

- [23] Y. He, M. Wu, and J. H. She, "An improved global asymptotic stability criterion for delayed cellular neural networks," *IEEE Trans. Neural Netw.*, vol. 17, no. 1, pp. 250–252, Jan. 2006.
- [24] Z. Wang, D. W. C. Ho, and X. Liu, "State estimation for delayed neural networks," *IEEE Trans. Neural Netw.*, vol. 16, no. 1, pp. 279–284, Jan. 2005.
- [25] V. T. S. Elanayar and Y. C. Shin, "Approximation and estimation of nonlinear stochastic dynamic systems using radial basis function neural networks," *IEEE Trans. Neural Netw.*, vol. 5, no. 4, pp. 594–603, Jul. 1994.
- [26] R. Habtom and L. Litz, "Estimation of unmeasured inputs using recurrent neural networks and the extended Kalman filter," in *Proc. Int. Conf. Neural Network*, Houston, TX, 1997, vol. 4, pp. 2067–2071.
- [27] F. M. Salam and J. Zhang, "Adaptive neural observer with forward co-state propagation," in *Proc. Int. Joint Conf. Neural Networks (IJCNN'01)*, Washington, DC, 2001, vol. 1, pp. 675–680.
- [28] Y. He, M. Wu, J. H. She, and G. P. Liu, "Delay-dependent robust stability criteria for uncertain neutral systems with mixed delays," *Syst. Contr. Lett.*, vol. 51, no. 1, pp. 57–65, 2004.
- [29] —, "Parameter-dependent Lyapunov functional for stability of time-delay systems with polytopic-type uncertainties," *IEEE Trans. Autom. Control*, vol. 49, no. 5, pp. 828–832, May 2004.
- [30] M. Wu, Y. He, J. H. She, and G. P. Liu, "Delay-dependent criteria for robust stability of time-varying delay systems," *Automatica*, vol. 40, no. 8, pp. 1435–1439, 2004.
- [31] M. Wu, Y. He, and J. H. She, "New delay-dependent stability criteria and stabilizing method for neutral systems," *IEEE Trans. Autom. Control*, vol. 49, no. 12, pp. 2266–2271, Dec. 2004.
- [32] Y. S. Moon, P. Park, W. H. Kwon, and Y. S. Lee, "Delay-dependent robust stabilization of uncertain state-delayed systems," *Int. J. Contr.*, vol. 74, no. 14, pp. 1447–1455, 2001.

## Face Recognition Using Kernel Scatter-Difference-Based Discriminant Analysis

Qingshan Liu, Xiaoou Tang, Hanqing Lu, and Songde Ma

**Abstract**—There are two fundamental problems with the Fisher linear discriminant analysis for face recognition. One is the singularity problem of the within-class scatter matrix due to small training sample size. The other is that it cannot efficiently describe complex nonlinear variations of face images because of its linear property. In this letter, a kernel scatter-difference-based discriminant analysis is proposed to overcome these two problems. We first use the nonlinear kernel trick to map the input data into an implicit feature space  $F$ . Then a scatter-difference-based discriminant rule is defined to analyze the data in  $F$ . The proposed method can not only produce nonlinear discriminant features but also avoid the singularity problem of the within-class scatter matrix. Extensive experiments show encouraging recognition performance of the new algorithm.

**Index Terms**—Face recognition, Fisher linear discriminant analysis, kernel scatter-difference-based discriminant analysis.

### I. INTRODUCTION

Principal component analysis (PCA) and Fisher linear discriminant analysis (FLDA) are two popular feature representation methods.

Manuscript received January 1, 2005; revised August 5, 2005. This work was supported by Natural Sciences Foundation of China under Grant 60405005 and 60135020, the joint fund of NSFC-RGC under Grant 60318003, and RGC N\_CUHK409/03.

Q. Liu, H. Lu, and S. Ma are with National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences, Beijing 100080, China (e-mail: qslu@nlpr.ia.ac.cn; luhq@nlpr.ia.ac.cn; masd@nlpr.ia.ac.cn).

X. Tang is with Department of Information Engineering, The Chinese University of Hong Kong, Hong Kong, China (e-mail: xtang@ie.cuhk.edu.hk).

