Kernel Weighted Scatter-Difference-Based Two Dimensional Discriminant Analyses for Face Recognition

Hythem Ahmed, Mohamed Jedra and Nouredine Zahid

Laboratory of Conception and Systems, Faculty of Sciences, Mohamed V University, Agdal, Avenue Ibn Batouta, B.P.1014, Rabat, Morocco
Hythem0101@hotmail.com, jedra@fsr.ac.ma, zahid@fsr.ac.ma, http://www.fsr.ac.ma

Abstract

In this paper, a novel image projection technique for face recognition application is proposed which is based on linear discriminant analysis (LDA) coined Kernel Weighted Scatter Two Dimensional Discriminant Analysis (KWS2DDA). The projection is performed through 2-direction which simultaneously works in row and column directions to solve the small sample size problem. This nonlinear dimensionality reduction algorithm has several interesting characteristics. It's overcomes the singularity problem, while achieving efficiency. In order to improve the performance of the proposed algorithm, we introduce Gaussian RBF kernel functions. We have performed multiple face recognition experiments to compare KWS2DDA with other dimensionality reduction methods showing that KWS2DDA consistently gives the best result than the other method.

Keywords: *Feature extraction, face recognition, LDA, PCA, 2DLDA, WS2DDA, KWS2DDA*

1. Introduction

Linear Discriminant Analysis [1, 2, 3, 4, 5] is a well-known method which projects the data onto a lower-dimensional vector space such that the ratio of the between-class distance to the within-class distance is maximized, thus achieving maximum discrimination. The optimal projection can be readily computed by applying the eigen-objective functions. All scatter matrices in question can be singular since the data is from a very high-dimensional space, and in general, the dimension exceeds the number of data points. This is known as the under sampled or singularity problem [6].

In recent years, many approaches have been brought to bear on such high-dimensional, under sampled problems, including pseudo-inverse LDA, PCA+LDA, and regularized LDA. More details can be found in [6]. Among these LDA extensions, PCA+LDA, has received a lot of attention, especially for face recognition [3]. In this two-stage algorithm, an intermediate dimension reduction stage using PCA is applied before LDA. The common aspect of previous LDA extension is the computation of Eigen-decomposition of certain large matrices, which not only degrades the efficiency but also makes it hard to scale them to large datasets.

The objective of LDA is to find the optimal projection so that the ratio of the determinants of the between-class and within-class scatter matrix of the projected samples reaches its maximum. However, concatenating 2D matrices into a 1D vector leads to a very high-dimensional image vector, where it is difficult to evaluate the scatter matrices accurately due to its large size and the relatively small number of training samples. Furthermore, the within-class scatter matrix is always singular, making the direct implementation of the LDA algorithm an intractable task. To overcome these problems, a new technique called 2-

dimensional LDA (2DLDA) was recently proposed [7]. This method directly computes the eigenvectors of the scatter matrices without matrix to vector conversion.

The scatter matrices in 2DLDA are quite small compared to the scatter matrices in LDA. The size of the 2DLDA matrix is proportional to the width of the image. 2DLDA evaluates the scatter matrices more accurately and computes the corresponding eigenvectors more efficiently than LDA or PCA. However, the main drawback of 2DLDA is that it needs more coefficients for image representation that the conventional PCA and LDA-based schemes. Tang et. al., [8] have introduced a weighting scheme to estimate the within-class scatter matrix using a so called relevance weights. Motivated by kernel machines, nonlinear extensions have been shown to improve linear methods by applying the ‘‘Kernel trick’’ [9]. Also kernel PCA (KPCA) [10, 11] combines the kernel trick with PCA to find nonlinear principal components in the feature space. However, as the same as PCA, KPCA captures the overall variance of all patterns which are inadequate for discriminating purposes. In the same spirit as the kernel extensions of PCA, kernel extension of LDA, called kernel discriminant analysis (KDA) [12–14], has also been developed and found to be more effective than PCA, KPCA and LDA for many classification applications due to its ability in extracting nonlinear features that exhibit high class separability. Chougdali, et all presents a kernel weighted scatter difference discriminant analysis (KWSDA) method for face recognition [15].

