

Research Article

Distributed Face Recognition Using Multiple Kernel Discriminant Analysis in Wireless Sensor Networks

Xiao-Zhang Liu¹ and Guan Yang²

¹ School of Computer Science, Dongguan University of Technology, Dongguan 523808, China

² School of Computer Science, Zhongyuan University of Technology, Zhengzhou 450007, China

Correspondence should be addressed to Guan Yang; guanyang.sysu@yahoo.com

Received 17 October 2013; Accepted 18 October 2013; Published 23 January 2014

Academic Editor: Fatos Xhafa

Copyright © 2014 X.-Z. Liu and G. Yang. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This paper proposes a module based distributed wireless face recognition system by integrating multiple kernel discriminant analysis with face recognition in wireless sensor networks. By maximizing the margin maximization criterion (MMC), we separately perform an iterative scheme for kernel parameter optimization for each module. The simulation on the FERET and CMU PIE face databases shows that our multiple kernel framework and the optimization procedure achieve high recognition performance, compared with single-kernel-based KDDA.

1. Introduction

Face recognition (FR) system is one of the most important biometric techniques and is used in a wide range of security applications such as access control, identification systems, and surveillance [1]. FR is a contactless biometric technique and has advantages of being natural and passive over other biometric techniques requiring cooperative subjects, such as fingerprint recognition and iris recognition [2]. A normal framework of FR system is shown in Figure 1, including procedures of enrollment and identification [3].

In recent years, FR systems combined with wireless sensor networks (WSNs) [4] have shown great interest, as WSNs are very helpful for contactless biometrics security applications. For example, Kim et al. implement a wireless face recognition system based on ZigBee protocol and principle components analysis (PCA) method with low energy consumption [5]. Muraleedharan et al. propose the use of a specific evolutionary algorithm to optimize routing in distributed time varying network for face recognition [6]. Chang and Aghajan focus on recovering face orientation for more robust face recognition in wireless image sensor networks [7]. Zaeri et al. propose application of face recognition for wireless surveillance systems [8].

As there exist many image variations such as pose, illumination, and facial expression, face recognition is a highly complex and nonlinear problem which could not be sufficiently handled by linear methods, such as principal components analysis (PCA) [9] and linear discriminant analysis (LDA) [10]. Therefore, it is reasonable to assume that a better solution to this inherent nonlinear problem could be achieved using nonlinear methods, such as the so-called kernel machine techniques [11]. Following the success of applying the kernel trick in support vector machines (SVMs) [12], many kernel-based PCA and LDA methods have been developed and applied in pattern recognition tasks, such as kernel PCA (KPCA) [13], kernel Fisher discriminant (KFD) [14], generalized discriminant analysis (GDA) [15], and kernel direct LDA (KDDA) [16].

It has been shown that the kernel-based LDA method is a feasible approach to solve the nonlinear problems in face recognition. However, the performance of the kernel-based LDA method is sensitive to the selection of a kernel function and its parameters. Kernel parameter selection to date can mainly be achieved by cross-validation [17], which is computationally expensive, and the selected kernel parameters cannot be guaranteed optimal. Furthermore, a single and fixed kernel can only characterize the geometrical

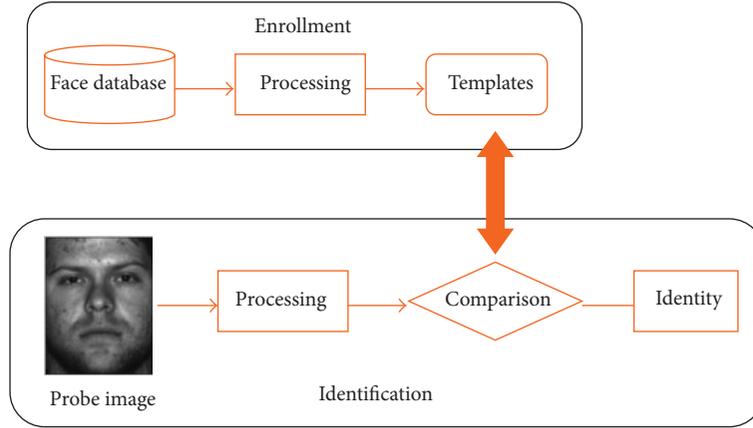


FIGURE 1: A normal framework of FR system.