Digital Object Identifier 10.1109/TNN.2006.875970

PCA generates a set of orthonormal basis vectors aiming at maximizing the variance over all the samples. A new face image can be represented by a linear combination of these basis vectors, i.e., Eigenface [1]. PCA is optimal for reconstruction but is not optimal for discrimination. For better discrimination, FLDA tries to find a linear projection to maximize between-class scatter  $S_B$  and minimize within-class scatter  $S_W$ , which can be obtained by maximizing the Fisher discriminant function  $J(w) = w^T S_B w / w^T S_W w$ . Mathematically, it is equivalent to solving leading eigenvectors of  $S_W^{-1} S_B$ . For the relationship between PCA and LDA, a detailed analysis is given in [2] and [3].

There exist two key problems for face recognition using FLDA. One is that  $S_W$  tends to be singular, since face recognition often has a small training sample size. This gives rise to a problem of unstable numerical computation. Several techniques have been developed to alleviate this problem [4]–[11]. For example, the well-known Fisherface employs PCA at first for dimension reduction, and then FLDA is performed [4], [5]. Belhumeur *et al.* [4] and Zhao *et al.* [5] showed that Fisherface gave a much higher recognition performance than Eigenface. Liu *et al.* [10] proposed simultaneous diagonalization of the within- and between-class scatter matrices to improve the FLDA. Direct LDA proposed by Yu and Yang focuses on between-class scatter [8], while the null space methods use information in the null space of within-class scatter only [7], [9]. In [11], Wang and Tang used random sampling strategy to handle the problem of small training sample.

Another problem with FLDA is that as a linear representation method it cannot well describe complex nonlinear variations of images with illumination, pose, and facial expression changes. Since the nonlinear kernel trick achieved a great success in support vector machine (SVM) [12], recently kernel based nonlinear analysis has attracted a great deal of attention. Mika *et al.* [13] and Baudat *et al.* [14] proposed a kernel-based nonlinear discriminant analysis, which combines the kernel trick with FLDA, i.e., kernel Fisher discriminant analysis (KFDA). However, the within-class scatter in the KFDA feature space is often singular due to small training sample size. Several techniques have been proposed to handle the problem of numerical computation [13]–[20].

In this letter, we propose a kernel scatter-difference-based discriminant analysis (KSDA) method to overcome both the matrix singularity problem and the nonlinear problem for face recognition. The kernel trick is first employed to construct an implicit feature space  $F$ , and then a new scatter-difference based discriminant rule is defined to analyze the data in  $F$  and produce nonlinear discriminating features. The scatter-difference-based discriminant rule is consistent with the principle of maximizing between-class scatter and minimizing within-class scatter. It can also avoid the matrix singularity problem. This letter is an extension of our previous work [15], [16], [18], and it differs from [15] and [16], in which a new discriminant rule is introduced in this letter and expands on [18] by including more experiments and an estimation and analysis of a balance factor  $M$  between between-class and within-class scatters. The proposed method approximates the null space method [7], [9] in  $F$  when  $M \rightarrow \infty$  [20]. The maximum margin criterion proposed in [21] can also be regarded as a special case of the method at  $M = 1$ . We conduct the experiments on the FERET, ORL, and YALE databases and give a comparison with six popular methods [1], [4], [15], [17], [19], [20].

The rest of this letter is arranged as follows. KFDA is reviewed in Section II. KSDA is described in Section III. Experiments are reported in Section IV, and conclusions are given in Section V.

## II. KERNEL FISHER DISCRIMINANT ANALYSIS

Before describing the proposed method, we briefly review KFDA in this section. The idea of KFDA is to solve the problem of FLDA in an implicit feature space  $F$  constructed by the nonlinear mapping  $\phi : x \in R^N \rightarrow \phi(x) \in F$ . In implementation, implicit feature vector  $\phi$  does not need to be calculated explicitly; instead it is embodied by computing the inner product of two vectors in  $F$  with a kernel function  $k(x, y) = (\phi(x) \cdot \phi(y))$ .

Let  $x$  be a vector of the input set with  $n$  elements and  $C$  classes, and the  $i$ th class has  $n_i$  samples. The mapping of  $x_i$  is noted as  $\phi_i = \phi(x_i)$ . Performing FLDA in  $F$  means maximizing the Fisher discriminant function  $J(w) = (w^T S_B^{\phi} w / w^T S_W^{\phi} w)$ , where  $S_B^{\phi} = \sum_{i=1}^C (\mu_i - \bar{\mu})(\mu_i - \bar{\mu})^T$  and  $S_W^{\phi} = \sum_{i=1}^C (1/n_i) \sum_{j=1}^{n_i} (\phi_j - \mu_i)(\phi_j - \mu_i)^T$  represent the between-class scatter and within-class scatter in  $F$  respectively, and  $\mu_i = (1/n_i) \sum_{j=1}^{n_i} \phi_j$  and  $\bar{\mu} = (1/n) \sum_{i=1}^C \phi_i$  are the mean of the  $i$ th class samples and the mean of all the samples in  $F$ , respectively.