2. Two Dimensional Linear Discriminant Analysis (2DLDA)

Suppose there are c known pattern classes $w_1, w_2, \ldots, w_c$ and N training samples $X = [x_j^i]$, i =1 , 2 , … ,I_j , j= 1, 2 , .. , c with (m × n) dimension . I_j are the number of training samples of class j and satisfies $\sum_{i=1}^{c} I_j = N$. The following steps summarize the process of 2DLDA:

1) Calculate the average matrix X of the N training images using:

$$\bar{x} = \frac{1}{N}\sum_{j=1}^{c}\sum_{i=1}^{I_j} x_j^i \quad (1)$$

2) Compute the mean $\overline{\mathbf{A_i}}$ of $\mathbf{i^{th}}$ class by:

$$\overline{x_i} = \left(\frac{1}{I_j}\right)\sum_{i=1}^{I_j} x_j^i \quad (2)$$

where i=1,2,…,I_j

3) Calculate the image between-class scatter matrix by:

$$S_b = \frac{1}{N}\sum_{j=1}^{c}(\overline{x}_j - \overline{x})\left(\overline{x}_j - \overline{x}\right)^T \quad (3)$$

4) Calculate the image within-class scatter matrix by:

$$S_w = \frac{1}{N}\sum_{j=1}^{c}\sum_{i=1}^{I_j}[x_j^i - \overline{x_j}][x_j^i - \overline{x_j}]^T \quad (4)$$

5) Find the optimal projection W so that the total scatter of the projected samples of the training images is maximized. The objective function of 2DLDA is defined by:

$$J(W) = \mathrm{argmax}_W \frac{W^T S_b W}{W^T S_w W} \quad (5)$$

It can be proven that the eigenvector corresponding to the maximum eigenvalue of $(S_w)^{-1}$ (S_b) is the optimal projection vectors which maximizes J(W). Generally, as it is not enough to have only one optimal projection vector, we usually look for d projection axes, say $(w_1, w_2, \ldots, w_d)$, which are the eigenvectors corresponding to the first d largest eigenvalues of $(S_w)^{-1}(S_b)$. In 2DLDA, once these projection vectors are computed, each training image x_j^i is then projected onto W to obtain the feature matrix Y_j^i with dimensions (m×d) of the training image x_j^i. So, during training, for each training image x_j^i a corresponding feature matrix is constructed and sorted for matching at the time of recognition.

6) For test image we project the test matrix onto the eigenvectors matrix to fined the new matrix of dimension (m × k):

$$B_j = A_T V \quad (6)$$

7) Calculate the face distance between two arbitrary feature matrix $\mathbf{B_i}$ and $\mathbf{B_j}$ defined by:

$$d(B_j, B_i) = \sum_{n=1}^{k} \left\| Y_n^j - Y_n^i \right\|_2 \quad (7)$$

If $d(B, B_i)$ = min $d(B_j, B_i)$ and $B_j \in \omega_k$, where ω_k identify the class and B is a test sample, then the resulting decision is $B \in \omega_k$.

3. Weighted Scatter Difference Discriminant Analysis

The maximum scatter difference (MSD) discriminant criterion [8] was a recently presented. Because the MSD utilizes the generalized scatter difference rather than the generalized Rayleigh quotient as a class separability measure, it avoids the singularity problem when addressing the small-sample size problem that troubles the Fisher discriminant criterion. Furthermore, studies have demonstrated that MSD classifiers based on this discriminant criterion have been quite effective on face-recognition tasks. We introduce weighted scatter matrices and thus define a weighted scatter-difference-based discriminant analysis criterion as follows:

$$J_M(W) = \mathrm{argmax}_w \mathrm{trace}(W^T(\hat{S}_b - M\hat{S}_w)W) \quad (8)$$

where,

$$\hat{S}_b = \sum_{i=1}^{c-1} \sum_{j=i+1}^{c} w(d_{ij}) P_i P_j (m_i - m_j)(m_i - m_j)^T \quad (9)$$

$$\hat{S}_w = \sum_{i=1}^{c} P_i r_i \sum_{j=1}^{n_i} (x_{ij} - m_i)(x_{ij} - m_i)^T \quad (10)$$

Equation (10) has extended the concept of weighting to estimate a within-class scatter matrix. Thus, by introducing a so-called relevance weights, a weighted within-class scatter matrix $\hat{S}_w$ is defined to replace a conventional within-class scatter matrix.

