alpha A$$

(30)

Substituting value of $\alpha A$ from equation (30) into (27), we can get:

$$f_{W_{pro}} = A^T \hat{K}_{W_{pro}} A = A^T (\hat{K}_{\phi} + \alpha I)A - b$$

$$= A^T \hat{K}_{\phi} A + A^T (\frac{1}{\lambda} (\hat{K}_{\phi} - \hat{K}_{W}))A - b$$

$$= 0$$

(31)

That

$$A^T \hat{K}_{\phi} A = \lambda b$$

(32)

From equation (27), we can also get:

$$A^T \left( \hat{K}_{W} + \alpha I \right) A = b$$

(33)

Dividing equation (32) by (33), we get:

$$\lambda = \frac{A^T \hat{K}_{\phi} A}{A^T \left( \hat{K}_{W} + \alpha I \right) A}$$

(34)

$$\lambda_{\text{max}} = \max \left( \frac{A^T \hat{K}_{\phi} A}{A^T \left( \hat{K}_{W} + \alpha I \right) A} \right) \approx \max \left( \frac{A^T \hat{K}_{\phi} A}{A^T \hat{K}_{\phi} A} \right)$$

$$= \text{largest eigenvalue of} \ (\hat{K}_{W})^{-1} \hat{K}_{\phi}$$

(35)

In order to optimal the proposed Fisher discriminant criterion, so we must set $\lambda$ equal to maximum of $J(A)$. We must have eigenvector $A$ to correspond to the maximum eigenvalue of $(\hat{K}_{W})^{-1} \hat{K}_{\phi}$, Thereby, the evaluation of $\alpha$ can be carried out from equation (30) by doing EVD of

$$\left( \frac{1}{\lambda} \hat{K}_{\phi} - \hat{K}_{W} \right)$$

where $\lambda = \lambda_{\text{max}}$. After evaluation the optimal parameter $\alpha$, orientation vector $A$ can be obtained by performing the EVD of

$$\left( \hat{K}_{W} + \alpha I \right)^{-1} \hat{K}_{\phi}$$

(36)

V. RESULTS AND DISCUSSION

In this Section, we present extensive experimental results on various public face databases to evaluate the effectiveness of the proposed KRRWDWA method. At first, we introduce several related competing algorithms and some experiment settings in Section 5.1. In Section 5.2, we will give the determination of the optimal parameters in KRRWDWA algorithm. In Section 5.3, we demonstrate the robustness of the proposed technique against illumination variations on the Multi-PIE face database and FRGC face database. In Section 5.4, we will evaluate the proposed new algorithm against pose and facial expression variations on the FERET and LFW face databases. In Section 5.5, we will discuss the Single training Sample Per Person(SSPP) problem on the above mentioned face databases. In Section 5.6, the computation complexity will be given.

A. The Competing Algorithms and Experimental Settings

In order to evaluate the performance of the proposed novel technique, we select several popular face recognition algorithms for comparison, including the baseline Eigen-face method, Fisher-face, OTF based CFA[32], Sparse Representation based Classification(SRC)[33], also some state-of-the-art local based FE methods including Block-FLD(B-FLD)[34], Cascade LDA(C-LDA)[35], Block based Bag of Words(BBOW)[36]. Of course, the RLDA technique and KRWDA algorithm also compared. For the proposed KRRWDWA algorithm, we evaluate the KRRWDWA(CV) method(using cross validation method for computing the regularized parameter $\alpha$) and KRRWDWA(DE) method(using the proposed deterministic method for computing the regularized parameter $\alpha$).

Each image in the face databases is normalized in order to extract the exact facial region that only contains the face. At first, the centers of the eyes are manually annotated,
secondly, rotation and scaling transformations align the centers of the eyes to predefined locations and fixed interocular distances, finally, a face image is cropped and resized to the size of 64x64 pixels. Histogram equalization is then applied to all face images for photometric normalization.

After feature extraction for both training set and test set, we employ the nearest neighbor classifier for final classification. The cosine similarity measure is used for all compared algorithms. For all face databases, a random subset that with \( t \) images per subset is taken from each databases to form the training set, so the rest of the database is used as the test set. For each \( t \), the experiments with randomly chosen subsets are performed 20 times. We report the average recognition rates as well as standard deviations over the randomly chosen tests as the final results.