Because  $w \in F$  must lie in the span of all the samples in  $F$ ,  $w$  can be represented by a linear combination of  $\phi_i$ , i.e.,  $w = \sum_{i=1}^n \alpha_i \phi_i$ . Thus, the problem of KFDA is converted into maximizing the function  $J(\alpha) = (\alpha^T K_B \alpha / \alpha^T K_W \alpha)$ , and its solution is the leading eigenvectors of  $K_W^{-1} K_B$ , where  $K_B = \sum_{i=1}^C (m_i - \bar{m})(m_i - \bar{m})^T$ ,  $K_W = \sum_{i=1}^C (1/n_i) \sum_{j=1}^{n_i} (\zeta_j - m_i)(\zeta_j - m_i)^T$ ,  $\zeta_j = (k(x_1, x_j), k(x_2, x_j), \dots, k(x_n, x_j))^T$ ,  $m_i = ((1/n_i) \sum_{j=1}^{n_i} k(x_1, x_j), (1/n_i) \sum_{j=1}^{n_i} k(x_2, x_j), \dots, (1/n_i) \sum_{j=1}^{n_i} k(x_n, x_j))^T$ , and  $\bar{m}$  is the mean of all of  $\zeta_i$ . For a new sample  $x$ , its projection onto  $w$  in  $F$  can be calculated by  $(w \cdot \phi(x)) = \sum_{i=1}^n \alpha_i k(x_i, x)$ .

## III. KERNEL SCATTER-DIFFERENCE-BASED DISCRIMINANT ANALYSIS

Though KFDA can produce nonlinear discriminating features, the problem of numerical computation for face recognition still exists, i.e., the matrix  $K_W$  cannot be guaranteed to be nonsingular. There are some similar techniques to deal with this problem in KFDA as in FLDA. For example, we used  $K_W + \mu I$  to replace  $K_W$  in our previous work [15], [16], where  $\mu$  is a very small constant and  $I$  is the identity matrix. Yang [17] made  $K_W$  nonsingular by performing PCA in the feature space first (we call this method KEDA for convenience in the following). Lu *et al.* adopted the idea of direct LDA to develop the KDDA method [19]; and in [20], the idea of null space was used to develop the KNDA method.

In this letter, KSDA is proposed for face recognition, in which nonlinear discriminating features are extracted without the numerical computation problem. First, the kernel trick is performed to construct an implicit feature space  $F$ . A scatter-difference-based discrimination rule is then defined in  $F$  to produce nonlinear discriminating features as follows:

$$J_M(w) = w^T \left( S_B^{\phi} - M \cdot S_W^{\phi} \right) w \quad (1)$$

where  $M$  is a nonnegative constant to balance  $S_B^{\phi}$  and  $S_W^{\phi}$ . In a sense, maximizing  $J_M(w)$  is also equivalent to maximizing  $S_B^{\phi}$  and minimizing  $S_W^{\phi}$ , so it is consistent with the Fisher discriminant rule. Moreover, scatter-difference-based discriminant rule can be regarded as a generalization of the techniques in [7], [9], [20], and [21]. When  $M = 0$ , it means to use the between-class scatter only, and it approximates the null-space methods at  $M \rightarrow \infty$  [7], [9], [20]. The maximum

margin criterion proposed in [21] is also a special case with  $M = 1$ . According to the description in Section II, we can rewrite (1) as

$$J_M(\alpha) = \alpha^T (K_B - M \cdot K_W) \alpha. \quad (2)$$

Assuming  $\|\alpha\| = 1$ , the maximization problem of  $J_M(\alpha)$  is equivalent to solving the maximum of the Lagrange function

$$L(\alpha, \lambda) = J_M(\alpha) - \lambda (\|\alpha\| - 1). \quad (3)$$

Let  $(\partial L(\alpha, \lambda) / \partial \alpha) = 0$ . We can have

$$(K_B - M \cdot K_W) \alpha = \lambda \alpha. \quad (4)$$

Thus, the problem of KSDA is translated into finding the leading eigenvectors of  $K_B - M \cdot K_W$ . Since no matrix inverse needs to be computed, KSDA successfully avoids any numerical computation problem.