Where P_i and P_j are the class priors, $w(d_{ij})$ is the Euclidean distance between the means of class i and class j. The weighting function $w(d_{ij})$ is generally a monotonically decreasing function:

$$w(d_{ij}) = \left\|m_i - m_j\right\|^{-2} \tag{11}$$

and r_i's $(0 < r_i \leq 1, \forall\, i)$ are the relevance based weights defined by

$$r_i = \sum_{j \neq i} w \tag{12}$$

Using the weighted scatter matrices $\hat{S}_b$ and $\hat{S}_w$ the criterion in equation (5) is weighted and the resulting algorithm is referred to as Weighted Scatter 2DLDA (WS-2DLDA) where:

$$M = \frac{\text{trace}(\hat{S}_b)}{\text{trace}(\hat{S}_w)} \tag{13}$$

4. Analysis Method

The key idea of kernel discriminant analysis is to solve the problem of LDA in an implicit feature space F, which is constructed by the kernel trick. Consider there is a feature mapping which maps the input data into a higher dimensional inner product space F, i.e., $\emptyset: R^N \rightarrow F$. Consequently, LDA can be performed in F. It is equivalent to maximizing the following criterion:

$$J(W) = \frac{\left|W^T S_b^{\emptyset} W\right|}{\left|W^T S_w^{\emptyset} W\right|} \tag{14}$$

The between-class scatter matrix $S_b^{\emptyset}$ and within-class scatter matrix $S_w^{\emptyset}$ in the feature space F can be expressed as:

$$S_b^{\emptyset} = \sum_{i=1}^{c} \frac{n_i}{N} (m_i^{\emptyset} - m^{\emptyset})(m_i^{\emptyset} - m^{\emptyset})^T \tag{15}$$

$$S_w^{\emptyset} = \frac{1}{N} \sum_{i=1}^{c} \sum_{j=1}^{n_i} (\emptyset(X_{ij}) - m_j^{\emptyset})(\emptyset(X_{ij}) - m_j^{\emptyset})^T \tag{16}$$

where $m_i^{\emptyset} = \frac{1}{n_i} \sum_{j=1}^{n_i} \emptyset(x_{ij})$ denotes the mean of class i in F and $m^{\emptyset} = \frac{1}{N} \sum_{i=1}^{c} \sum_{j=1}^{n_i} \emptyset(x_{ij})$ the total mean. Referring to Eq. (14), any column of the solution W must lie in the span of all the samples in F. So, there exist coefficients α_{ij} such that:

$$w = \sum_{i=1}^{c} \sum_{j=1}^{n_i} \alpha_{ij} \emptyset(x_{ij}) \tag{17}$$

where w represents any one column of the projection matrix W. In other words, by combining Eq. (17) and the definition of $m_i^{\emptyset}$ we can project each mean of class i onto an axis of F as follows:

$$W^T m_i^{\emptyset} = \frac{1}{n_j} \sum_{i=1}^{c} \sum_{j=1}^{n_i} \sum_{k=1}^{n_i} \alpha_{ij} k\left(x_{ij}, x_{jk}\right) = \alpha^T M_j \tag{18}$$

where $\alpha = (\alpha_{11}, \alpha_{1n_i}, \ldots, \alpha_{ij}, \ldots, \alpha_{c1}, \alpha_{cn_c})^T$ and $k(x_{ij}, x_{jk}) = \langle \emptyset(x_{ij}), \emptyset(x_{jk}) \rangle, \langle \, , \rangle$ is the inner-product.

Thus, by using the definitions of $S_b^{\emptyset}, S_w^{\emptyset}$ and equation (18), the numerator and denominator of Eq. (14) can be rewritten as follows:

$$W^T S_b^{\emptyset} W = \alpha^T K_b \alpha \text{ and } W^T S_b^{\emptyset} W = \alpha^T K_w \alpha \quad (19)$$

where

$$K_b = \frac{n_i}{N} \sum_{i=1}^{c} (M_i - M)(M_i - M)^T \quad (20)$$

and,

$$K_w = \frac{1}{N} \sum_{i=1}^{c} \sum_{j=1}^{n_i} (\xi_{ij} - M_i)(\xi_{ij} - M_i)^T \quad (21)$$