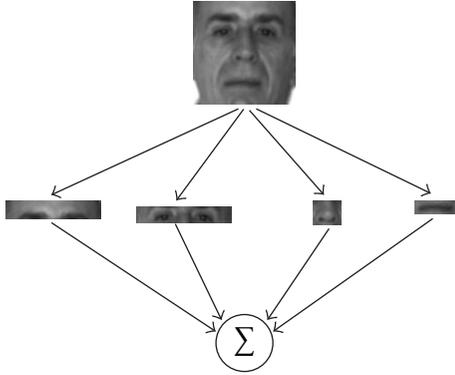


FIGURE 2: Submodules of a face image.

structure of some aspects for the input data, and, thus, not always be fit for the applications which involve data from multiple, heterogeneous sources [18, 19].

Recent applications and developments based on SVMs [20, 21] have shown that using multiple kernels (i.e., a combination of several “base kernels”) instead of a single fixed one can enhance classifier performance, which raised the so-called multiple kernel learning (MKL) method. With m kernels, input data can be mapped into m feature spaces, where each feature space can be taken as one view of the original input data [19]. Each view is expected to exhibit some geometrical structures of the original data from its own perspective such that all the views can complement for the subsequent learning task. It has been proven that MKL can offer some needed flexibility and well manipulate the case that involves multiple, heterogeneous data sources [18, 22, 23]. However, MKL is proposed for SVMs, and there have been few reports on performance of the kernel-based LDA method with multiple kernels. Liu and Feng propose multiple kernel Fisher discriminant analysis (MKFD) with an iterative scheme for weight optimization [24], in which the constructed kernel is a linear combination of several base kernels with a constraint on their weights.

In this paper, we integrate multiple kernel discriminant analysis with face recognition in wireless sensor networks, and propose a module based distributed wireless face recognition system. We consider separate cluster head for each module, that is, forehead, eye, lips, and nose. Only the local cluster is responsible for internal module processing for both training and recognition.

The rest of this paper is organized as follows. First we describe the module based distributed wireless face recognition system in Section 2. Then in Section 3, the optimization scheme for the multikernels is presented. The simulation results are reported in Section 4, while we draw our conclusion in Section 5.

2. Module Based Distributed Wireless Face Recognition

We present a face recognition system in wireless sensor networks where training and recognition are performed both in distributed environment. The image is divided into four submodules, that is, forehead, eyes, nose, and lips, as shown in Figure 2. For face recognition tasks, enrolment and identification of each submodule are performed in separate cluster heads, and the computations are carried out in kernel feature space [12]. Each cluster head is responsible for processing its submodule and communicating with the sink cluster which performs the score level fusion.

The following describes the score level fusion criterion at the sink node. Given N images belonging to C subjects. Denote the membership degree of the n th image belonging to the c th subject as μ_n^c , $n = 1, 2, \dots, N$, $c = 1, 2, \dots, C$. μ_n^c is obtained as follows:

$$\mu_n^c = \alpha_1 s_n^c(1) + \alpha_2 s_n^c(2) + \alpha_3 s_n^c(3) + \alpha_4 s_n^c(4), \quad (1)$$

where $s_n^c(k)$ denotes the score of the k th submodule from the n th image with regard to the c th subject, and α_k is the corresponding weight value, $k = 1, 2, 3, 4$. The n th image is assigned to the I th subject if and only if $\mu_n^I = \max_{1 \leq c \leq C} \mu_n^c$.