The factor  $M$  is used to balance the matrix  $K_B$  and  $K_W$ , and its value depends on the training data. Practically  $M$  can be regarded as balancing the energy variations of  $K_B$  and  $K_W$ . Assuming the energy variation ratio between the matrix  $K_B$  and  $K_W$  is  $\beta$ , we can use the product of  $\beta$  and a nonnegative constant  $D$  to approximate  $M$ , i.e.,  $M = D \cdot \beta$ , where  $D$  is independent of the matrix and  $K_B$  and  $K_W$ . Then (4) is converted into

$$(K_B - D \cdot \beta \cdot K_W) \alpha = \lambda \alpha. \quad (5)$$

Given a database,  $K_B$ ,  $K_W$ , and  $\beta$  are known, so the optimal discriminating features are only related to the independent factor  $D$ . Therefore we can use a good  $D$  value estimated from a database as an empirical estimation of  $D$  for other databases. Certainly, how to define and evaluate the energy variation ratio  $\beta$  is not trivial. For simplicity, we use the largest eigenvalues of  $K_B$  and  $K_W$  to approximate their energy variations in this letter, noted as  $\lambda_{max}(K_B)$  and  $\lambda_{max}(K_W)$ , i.e.,  $\beta = \lambda_{max}(K_B) / \lambda_{max}(K_W)$ . The following experiments show that such a simplification is suitable.

## IV. EXPERIMENTS

We conduct experiments on three databases, i.e., the FERET, ORL, and YALE databases. On the FERET database, the FERET testing method is used. We use the "leave one out" statistical testing strategy on the ORL and YALE database. The proposed method is compared with KFDA [15], KEDA [17], KDDA [19], KNDA [20], and two well-known linear methods, i.e., Fisherface [4] and Eigenface [1]. In our experiments, the polynomial kernel is chosen,  $k(x, y) = (\phi(x) \cdot \phi(y)) = (a(x \cdot y) + b)^d$ , for it gave good performances in previous experiments [15]–[18], [20]. Its parameters are set as  $\alpha = 0.001$ ,  $b = 0$ , and  $d = 2$ , the same as in [15], [16], [19], and [20]. In [15] and [16],  $K_W$  is replaced by  $K_W + \mu I$  to deal with the problem of numerical computation. As in [15] and [16], we set  $\mu = 10^{-4}$  in the experiments.

### A. On the FERET Database

The FERET database has been widely used to evaluate face recognition methods [6]. We test the experiments on the FA and FB images of 1195 persons. Each person has two images with expression variation. One is in the FA set and the other is in the FB set. All the images are normalized to  $48 \times 54$  by eye locations. The coordinates of the two

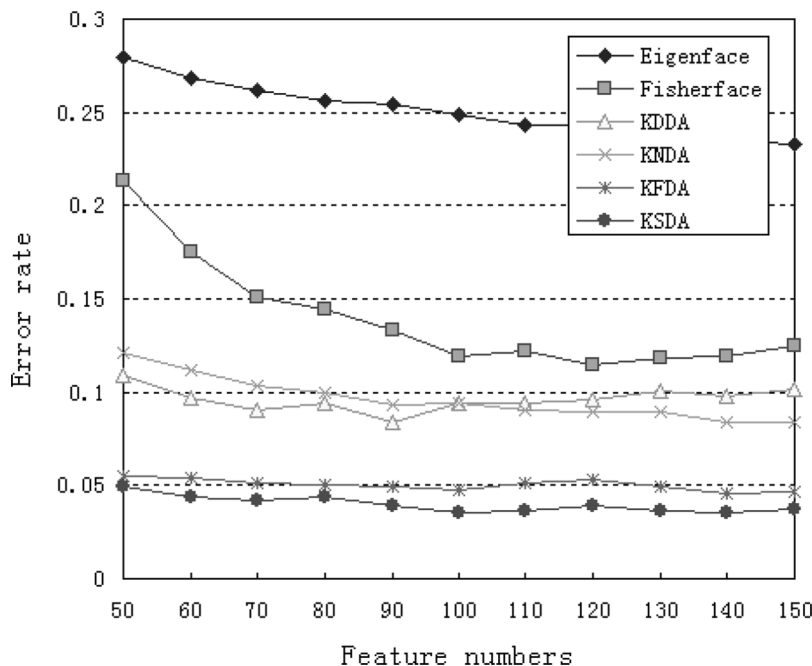


Fig. 1. Experimental results based on the Euclidean distance.