Where

$$M_i = \begin{pmatrix} \frac{1}{n_i} \sum_{k=1}^{n_i} k(x_{11}, x_{ik}), \ldots, \frac{1}{n_i} \sum_{k=1}^{n_i} k(x_{1n_i}, x_{ik}), \ldots, \\ \frac{1}{n_i} \sum_{k=1}^{n_i} k(x_{c1}, x_{ik}), \ldots, \frac{1}{n_i} \sum_{k=1}^{n_i} k(x_{cn_c}, x_{ik}) \end{pmatrix}^T \quad (22)$$

$$M = (\frac{1}{N} \sum_{j=1}^{N} k(x_1, x_j), \ldots, \frac{1}{N} \sum_{j=1}^{N} k(x_N, x_j))^T \quad (23)$$

$$\xi_{ij} = \begin{pmatrix} k(x_{11}, x_{ij}), \ldots, k(x_{1n_i}, x_{ij}), \ldots, \\ k(x_{c1}, x_{ij}), \ldots, k(x_{cn_c}, x_{ij}) \end{pmatrix}^T \quad (24)$$

In this way, maximizing expression in equation (14) is converted to maximize:

$$J = \frac{|\alpha^T K_b \alpha|}{|\alpha^T K_w \alpha|} \quad (25)$$

Similar to original LDA, the optimal solution of equation (25) can be computed by finding the leading (ℓ) eigenvectors $\{a_i\}_{i=1,\ldots,\ell}$of $K_w^{-1} K_b$ corresponding to the nonzero eigenvalues. Once A=$[a_1, \ldots, a_\ell]$ is obtained, for a given pattern x, we can map it to a ℓ $-$dimensional space spanned by the columns of A. This projection is given by $y = (a_1, \ldots, a_\ell)^T \xi_{ij}$.

5. Kernel Weight Scatter Two Dimensional Discriminant Analysis (KWS2DDA)

Recently, Liu et. al., [16] have proposed a method called kernel scatter-difference based discriminant analysis (KWS2DDA) to extract non-linear discriminating features without numerical computation problem. Hence, a new scatter-difference-based discriminant rule has been defined to analyses data in an implicit feature space F and produce non-linear discriminant features by solving the following criterion:

$$J_M(W) = W^T (S_b^{\emptyset} - M S_w^{\emptyset}) W \quad (26)$$

Where M is a non-negative constant to balance $S_w^{\emptyset}$ and $S_b^{\emptyset}$ according to the description in section (3), the maximization problem of equation (26) is converted to maximize:

$$J_M(\alpha) = \alpha^T (K_b - MK_w)\alpha \tag{27}$$

The solution of this problem can be reached by finding the leading eigenvectors of $K_b - MK_w$.

Tang et. al., [8] have introduced a weighting scheme to estimate the within-class scatter matrix using a so called relevance weights. Hence, taking into account the previous ideas, we introduce weighted scatter matrices and thus define a weighted scatter-difference-based discriminant analysis criterion as follows:

$$J_M(W) = \text{argmax}_w \text{trace}(W^T\left(\hat{S}_b^{\phi} - M\hat{S}_w^{\phi}\right)W) \tag{28}$$

where,

$$\hat{S}_b^{\phi} = \sum_{i=1}^{C-1} \sum_{j=i+1}^{C} wP_iP_j \left(m_i^{\phi} - m_j^{\phi}\right)(m_i^{\phi} - m_j^{\phi})^T \tag{29}$$

$$\hat{S}_w^{\phi} = \sum_{i=1}^{C} P_i r_i \sum_{j=1}^{n_i} (\phi_j - m_i^{\phi})(\phi_j - m_i^{\phi})^T \tag{30}$$

where w, P_i, P_j and r_i are defined in section (III). As same as traditional KDA presented in previous section, using kernel trick, the optimization criterion in equation (26) can be transformed as follows:

$$\hat{J}_M(\alpha) = \alpha^T\left(\hat{K}_b - M\hat{K}_w\right)\alpha \tag{31}$$

where

$$\hat{K}_b = \sum_{i=1}^{C-1} \sum_{j=i+1}^{C} wP_iP_j\left(m_i - m_j\right)(m_i - m_j)^T \tag{32}$$

and

$$\hat{K}_w = \sum_{i=1}^{C} P_i r_i \sum_{j=1}^{n_i} (\xi_{ij} - m_i)(\xi_{ij} - m_i)^T \tag{33}$$

With m_i and ξ_{ij} having the same expression as in equations (22) and (24). Finally we find equation (31) where M having the same expression in equation (13).