In this paper, we focus on the SSS problem, which is one of the most challenging issues in face recognition. In order to evaluate the effectiveness of different feature extraction methods to solve the SSS problem, the value of \( t \) is set to 2 ~ 5 for all face databases.

### B. Determining the Related Parameters

In KRRWDA technique, two important parameters named regularized parameter \( \alpha \) and kernel functions will be determined at first which have an influence on the final face recognition accuracy. The regularized parameter should be carefully set that the purpose of the regularization process is to reduce the high variance related to the estimation of the covariance matrix, which is caused by the SSS problem. On the other hand, the kernel function is crucial for the kernel trick, in this part, we will compare the performance between Polynomial kernel function, Gaussian RBF kernel function and Fractional Polynomial kernel function.

At first, we determine the regularized parameter \( \alpha \) which using in the KRRWDA technique, Table 1 shown the value of \( \alpha \) using in different face databases.

<table>
<thead>
<tr>
<th>Database</th>
<th>AR</th>
<th>Multi-PIE</th>
<th>FRGC</th>
<th>FERET</th>
<th>LFW</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \alpha )</td>
<td>10.5715.6</td>
<td>6.54 x 10^9</td>
<td>1.11 x 10^11</td>
<td>1.24 x 10^9</td>
<td>2.98 x 10^10</td>
</tr>
</tbody>
</table>

Second, for kernel trick, different kernel functions will obtain different performance, so in this part, we compare the performance between above mentioned kernel functions in AR face database using KRRWDA (CV) algorithms. The AR face database contains over 4000 face images of 126 subjects (70 men and 56 women). The AR database characterizes the divergence from ideal conditions by incorporating various facial expressions, illuminations changes, and occlusion modes. A subset that contains 120 subjects which each subject has 14 images with only facial expression and illumination changes is used in this expressions, see Figure 1 for some example.

In this experiment, we only focus on the performance influence by kernel functions, so the computation of the parameter \( \alpha \) using the cross validation technique (CV) in this experiment. The definition of the three general used kernel function are given in Table 2. In Table 3 to Table 5 we report the results of the performance.
TABLE 2. KERNEL FUNCTIONS

<table>
<thead>
<tr>
<th>Polynomial Kernel Function</th>
</tr>
</thead>
<tbody>
<tr>
<td>$K(x,y) = (1 + xy^d)$</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Gaussian RBF kernel function</th>
</tr>
</thead>
<tbody>
<tr>
<td>$K(x,y) = \exp\left(-\frac{</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Fractional Polynomial Kernel Function</th>
</tr>
</thead>
<tbody>
<tr>
<td>$K(x,y) = (1 + xy^d)$</td>
</tr>
</tbody>
</table>

TABLE 3. THE PERFORMANCE OF KRRWDA (CV) ALGORITHM UNDER POLYNOMIAL KERNEL FUNCTION

<table>
<thead>
<tr>
<th>Tab 3. Polynomial Kernel Function</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\sigma^2$</td>
</tr>
<tr>
<td>------------------</td>
</tr>
<tr>
<td>2</td>
</tr>
<tr>
<td>3</td>
</tr>
<tr>
<td>4</td>
</tr>
<tr>
<td>5</td>
</tr>
</tbody>
</table>

TABLE 4. THE PERFORMANCE OF KRRWDA(CV) ALGORITHM UNDER GAUSSIAN RBF KERNEL FUNCTION

<table>
<thead>
<tr>
<th>Gaussian RBF Kernel Function</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\sigma^2$</td>
</tr>
<tr>
<td>------------------</td>
</tr>
<tr>
<td>2</td>
</tr>
<tr>
<td>3</td>
</tr>
<tr>
<td>4</td>
</tr>
<tr>
<td>5</td>
</tr>
</tbody>
</table>

TABLE 5. THE PERFORMANCE OF KRRWDA (CV) ALGORITHM UNDER FRACTIONAL POLYNOMIAL KERNEL FUNCTION

<table>
<thead>
<tr>
<th>Fractional Polynomial Kernel Function</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\sigma^2$</td>
</tr>
<tr>
<td>------------------</td>
</tr>
<tr>
<td>2</td>
</tr>
<tr>
<td>3</td>
</tr>
</tbody>
</table>

From the recognition results shown in Table 3 to Table 5, we find the best recognition result is obtained using Gaussian RBF kernel function when $\sigma^2 = 10^5$. So in the rest of the paper, we use Gaussian RBF kernel function which $\sigma^2 = 105$.