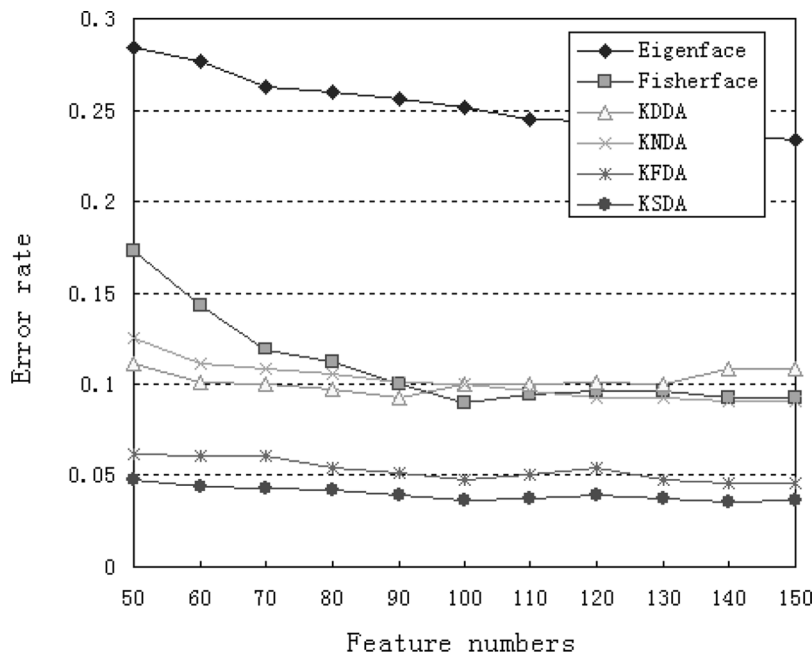


Fig. 2. Experimental results based on the cosine distance.

eyes are fixed at (12,14) and (36,14). No other preprocessing except histogram equalization is performed.

The experiments are performed according to the FERET testing protocol [6]. We randomly select 400 subjects, i.e., 800 images, from FA and FB images for learning basis vectors. We denote the number of basis vectors as feature numbers in the following figures and tables. For the other 795 persons, FA images are used as the gallery set and FB images are used as the probe set. Two popular distance measures are computed to test the performance of the proposed method, i.e., Euclidean distance  $d = \|x - y\|^2$  and cosine distance  $d = 1 - (x \cdot y / \|x\| \|y\|)$ .

Figs. 1 and 2 report the comparison results based on the Euclidean distance and the cosine distance, respectively. Here, we set  $\beta = 25$

due to  $\lambda_{max}(K_B) / \lambda_{max}(K_W) \approx 25$ , and we set  $D = 8$ , i.e.,  $M = 25 \times 8 = 200$  for KSDA. The performance of KSDA with different  $D$  values will be evaluated later. We can see that KSDA always shows better performance than KFDA, KEDA, KDDA, and KNDA. It implies that the numerical computation problem does affect the discriminative ability of KFDA. KDDA assumes that the null space of  $K_B$  contains no discriminating information and is removed first. It is demonstrated that this assumption is not true in [22], since the null space of  $K_B$  also has some discriminating information. KNDA only uses the information of the null space of  $K_W$ , so the discriminating information outside the null space of  $K_W$  is discarded. KEDA discards the information of the null space of  $K_W$ . KFDA adds a perturbation to make  $K_W$  nonsin-

TABLE I  
ERROR RATE (%) OF KSDA WITH DIFFERENT  $D$  VALUES

$D$ Values	$M = 1$	$D = 1$	$D = 2$	$D = 4$	$D = 6$	$D = 8$	$D = 10$	$D = 12$	$D = 20$	$D = 40$
Euclidean	30.82	10.19	10.06	7.92	4.53	<b>3.52</b>	<b>3.52</b>	<b>3.52</b>	3.65	3.90
Cosine	30.94	10.31	10.06	8.05	4.40	<b>3.52</b>	<b>3.52</b>	<b>3.52</b>	3.65	4.03