Lastly, the solution of KWS2DDA optimization problem in equation (31) can be easily obtained by the leading eigenvectors of the matrix $\hat{K}_b - M\hat{K}_w$. So, in order to further improve the performance of our proposed algorithm, we utilize the kernel function called Gaussian RBF kernel function as:

$$k(x, y) = \exp(-\frac{\|x-y\|^2}{2\sigma^2}) \tag{34}$$

To achieve the best solution we make change in this function to become:

$$k(x, y) = \exp(-\frac{\sqrt{\|x-y\|^{2h}}}{2\sigma^2}) \tag{35}$$

Where h=3 and $\sigma^2 = 10^7$.

Note that with h=2 the extended expression given by equation (35) refuses to the Gaussian RBF given by the equation (34).

Table 1. Kernel Functions

1. Gaussian RBF Kernel

   $k(x, y) = \exp(-\|x-y\|^2)/2\sigma^2$

2. Power distance Kernel

   $k(x, y) = -\|x-y\|^{\beta}$ where $0 < \beta \leq 1$

3. Logarithmic Kernel

   $k(x, y) = -\log(1 + \|x-y\|^{\beta})$ where $0 < \beta \leq 2$

6. Experiment and Results

To illustrate the effectiveness of the KWS2DDA algorithm a set of experiments are performed. The ORL, Yale and the extended Head pose face database were used. In our experiments, we have used a 2.09 GHZ Mac Book computer with 2GHz RAM and the Open CV with visual C++ development environment. The detailed results of these experiments will be presented in this section.

6.1. The Experiments on the ORL Face Base

For showing the effect of KWS2DDA, we use ORL database [17]. This base contains images from 40 individuals, each providing 40 different images. For some subjects, the images were taken at different times. The facial expressions (open or closed eyes, smiling or non-smiling) and facial details (glasses or no glasses) also vary. The images were taken with a tolerance for some titling and rotation of the face of up to 20 degrees. Moreover, there is also some variation in the scale of up to about 10 percent. All images are grayscale and normalized to a resolution of 112×92 pixels. Example images are shown in Figure 1.

Figure 1. Examples of ORL Database

Here we use between 3 to 9 images sample per class for training, and the remaining images for the test and we take five cases with different input and find the mean of this five cases. Table 2 shows the comparisons result on recognition accuracy. We can see from this table that the result of testing KWS2DDA is better than the results of other method in some cases.

For the comparison of cup time(s) for feature extraction of ORL databases, it can be seen from Table 3 that all algorithms take much less time than KWS2DDA, because KWS2DDA is very complex in computational and take many steps. To illustrate the effect of kernel function choice, in the delivered performance of nonlinear algorithms, we have performed KWS2DDA with different kernel functions. For Gaussian RBF kernel, the value $\sigma^2 = 10^7$ gives maximum recognition rate compared with the others values ofσ^2this shown in Figure 2. In Table 4 we report the results for kernel functions commonly used in the face recognition literature namely Gaussian RBF kernel by different h parameters.

Table 2. Comparisons Results of Recognition Accuracy on ORL Database

k Algorithm	3	4	5	6	7	8	9
PCA+LDA	91.66	94.36	96	97.5	98.87	99.17	99.17
2DPCA	90.57	94.65	96.15	97.71	98.87	99.17	99.17
2DLDA	92.1	96.89	97.33	98.75	99.16	99.17	100
WS2DLDA	93.31	97.73	98.67	98.9	99.2	99.37	100
KWS2DDA	93.5	98.44	98.83	99.58	99.72	99.88	100

Table 3. Recognition Times of ORL Database

k Algorithm	3	4	5	6	7	8	9
PCA+LDA	7.7	8.5	10.6	9.7	9.9	9.5	7
2DPCA	7	7.4	7	7.6	7	5.5	4
2DLDA	4	4.5	5	5	4.5	4.2	4
WS2DLDA	5.4	5.7	5.4	5.4	4.7	4.3	4
KWS2DDA	60	70	81	80	82	68	52

Table 4. Comparisons Results of Recognition Accuracy on ORL Database by Gaussian RBF Kernel

K	3	4	5	6	7	8	9
h = 1	86.21	93.24	96	97.81	98.72	99.58	100
h = 2	87.13	94.29	95.75	97.81	98.7	98.8	100
h = 3	93.5	98.44	98.83	99.58	99.72	99.88	100