C. Robustness to Illumination Variations

One of the most fundamental challenges is face recognition is significant facial appearance variations due to illumination changes. In this section, we evaluate the performance of the proposed algorithm against illumination variations on the Multi-PIE face database and the FRGC face database.

The Multi-PIE database contains more than 750,000 images of 337 subjects which captured in 4 sessions with variations in pose, illumination and facial expression. Figure 2 shows the face images of one subject of this face database. The FRGC face database consists of controlled images, uncontrolled images and three-dimensional face images for each subject. In this experiment, we select a subset containing 6000 images of 300 subjects. Figure 3 shows the face images of one subject on the FRGC database which used in this paper.

Table 6 and Table 7 show the average recognition performance obtained by the different algorithms on the Multi-PIE and FRGC face databases respectively. From these two tables, we can see that the proposed KRRWDA (DE) algorithm consistently achieves better recognition accuracies than the other competing methods. Comparing with KRRWDA(CV) algorithm, the proposed KRRWDA(DE) improve the recognition performance by about 4-5%. SRC obtains better results than B-FLD which shows that SRC is robust in dealing with illumination variations. B-FLD constructs multiple training patterns from a single images but it does die consider the relationships among different face sub-regions.

Fig. 2 The face images of Multi-PIE database
D. Robustness to Pose and Facial Expression Variations

In this section, we evaluate the influence of pose and facial expression variations on the performance of the proposed algorithms by using FERET face database and the LFE face database.

The FERET face database is a standard face database for evaluating the performance of face recognition algorithms. A subset of the FERET database which includes 1400 images of 200 subjects is used, as shown in Figure 4. This subset involves challenges, such as variations in facial expression and pose. Besides, we also perform an experiment on a more realistic face database captured in

---

**TABLE 6. THE PERFORMANCE OBTAINED BY THE DIFFERENT ALGORITHMS ON THE MULTI-PIE DATABASE**

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>( t = 2 )</th>
<th>( t = 3 )</th>
<th>( t = 4 )</th>
<th>( t = 5 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Eigen-face</td>
<td>72.24 ± 1.5</td>
<td>78.54 ± 1.7</td>
<td>82.13 ± 1.8</td>
<td>85.41 ± 1.7</td>
</tr>
<tr>
<td>Fisher-face</td>
<td>76.79 ± 1.1</td>
<td>86.63 ± 1.3</td>
<td>88.95 ± 1.4</td>
<td>92.07 ± 1.5</td>
</tr>
<tr>
<td>CFA-OTF</td>
<td>83.15 ± 0.8</td>
<td>88.05 ± 0.8</td>
<td>90.17 ± 0.6</td>
<td>93.10 ± 0.5</td>
</tr>
<tr>
<td>SRC</td>
<td>82.24 ± 1.2</td>
<td>86.59 ± 1.3</td>
<td>89.98 ± 1.2</td>
<td>93.15 ± 0.9</td>
</tr>
<tr>
<td>B-FLD</td>
<td>81.17 ± 1.0</td>
<td>82.84 ± 1.4</td>
<td>88.77 ± 1.1</td>
<td>89.73 ± 1.0</td>
</tr>
<tr>
<td>C-LDA</td>
<td>83.25 ± 0.9</td>
<td>85.77 ± 0.8</td>
<td>89.95 ± 0.5</td>
<td>90.07 ± 0.8</td>
</tr>
<tr>
<td>BBOW</td>
<td>83.58 ± 0.8</td>
<td>87.25 ± 0.9</td>
<td>91.27 ± 0.9</td>
<td>92.66 ± 0.7</td>
</tr>
<tr>
<td>KRRWDA (CV)</td>
<td>82.51 ± 1.1</td>
<td>86.24 ± 1.3</td>
<td>90.05 ± 0.8</td>
<td>91.17 ± 0.6</td>
</tr>
<tr>
<td>KRRWDA (DE)</td>
<td>86.87 ± 0.6</td>
<td>92.07 ± 0.7</td>
<td>94.15 ± 0.5</td>
<td>96.65 ± 0.4</td>
</tr>
</tbody>
</table>