TABLE II  
ERROR RATES (%) OF SIX METHODS ON THE ORL DATABASE

Method	KSDA	KFDA	KEDA	KNDA	KDDA	Fisherface	Eigenface
Euclidean	<b>1.0</b>	2.0	2.5	2.0	2.75	2.5	3.0
Cosine	<b>1.25</b>	1.75	2.5	1.75	2.5	2.5	2.75
Feature numbers	39	39	39	39	39	39	80

TABLE III  
ERROR RATES (%) OF SIX METHODS ON THE YALE DATABASE

Method	KSDA	KFDA	KEDA	KNDA	KDDA	Fisherface	Eigenface
Euclidean	<b>2.42</b>	4.24	3.64	6.06	21.21	7.27	20.61
Cosine	<b>2.42</b>	4.24	3.64	6.67	21.21	4.84	21.21
Feature numbers	14	14	14	14	14	14	75

TABLE IV  
ERROR RATES (%) OF KSDA WITH DIFFERENT  $D$  VALUES ON THE ORL DATABASE

$D$ Values	$M = 1$	$D = 1$	$D = 2$	$D = 4$	$D = 6$	$D = 8$	$D = 10$	$D = 12$	$D = 20$	$D = 40$
Euclidean	2.5	1.5	1.25	<b>1.0</b>	<b>1.0</b>	<b>1.0</b>	<b>1.0</b>	<b>1.0</b>	1.25	1.75
Cosine	2.5	2.25	1.25	<b>1.25</b>	<b>1.25</b>	<b>1.25</b>	<b>1.25</b>	<b>1.5</b>	1.5	1.75

TABLE V  
ERROR RATES (%) OF KSDA WITH DIFFERENT  $D$  VALUES ON THE YALE DATABASE

$D$ Values	$M = 1$	$D = 1$	$D = 2$	$D = 4$	$D = 6$	$D = 8$	$D = 10$	$D = 12$	$D = 20$	$D = 40$
Euclidean	13.33	13.33	5.45	3.03	<b>2.42</b>	<b>2.42</b>	<b>2.42</b>	3.03	3.64	3.64
Cosine	10.30	10.30	5.45	2.42	<b>2.42</b>	<b>2.42</b>	<b>2.42</b>	3.03	3.64	4.24

gular. Because KSDA does not need to compute any matrix inverse for generating discriminating features, it successfully avoids the numerical computation problem. The performance of Fisherface with the cosine distance is better than the Euclidean distance, since the cosine distance is in a sense a nonlinear distance measure. This also implies that the nonlinear variations exist in real face images. Compared with the other six methods, Eigenface is an unsupervised and linear method, and PCA is only optimal for reconstruction, so it is understandable for Eigenface to give the worst performance.

As for KSDA, the balance factor  $M$  plays an important role on its performance. We decompose  $M$  into a production of  $\beta$  and  $D$ , where  $\beta$  represents the energy variation ratio between the matrix  $K_B$  and  $K_W$ . Theoretically  $\beta$  is known for a given database. Thus, we investigate the performance of KSDA with different  $M$  values, i.e., different  $D$  values. However, it is hard to give an accurate definition of  $\beta$ . In this letter, we use the largest eigenvalues of  $K_B$  and  $K_W$  to approximate their energy variations, and get  $\beta = 25$ . The experimental results are reported in Table I. It can be seen that the performances of KSDA for  $D = 8$  to 12 ( $M = 200$  to 300) are very good. The result of maximum margin criterion [21], i.e.,  $M = 1$ , is also given in Table I. We can see that the performance of KSDA is poor at  $M = 1$ , with the lowest error rate around 30%. It seems that maximum margin criterion [21] is not suitable for face recognition.

### B. On the ORL and YALE Databases

There are 40 persons in the ORL database with ten different images for each person, including variations in pose, facial expression, and

with or without glasses, but there is little illumination variation. The YALE database contains 11 subjects, and each subject has 11 different front view images that include variations in facial expression (normal, happy, sad, sleepy, surprised, and wink), illumination (left-light, center-light, right-light) and with or without glasses. We down-sample the image size  $45 \times 56$  to reduce computational complexity. No other pre-processing except histogram equalization is performed.