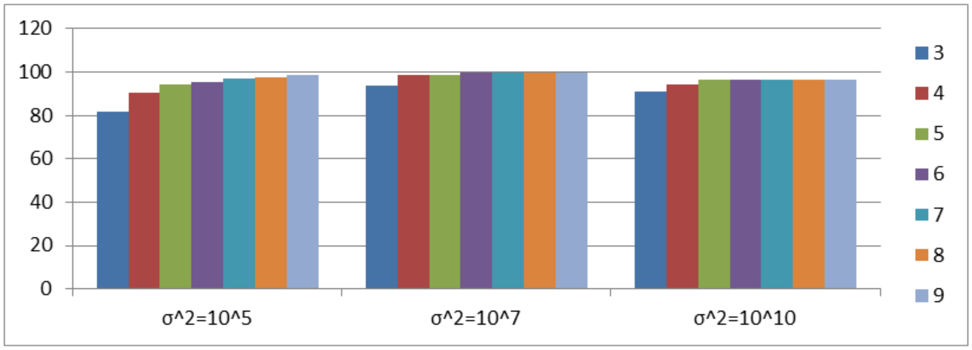

Figure 2. Recognition Rate for Different σ^2 on Oral Database

6.2 Experiment on the Yale Database

The next experiment is performed using the Yale face database [18], which contains 165 images of 15 individuals (each person has 11 different images) under various facial expressions and lighting conditions. Each image is manually cropped and resized 92 × 112 pixel in this experiment. We use a number of training images between 3 to 10 and the remaining images for test and we take five cases with different input and find the mean of this five cases. From the experimental results are listed in Table 5, we can see that the recognition rate of KWS2DDA is superior to other methods. Example of images is shown in Figure 3.

Figure 3. Examples of Yale Database

Hence, we can show that for power distance logarithmic and power distance kernel the performance decrease with the parameter β and globally gives less result than Gaussian RBF. The performance of fractional power distance kernel with value d = 0.5 is good but it is less than that of logarithmic kernel at β = 1 and β=2 On the other hand, Table 6 gives the average recognition rates for KWS2DDA when logarithmic and power distance kernel functions are used. We note that the logarithmic power distance kernel at β= 1 and β=2 provides a best performance compared to the other kernel functions. The power distance kernel gives acceptable but relatively lower rates.

Table 5. Comparisons Results of Recognition Accuracy on YAL B Database

k / Algorithm	3	4	5	6	7	8	9	10
PCA+LDA	83.22	89.26	95.86	92.89	95	98.51	97.78	97.78
2DPCA	83.38	89.89	96.24	96	95.56	98.51	97.78	97.78
2DLDA	90.5	90.51	95.51	94.67	96.11	99.24	97.78	100
WS2DLDA	93.45	92.51	98.13	98.22	98.33	99.24	98.89	100
KWS2DDA	94.4	97.67	99.15	99.4	99.4	99.4	99.67	100

Table 6. Comparisons Results of Recognition Accuracy on YAL B Database for Kernel Methods

K	4	5	6	7	8	9	10
Power distance kernel							
B = 0.5	83.26	94.7	91.33	97.5	99.24	98.9	97.7
B =0.7	81.06	91.5	90	97.7	98.5	98.9	97.8
B =1	82.1	97.63	96.33	97.5	98.5	97.8	97.8
Logarithmic kernel							
B =0.5	84.24	91.5	90	98.33	99.24	98.9	97.7
B =1	93.23	91.5	90	98.33	99.24	98.9	97.7
B =2	93.23	91.5	90	98.33	99.24	98.9	97.7

6.3 Experiment on the Head Pose Database

Next we considered and experiment with regard to the head pose database [19]. This database has a benchmark of 2790 monocular face images of 15 persons with variations of pan and tilt angles in the range (-90 , 90) degrees and 2 series of 93 images for every person (93 different poses) are available. The purpose of having 2 series per person is to be able to train and test algorithms on known and unknown faces. People having various skin colors in the database were classified as either wear glasses or not. Background is willingly neutral and uncluttered in order to focus on face operations. To obtain different poses, markers are posed in the whole room. Each marker corresponds to a 2D pose (pan, tilt) are used as markers. The whole set of post-it covers a half-sphere in front of the person see Table 7. Each image is manually cropped and resized by 52 $\times$72 pixel. We use a number of training images between 10 to100 and 88 different images for test.