**TABLE 7. THE PERFORMANCE OBTAINED BY THE DIFFERENT ALGORITHMS ON THE FRGC DATABASE**

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>( t = 2 )</th>
<th>( t = 3 )</th>
<th>( t = 4 )</th>
<th>( t = 5 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Eigen-face</td>
<td>45.38 ± 1.3</td>
<td>53.10 ± 1.2</td>
<td>64.35 ± 1.1</td>
<td>70.26 ± 1.5</td>
</tr>
<tr>
<td>Fisher-face</td>
<td>48.17 ± 1.1</td>
<td>52.43 ± 1.3</td>
<td>66.78 ± 1.5</td>
<td>69.06 ± 1.7</td>
</tr>
<tr>
<td>CFA-OTF</td>
<td>54.35 ± 0.8</td>
<td>62.17 ± 0.8</td>
<td>65.99 ± 0.9</td>
<td>73.91 ± 1.0</td>
</tr>
<tr>
<td>SRC</td>
<td>57.72 ± 1.1</td>
<td>65.14 ± 1.2</td>
<td>72.28 ± 1.2</td>
<td>81.18 ± 0.9</td>
</tr>
<tr>
<td>B-FLD</td>
<td>53.13 ± 1.0</td>
<td>62.28 ± 1.3</td>
<td>66.77 ± 0.9</td>
<td>70.20 ± 1.0</td>
</tr>
<tr>
<td>C-LDA</td>
<td>55.72 ± 1.1</td>
<td>66.11 ± 0.8</td>
<td>72.24 ± 1.1</td>
<td>76.89 ± 1.2</td>
</tr>
<tr>
<td>BBOW</td>
<td>58.57 ± 1.4</td>
<td>71.90 ± 1.2</td>
<td>73.10 ± 0.7</td>
<td>78.43 ± 0.7</td>
</tr>
<tr>
<td>KRRWDA (CV)</td>
<td>59.86 ± 1.2</td>
<td>70.66 ± 1.3</td>
<td>78.31 ± 0.8</td>
<td>85.53 ± 0.6</td>
</tr>
<tr>
<td>KRRWDA (DE)</td>
<td>63.99 ± 0.8</td>
<td>75.24 ± 0.9</td>
<td>82.21 ± 0.5</td>
<td>88.58 ± 0.6</td>
</tr>
</tbody>
</table>
unconstrained environments which named LFW face database that usually used to evaluate face recognition algorithms in real scenarios. It contains the images of 5749 different individuals collected from the web. LFW-a is a version of LFW after face alignment and a subset with 150 subjects is chosen. Figure 5 shows the samples of one subject on the LFW database used in this experiment. Table 8 and Table 9 show the experiments results on the FERET face database and LFW database respectively. 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 those obtained by KRRWDA(DE), this is due to the face that the usage of the whole face region makes CFA-OTF sensitive to pose variation. In contrast, KRRWDA(DE) alleviate this problem by using kernel trick and regularized LDA technique. Furthermore, BBOW obtains lower recognition rate than KRRWDA(DE) in the LFW face database, which indicates that BBOW can not effectively capture the intrinsic discriminative information when the training set contains variations in pose and facial expression.

Compared with the recognition performance obtained on other face databases. The proposed KRRWDA(DE) algorithm obtains lower recognition accuracies on the LFW database. There are two main reasons: (1) some face images contains the surrounding background which decreases the discriminability of features extracted by our proposed algorithm.(2) the mismatching of face images between the training samples and test samples can occur when dealing with large pose variations.