The experiments are performed with the “leave-one-out” statistical strategy, i.e., only one image is selected from the database as testing and the remaining images are used for training. We first compare KSDA with the other five methods. Table II shows their best results on the ORL database, where we set  $\beta = 4$  for KSDA because of  $\lambda_{\max}(K_B)/\lambda_{\max}(K_W) \approx 4$ . We set  $D = 8$  following the results on the FERET database, so we have  $M = 4 \times 8 = 32$ . Table III gives the best results on the YALE database, where we set  $\beta = 1$  since  $\lambda_{\max}(K_B)/\lambda_{\max}(K_W)$  approximates one, so we have  $M = 1 \times 8 = 8$  for KSDA. We can see that these results are consistent with the results on the FERET database, i.e., KSDA gives a better performance than the other six methods with both the Euclidean and cosine distances. In the case of the Euclidean distance, KSDA gives a recognition error rate of 1.0%, meaning that only four samples are incorrectly recognized among 400 samples. On the YALE database, KSDA has an error rate of 2.42%. The error rate of KDDA is very high on the YALE database, because the YALE database only has 15 classes, and KDDA discards the null space of  $K_B$ , i.e., the dimensions of  $K_B$  and  $K_W$  are first reduced to  $\text{ClassNum} - 1 = 14$ ; thus too much discriminating information is lost. In addition, each database embodies different variations in our experiments. The FERET database includes illumination and facial expression variations, and the ORL database

has pose and facial expression variations. The variations in the Yale database are illumination and facial expression. The experimental results show that KSDA is robust to all these variations.

We also investigate the performance of KSDA with different  $D$  values on these two databases. Tables IV and V illustrate the experimental results on the two databases, respectively. On the ORL database, KSDA gives a better performance in the range of  $D = 4$  to  $D = 12$ , i.e.,  $M$  is from 24 to 48. On the YALE database, KSDA performs well for the range of  $D = 6$  to 10, i.e.,  $M$  is from six to ten. These results are also consistent with the results on the FERET database. The performance of KSDA with  $M = 1$  is poor, as are the results on the FERET database.

## V. CONCLUSIONS

In this letter, we propose a KSDA algorithm for face recognition. First, the nonlinear kernel trick is employed to map the input data into an implicit feature space  $F$ . Then a scatter-difference-based discriminant rule is defined to analyze the data in  $F$  to produce nonlinear discriminating features. KSDA can not only produce nonlinear discriminating features based on the principle of maximizing the between-class scatter and minimizing the within-class scatter but also avoid the problem of numerical computation. Experiments on the FERET, ORL, and YALE databases show that KSDA can give a higher recognition rate than existing methods.

## REFERENCES

- [1] M. Turk and A. Pentland, "Eigenfaces for recognition," *J. Cogn. Neurosci.*, vol. 3, no. 1, pp. 71–86, 1991.
- [2] X. Wang and X. Tang, "A unified framework for subspace face recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 26, no. 9, pp. 1222–1228, Sep. 2004.
- [3] —, "Unified subspace analysis for face recognition," in *Proc. Int. Conf. Computer Vision*, 2003, pp. 679–686.
- [4] P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman, "Eigenfaces vs. fisherfaces: recognition using class specific linear projection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 19, no. 7, pp. 711–720, Jul. 1997.
- [5] W. Zhao, R. Chellappa, and P. J. Phillips, Subspace linear discriminant analysis for face recognition Center Automation Res., Univ. Maryland, Tech. Rep. CAR-TR-914, 1998.
- [6] P. J. Phillips, H. Wechsler, J. Huang, and P. Rauss, "The feret database and evaluation procedure for face recognition algorithms," *Image Vision Comput.*, vol. 16, no. 5, pp. 295–306, 1998.
- [7] L. F. Chen, H. M. Liao, J. C. Lin, M. T. Ko, and G. J. Yu, "A new LDA-based face recognition system which can solve the small sample size problem," *Pattern Recognit.*, vol. 33, no. 10, pp. 1713–1726, 2000.
- [8] H. Yu and J. Yang, "A direct LDA algorithm for high-dimensional data—With application to face recognition," *Pattern Recognit.*, vol. 34, no. 10, pp. 2067–2070, 2001.
- [9] R. Huang, Q. S. Liu, H. Q. Lu, and S. D. Ma, "Solving the small sample size problem of LDA," in *Proc. Int. Conf. Pattern Recognition*, 2002, vol. 3, pp. 29–32.
- [10] C. Liu and H. Wechsler, "Enhanced fisher linear discriminant models for face recognition," in *Proc. Int. Conf. Pattern Recognition*, 1998, vol. 2, pp. 1368–1372.
- [11] X. Wang and X. Tang, "Random sampling LDA for face recognition," in *Proc. Int. Conf. Computer Vision Pattern Recognition*, 2004, vol. 2, pp. 259–265.
- [12] E. Osuna, R. Freund, and F. Girosi, Support vector machines: Training and applications AI Lab, Massachusetts Inst. Tech., Cambridge, Tech. Rep., 1997.
- [13] S. Mika, G. Ratsch, and J. Weston, "Fisher discriminant analysis with kernels," in *Proc. Neural Networks Signal Processing Workshop*, 1999, pp. 41–48.
- [14] G. Baudat and F. Anouar, "Generalized discriminant analysis using a kernel approach," *Neural Comput.*, vol. 12, no. 10, pp. 2385–2404, 2000.