In this experiment compares KWS2DDA with different kernel functions for face recognition. The results are given in Tables 8 and 9. One can note that Recognition rates are relatively high when Gaussian RBF kernel function is used especially at degrees h = 3 and $\sigma^2 = 10^7$, we can see that the recognition rate of the proposed KRW2DDA is superior to other methods. Face positions on each image are labeled in an individual text file. In Figure 4 a small sample of this database.

Table 7. Pan and Tilt Angles of Head Pose Database

	Negative values	Positive values
Pan angel	Bottom	Top
Tilt angel	Left	Right
Pan (Vertical angle)	{-90, -75, -60, -45, -30, -15, 0, +15, +30, +45, +60, +75, +90}	
Tilt (Horizontal angle)	{-90, -60, -30, -15, 0, +15, +30, +60, +90}	

Figure 3. A Small Sample of Head Pose Database

Table 8. Comparison Result of Recognition Accuracy on Head Pose Database

k Algorithm	10	40	50	60	70	80	90	100
PCA+LDA	87.8	96.2	96.7	97.5	98.1	99.4	99.4	99.4
2DPCA	88.5	91.06	93.1	96	98	99.4	99.4	99.4
2DLDA	87.9	96.4	97.1	97.8	98.5	99.4	99.4	99.4
WS2DLDA	88.2	96.5	97.53	97.93	98.5	99.3	99.6	99.6
KWS2DDA	92.3	98.5	98.6	98.7	98.6	99.4	99.7	99.85

Table 9. Comparisons Results of Recognition Accuracy on Head Pose Database for Kernel Methods

K	10	40	50	60	70	80	90	100
Power distance kernel								
B = 0.5	89.8	95.7	95.5	96.6	98.4	98.8	99	99
B =0.7	89.4	95.3	96.1	96.6	98.4	98.7	98.9	98.9
B =1	87.4	95.5	96	96.5	98.5	98.1	98.9	98.9
Logarithmic kernel								
B =0.5	93.1	96.7	96.9	97.1	98.7	98.9	98.9	98.9
B =1	92.4	96.6	96.8	97.1	98.7	99	98.9	98.9
B =2	92.4	96.6	96.8	97.2	98.7	99	98.9	98.9

7. Conclusion

A novel feature extraction method, kernel weighted scatter two dimensional discriminant analysis (KWS2DDA) is proposed in this paper. The faces used during experimentation are ORl face database, Yale face database and Head Pose database demonstrate that proposed method gives good properties of face recognition rate.

We came to state the following facts, the recognition rate of KWS2DDA is better than the others method, the execution time of WS2DLDA and 2DLDA is less than KWS2DDA. Finally we prove that the cup time depends on the size of the face images and the number of class. The proposed algorithm is relatively more effective in face recognition as compared with the other competing methods.