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>$t = 2$</th>
<th>$t = 3$</th>
<th>$t = 4$</th>
<th>$t = 5$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Eigen-face</td>
<td>53.27 ± 3.0</td>
<td>60.12 ± 2.9</td>
<td>65.50 ± 2.8</td>
<td>70.22 ± 2.1</td>
</tr>
<tr>
<td>Fisher-face</td>
<td>66.63 ± 1.8</td>
<td>67.79 ± 1.7</td>
<td>76.23 ± 1.4</td>
<td>77.54 ± 1.3</td>
</tr>
<tr>
<td>CFA-OTF</td>
<td>58.96 ± 1.7</td>
<td>65.53 ± 1.5</td>
<td>74.18 ± 1.1</td>
<td>78.97 ± 1.4</td>
</tr>
<tr>
<td>SRC</td>
<td>66.21 ± 2.1</td>
<td>67.14 ± 2.3</td>
<td>71.16 ± 1.2</td>
<td>73.36 ± 1.9</td>
</tr>
<tr>
<td>B-FLD</td>
<td>67.57 ± 1.8</td>
<td>69.95 ± 1.4</td>
<td>73.28 ± 1.7</td>
<td>80.95 ± 1.6</td>
</tr>
<tr>
<td>C-LDA</td>
<td>68.83 ± 0.9</td>
<td>70.17 ± 2.3</td>
<td>75.36 ± 2.4</td>
<td>83.27 ± 2.3</td>
</tr>
<tr>
<td>BBOW</td>
<td>74.15 ± 0.8</td>
<td>77.42 ± 1.2</td>
<td>86.00 ± 1.3</td>
<td>92.34 ± 1.5</td>
</tr>
<tr>
<td>KRRWDA (CV)</td>
<td>75.10 ± 1.9</td>
<td>81.14 ± 1.8</td>
<td>90.25 ± 1.1</td>
<td>92.11 ± 1.4</td>
</tr>
<tr>
<td>KRRWDA (DE)</td>
<td>80.60 ± 1.4</td>
<td>84.72 ± 1.3</td>
<td>94.26 ± 1.2</td>
<td>95.85 ± 1.1</td>
</tr>
</tbody>
</table>

Fig.4 The face images of FERET database.

Fig.5 The face images of LFW database.

TABLE 8. THE PERFORMANCE OBTAINED BY THE DIFFERENT ALGORITHMS ON THE FERET DATABASE
TABLE 9. THE PERFORMANCE OBTAINED BY THE DIFFERENT ALGORITHMS ON THE LFW DATABASE

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>$t = 2$</th>
<th>$t = 3$</th>
<th>$t = 4$</th>
<th>$t = 5$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Eigen-face</td>
<td>24.15 ± 3.2</td>
<td>28.10 ± 3.2</td>
<td>32.23 ± 3.5</td>
<td>37.00 ± 3.7</td>
</tr>
<tr>
<td>Fisher-face</td>
<td>27.89 ± 2.8</td>
<td>33.42 ± 1.3</td>
<td>38.42 ± 2.4</td>
<td>44.25 ± 2.5</td>
</tr>
<tr>
<td>CFA-OTF</td>
<td>25.27 ± 3.5</td>
<td>30.17 ± 0.8</td>
<td>32.17 ± 1.6</td>
<td>35.24 ± 3.5</td>
</tr>
<tr>
<td>SRC</td>
<td>30.25 ± 2.5</td>
<td>35.24 ± 2.3</td>
<td>39.98 ± 1.2</td>
<td>45.15 ± 1.9</td>
</tr>
<tr>
<td>B-FLD</td>
<td>32.53 ± 2.3</td>
<td>36.78 ± 2.4</td>
<td>40.12 ± 1.9</td>
<td>45.24 ± 1.5</td>
</tr>
<tr>
<td>C-LDA</td>
<td>31.10 ± 2.2</td>
<td>35.41 ± 2.1</td>
<td>38.82 ± 1.5</td>
<td>44.99 ± 1.8</td>
</tr>
<tr>
<td>BBOW</td>
<td>31.27 ± 1.9</td>
<td>33.41 ± 1.9</td>
<td>41.27 ± 1.9</td>
<td>48.21 ± 1.7</td>
</tr>
<tr>
<td>KRRWDA (CV)</td>
<td>32.10 ± 1.1</td>
<td>42.17 ± 2.1</td>
<td>47.05 ± 1.8</td>
<td>50.72 ± 1.3</td>
</tr>
<tr>
<td>KRRWDA (DE)</td>
<td>38.20 ± 1.6</td>
<td>43.10 ± 1.7</td>
<td>48.58 ± 1.4</td>
<td>52.20 ± 1.4</td>
</tr>
</tbody>
</table>