- [15] Q. S. Liu, R. Huang, H. Q. Lu, and S. D. Ma, "Face recognition using kernel based Fisher discriminant analysis," in *Proc. Int. Conf. Automatic Face Gesture Recognition*, 2002, pp. 197–201.
- [16] Q. S. Liu, H. Q. Lu, and S. D. Ma, "Improving kernel Fisher discriminant analysis for face recognition," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 14, no. 1, pp. 42–49, Jan. 2004.
- [17] M. H. Yang, "Kernel Eigenfaces vs. kernel Fisherfaces: Face recognition using kernel methods," in *Proc. Int. Conf. Automatic Face Gesture Recognition*, 2002, pp. 215–220.
- [18] Q. S. Liu, X. Tang, H. Q. Lu, and S. D. Ma, "Kernel scatter-difference based discriminant analysis for face recognition," in *Proc. Int. Conf. Pattern Recognition*, 2004, vol. 2, pp. 419–422.
- [19] J. Lu, K. N. Plataniotis, and A. N. Venetsanopoulos, "Face recognition using kernel direct discriminant analysis algorithms," *IEEE Trans. Neural Netw.*, vol. 14, no. 1, pp. 117–126, Jan. 2003.
- [20] W. Liu, Y. H. Wang, S. Z. Li, and T. N. Tan, "Null space-based kernel Fisher discriminant analysis for face recognition," in *Proc. Int. Conf. Automatic Face Gesture Recognition*, 2004, pp. 369–374.
- [21] H. F. Li, T. Jiang, and K. S. Zhang, "Efficient and robust feature extraction by maximum margin criterion," in *Proc. Int. Conf. Advance Neural Information Processing Systems*, 2004, pp. 97–104.
- [22] X. Wang and X. Tang, "Dual-space linear discriminant analysis for face recognition," in *Proc. Int. Conf. Computer Vision Pattern Recognition*, 2004, vol. 2, pp. 564–569.

## Analog Neural Network for Support Vector Machine Learning

Renzo Perfetti and Elisa Ricci

**Abstract**—An analog neural network for support vector machine learning is proposed, based on a partially dual formulation of the quadratic programming problem. It results in a simpler circuit implementation with respect to existing neural solutions for the same application. The effectiveness of the proposed network is shown through some computer simulations concerning benchmark problems.

**Index Terms**—Analog circuits, quadratic optimization, recurrent neural networks, support vector machines.

## I. INTRODUCTION

Support vector machines (SVMs) have been widely used as a robust tool for classification and regression [1]. Recently, an efficient neural network realization for SVM learning has been proposed in [2] and extended to more general optimization problems in [3]. It consists of a one-layer recurrent neural network whose steady-state solution satisfies the Karush–Kuhn–Tucker (KKT) conditions of the dual quadratic programming problem (QP). The neural network proposed in [2] has a lower complexity with respect to previous solutions [4]–[7] and can be used for the analog VLSI realization of SVM learning in real-time applications. In this letter, we propose a slightly different scheme, with comparable performance, which results in a simpler and more efficient circuit implementation. The letter is organized as follows. In Section II, we briefly review the SVM learning problem in the partially dual formulation. In Section III, we derive the proposed neural network and

Manuscript received August 23, 2005 ; revised February 27, 2006.

The authors are with the Department of Electronic and Information Engineering, University of Perugia, Perugia 06125, Italy (e-mail: perfetti@diei.unipg.it; elisa.ricci@diei.unipg.it).

Digital Object Identifier 10.1109/TNN.2006.875967