References

[1] R. O. Duda, P. E. Hart and D. Stork, “Pattern Classification”, Wiley, **(2000)**.
[2] K. Fukunage, “Introduction to statistical Pattern Classification”, Academic Press, San Diego, California, USA, **(1990)**.
[3] P. N. Belhumeour, J. P. Hespanha and D. J. Kriegman, “Eienfaces vs. Fisherfaces: Recognition using class specific linear projection”, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 19, no. 7, **(1997)**, pp. 711-720.
[4] D. l. Swets and J. Weng, “Using discriminant eigenfeatures for image retrieval”, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 18, no. 8, **(1996)**, pp. 831-836.
[5] S. Dudoit, J. Fridlyand and T. P. Speed, “Comparison of discrimination methods for the classification of tumors using gene expression data”, Journal of the American Statistical Association, vol. 97, no. 457, **(2002)**, pp. 77-87.
[6] W. J. Krzanowski, P. Jonathan, W. V. McCarthy and M. R. Thomas, “Discriminant analysis with singular covariance matrices: methods and applications to spectroscopic data”, Applied Statistics, vol. 44, **(1995)**, pp. 101-115.
[7] M. Li and B. Yuang, “2D-LDA: A Novel Statistical Linear Discriminant Analysis for Image Matrix”, Pattern Recognition. Lett., vol. 26, no. 5, **(2005)**, pp. 527-532.
[8] E. K. Tang, P. N. Suganthan, X. Yao and A. K. Qin, “Linear dimensionality reduction using relevance weighted LDA”, Pattern recognition, vol. 38, **(2005)**, pp. 485–493.
[9] G. Baudat and F. Anouar, “Generalized discriminant an alysisusing a kernel approach”, Neural Comput., vol. 12, no. 10, **(2000)**, pp. 2385–2404.
[10] B. Scholkopf, A. Smola and K. R. Muller, “Nonlinear component analysis as a Kernel eigenvalue problem”, Neural Comput., vol. 10, no. 5, **(1998)**, pp. 1299–1319.
[11] M. H. Yang, N. Ahuja and D. Kriegman, “Face recognition using kernel eigenfaces”, In: Proceedings of international conference on image processing, **(2000)**, pp. 37–40.
[12] S. Mika, G. Ratsch and J. Weston, “Fisher discriminant analysis with kernels”, Neural Netw Signal Process, vol. 9, **(1999)**, pp. 41–48.
[13] Q. Liu, R. Huang, H. Lu and S. Ma, “Face recognition using kernel-based Fisher discriminant analysis”, In: Proceedings of the 5th IEEE international conference on automatic face and gesture recognition, **(2002)**, pp. 205–211.
[14] M. H. Huang, “Kernel eigenfaces vs. Kernel Fisherfaces: face recognition using kernel methods”, In: Proceedings of the 5th IEEE international conference on automatic face and gesture recognition, **(2002)**, pp. 215–220.
[15] Chougdali, Jedra et Zahid, “Kernel weighted scatter difference discriminant analysis”, Proceeding in ICIAR, **(2008)**, pp. 977–983.
[16] Q. Liu, X. Tang, H. Lu and S. Ma, “Face recognition using Kernel Scatter-Difference based Discriminant analysis”, IEEE Trans. Neural Networks, vol. 17, **(2006)**, pp. 1081–1085.
[17] http://www.cam-orl.co.uk.
[18] ftp://plucky.cs.yale.edu/CVC/pub/images/y B/TarSets.
[19] http://www.prima.inrialpes.fr/perso/Gourier/Faces/HPDatabase.html.

Authors

Hythem Ahmed Abdalla was born in Aldamer, Sudan in 1976. He graduated from the Faculty of science and Technology, university of AL NEELAN in Khartoum, Sudan, and received the M.Sc degree in computer science in 2003. Currently, he is preparing his Doctor at National degree in computer science. His research interests are biometrics, pattern recognition, machine intelligence and image processing.

Mohamed Jedra received the « Doctorat deTroisième Cycle » and « Doctorat d'Etat »degrees in Electronics Engineering and Informatics all from Mohammed V University in Rabat, Morocco. From 1990 to 1999, he was the Network and Internet Center Manager in the Faculty of Sciences and Assistant Professor at the Department of Physics. In 1999, he became a Professor Habilité of Informatics in the same Department and in 2003 he was promoted to the position of Professor. From 1987 to the present, he was a member of the Conception & Systems Laboratory. He is the co-founder and the current Chair of the Architecture of Informatics Systems UFR/ Formation and Research Unit in the Faculty of Sciences in Rabat since 2003. His main research interests include computational intelligence, pattern recognition and biometric. He was a reviewer of several conferences dealing with topics related to computational intelligence and pattern recognition. He has published numerous refereed papers in specialized journals and conferences. Mohamed Jedra is a member of IEEE since 1995. He is also a member of IEEE Computer Society, IEEE Computational Intelligence Society and IEEE Signal Processing Society.

Noureddine Zahid holds « Doctorat de Troisième Cycle » and « Doctorat d'Etat » degrees in Electronics Engineering and Informatics; all form the Mohammed V University in Rabat, Morocco. He was an Assistant Professor in computer science at the Department of Physics between 1988 and 1999, and Professor "Habilité" between 1999 and 2003. In 2003, he was promoted to the position of Professor in the same department. He is a member of Conception & Systems Laboratory. He is the cofounder and the current Chair of the Informatics Telecommunications and Imagery master (ITI) in the Faculty of Science in Rabat. He has presented and published many articles in scientific journals and conferences. He has reviewed numerous papers for international journals in the field of fuzzy systems and pattern recognition. His research interests lie in the areas of fuzzy logic, pattern recognition, machine learning and biometric.