E. Single Training Sample Face Recognition

In this section, we test the performance of the competing algorithms on all above mentioned face databases with a Single training Sample Per Person (SSPP, which is an extreme case of the SSS problem that severely challenges conventional face recognition algorithms). In such case, the traditional supervised learning techniques, such as LDA may not be applicable since the intra subject information cannot be obtained from one training sample that one possible solution is to use a generic training set. Note that since Fisher-face cannot deal with the SSPP problem, its performance not reported in this section. Table 10 shows the average recognition accuracies obtained by the competing algorithms in dealing with the SSPP problem. Among the competing algorithms, the proposed KRRWDA (DE) technique obtains comparable results on most face databases. Specially, KRRWDA (DE) outperforms most of the compared local based algorithms, such as B-FLD and BBOW. Furthermore, compared with the SRC method, KRRWDA (DE) still achieves better performance which clearly demonstrates the desirable classification ability of the proposed algorithm.

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>AR</th>
<th>Multi-PIE</th>
<th>FRGC</th>
<th>FERET</th>
<th>LFW</th>
</tr>
</thead>
<tbody>
<tr>
<td>Eigenface</td>
<td>35.77 ± 3.5</td>
<td>50.12 ± 3.5</td>
<td>22.42 ± 4.1</td>
<td>33.70 ± 3.8</td>
<td>11.13 ± 3.8</td>
</tr>
<tr>
<td>CFA-OTF</td>
<td>38.54 ± 3.4</td>
<td>55.53 ± 2.5</td>
<td>40.18 ± 3.8</td>
<td>31.00 ± 3.4</td>
<td>13.21 ± 2.8</td>
</tr>
<tr>
<td>SRC</td>
<td>45.27 ± 3.2</td>
<td>57.89 ± 1.8</td>
<td>38.28 ± 3.3</td>
<td>43.32 ± 3.2</td>
<td>15.26 ± 2.7</td>
</tr>
<tr>
<td>B-FLD</td>
<td>48.81 ± 2.8</td>
<td>56.95 ± 1.4</td>
<td>45.17 ± 2.7</td>
<td>50.47 ± 2.8</td>
<td>18.78 ± 2.5</td>
</tr>
<tr>
<td>BBOW</td>
<td>64.21 ± 2.5</td>
<td>55.98 ± 1.8</td>
<td>46.31 ± 2.7</td>
<td>60.52 ± 2.5</td>
<td>17.37 ± 3.2</td>
</tr>
<tr>
<td>KRRWDA (CV)</td>
<td>65.40 ± 2.3</td>
<td>61.11 ± 1.8</td>
<td>48.94 ± 3.1</td>
<td>64.25 ± 2.4</td>
<td>21.15 ± 2.9</td>
</tr>
<tr>
<td>KRRWDA (DE)</td>
<td>66.13 ± 2.4</td>
<td>62.81 ± 1.3</td>
<td>52.74 ± 2.8</td>
<td>66.60 ± 2.1</td>
<td>22.17 ± 2.8</td>
</tr>
</tbody>
</table>

Note that the results obtained by some competing algorithms in our experiments are different from the reported results. This is because the experimental settings in our paper and the original paper are different. In our face recognition experiment in this section, we only use a single sample per person for training in our paper.

F. Computational Complexity of the Proposed Algorithm

In this section, we compare the computational time of the proposed KRRWDA (DE) algorithm with that of some representative feature extraction algorithms including Eigen-face, Fisher face, CFA-OTF, KRRWDA (CV). All the
computational time is reported on a workstation with 2 inter
Xeon E5620 CPUs on the MATLAB platform. Table 11
shows the computational time spent on the training set and
test stages by these algorithms on the CAS-PEAL face
database.

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>Training Time</th>
<th>Recognition Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>Eigen-face</td>
<td>51.41</td>
<td>70.61</td>
</tr>
<tr>
<td>Fisher-face</td>
<td>83.74</td>
<td>20.62</td>
</tr>
<tr>
<td>CFA-OTF</td>
<td>522.74</td>
<td>32.88</td>
</tr>
<tr>
<td>KRRWDA(CV)</td>
<td>3134.47</td>
<td>123.80</td>
</tr>
<tr>
<td>KRRWDA(DE)</td>
<td>2202.78</td>
<td>82.64</td>
</tr>
</tbody>
</table>

As shown in Table 11, the computational time of the
proposed KRRWDA (DE) used for training is higher than
that of the other algorithms except KRRWDA (CV)
technique. As the training stage is usually performed offline,
the computational complexity of the proposed algorithm will
not constrain its application to real world tasks.

VI. CONCLUSION

In this paper, we have presented an effective feature
extracted method called kernel regularized relevance
weighted discriminant analysis (KRRWDA) and applied it to
the task of face recognition in order to solve the SSS
problem. In order to solve the problem of the computation
efficiency of the regularized parameter $\alpha$, a deterministic
way computation method is given in this paper. Aimed to test
the performance of this novel feature extraction technique,
we have evaluated KRRWDA (DE) method under different
conditions, including variations in illumination, facial
expression, and pose, as well as dealing with SSPP
problem. The experimental results have shown that
KRRWDA (DE) outperforms most state-of-the-art feature
extraction algorithms on public face databases for solving
the SSS problem.

We can intuitively explain the KRRWDA (DE)
performing than other algorithms for dealing with SSS
problem as follows. In the KRRWDA (DE) technique, we
are maximizing the modified Fisher criterion by the ratio of
between class scatter and within class scatter. To get the $\alpha$
parameter, we are maximizing the difference between
the between class scatter and within class scatter. Thus, we are
combining these two different philosophies of LDA
mechanism in our KRRWDA (DE) technique and this is
helping us in getting better recognition performance.

As mentioned in our experiments, the method proposed
in this paper cannot handle face recognition with very large
pose variations well, recent study has demonstrate that the
Deep Learning method (DP) can be helpful to improve the
recognition performance. Hence, how to design a effective
face recognition based on DP and space method under large
pose variations is an interesting direction of our future work.
In addition, we also interesting to excavate the effective
computation method of the regularized parameter $\alpha$.

ACKNOWLEDGMENT

The author would like to thank the anonymous reviewers
for their constructive comments. The work was supported by
the National Natural Science Foundation of China under
Grant 61263031 and 61563030, by the by the National
Natural Science Foundation of Hunan Province of China
under Grant 14JJ6041and 16JJ6025, by the Key Project of
Education Department of Hunan Province of China under
Grant 15A044.

COMPETING Interests

The authors declare that they have no competing
interests.

REFERENCES

recognition using random forest based fusion of a-trous wavelet
transform coefficients from thermal and visible images[J].AEU -
AND SOME RECENT DEVELOPMENTS[R].Biometrics
[3] Heikila M, Pietikainen M. Description of interest regions with local
binary patterns[J].Pattern Recognition,2009,42(3):425-436
regularized linear discriminant analysis[J].Neurocomputing,
[6] Ramirez Rivera A , Rojas Castillo J , OksamChae etc. Local
directional number pattern for face analysis : Face and Expression
for robust face recognition". Pattern Recognition, November 2014.
[10] ZHONG Fujin , ZHANG Jiashu. Face recognition with improved
local directional patterns [J]. Journal of Optoelectronics , 2012 ,
23(8):1570-1575.

DOI 10.5013/IJSSST.a.17.35.32    32.11    ISSN: 1473-804x online, 1473-8031 print
DI WU: A ROBUST FACE RECOGNITION ALGORITHM BASED ON KERNEL REGULARIZED RELEVANCE …