Now we explain score $s_n^c(k)$, taking the forehead module as an example. In the forehead module cluster, there are also

N samples belonging to C classes. For the n th sample \mathbf{x}_n , we can compute the squared kernel distance between \mathbf{x}_n and the center of the c th class \mathbf{m}_c as follows:

$$\begin{aligned} d_n^c &= \|\Phi(\mathbf{x}_n) - \Phi(\mathbf{m}_c)\|^2 \\ &= \Phi(\mathbf{x}_n)^T \Phi(\mathbf{x}_n) - 2\Phi(\mathbf{x}_n)^T \Phi(\mathbf{m}_c) + \Phi(\mathbf{m}_c)^T \Phi(\mathbf{m}_c) \\ &= k(\mathbf{x}_n, \mathbf{x}_n) - 2k(\mathbf{x}_n, \mathbf{m}_c) + k(\mathbf{m}_c, \mathbf{m}_c), \end{aligned} \quad (2)$$

where $\Phi : \mathbf{x} \in \mathbf{R}^t \rightarrow \Phi(\mathbf{x}) \in \mathbf{F}$ is a nonlinear mapping, which is implicitly defined by a Mercer kernel function $k(\mathbf{x}, \mathbf{y}) = \Phi(\mathbf{x})^T \Phi(\mathbf{y})$ [12, 25]. Then we sort the C distances $\{d_n^c \mid c = 1, 2, \dots, C\}$ in ascending order and denote the squared kernel distances after sorting as $\{d_n^{i(c)} \mid c = 1, 2, \dots, C\}$; that is, $i(c)$ is the order number of the c th largest squared kernel distance in $\{d_n^c \mid c = 1, 2, \dots, C\}$. Let

$$s_n^{i(c)}(1) = C - c, \quad (3)$$

$c = 1, 2, \dots, C$, then we get $\{s_n^c(1) \mid c = 1, 2, \dots, C\}$. Thus, given the n th sample, the smaller the kernel distance between it and the center of the c th class, the greater $s_n^c(1)$ is, $c = 1, 2, \dots, C$.

Scores $s_n^c(2) \sim s_n^c(4)$ can be obtained from the eyes, nose, and lips modules, respectively, using a similar way to score $s_n^c(1)$.

3. Optimization of Multiple Kernels

From Section 2, we can see the computations in the proposed distributed wireless face recognition frame are based on kernel techniques. The framework has four computing modules, for forehead, eyes, nose, and lips, respectively. These four kinds of data are so different that it is hard to imagine that an excellent classification can be reached by a single kernel. We propose the use of multiple kernels for these four modules. Specifically, we use the Gaussian RBF kernel

$$k(\mathbf{x}, \mathbf{y}) = \exp\left(-\frac{\|\mathbf{x} - \mathbf{y}\|^2}{\sigma^2}\right), \quad (4)$$

but with different values of parameter σ for the different computing modules.

Every module should have its own optimal kernel parameter. For each module, we separately perform the following kernel parameter optimization procedure.

3.1. Some Notations on Kernel Discriminant Analysis. For a certain module X , there are N samples belonging to C classes. Assume the i th class X_i contains N_i samples; that is, $X_i = \{\mathbf{x}_1^i, \mathbf{x}_2^i, \dots, \mathbf{x}_{N_i}^i\}$, $i = 1, 2, \dots, C$, so $N = \sum_{i=1}^C N_i$. Denote the kernel matrix ($N \times N$) as

$$\mathbf{K} = [k(\mathbf{x}_j^i, \mathbf{x}_h^l)]_{\substack{i=1, \dots, C, j=1, \dots, N_i, \\ l=1, \dots, C, h=1, \dots, N_l}}, \quad (5)$$

where $k(\mathbf{x}_j^i, \mathbf{x}_h^l) = \Phi(\mathbf{x}_j^i)^T \Phi(\mathbf{x}_h^l)$; then the kernel matrix corresponds to the nonlinear mapping

Under the nonlinear mapping Φ , the i th mapped class and the mapped sample set are, respectively, given by

$$\begin{aligned} \Phi(X_i) &= \{\Phi(\mathbf{x}_1^i), \Phi(\mathbf{x}_2^i), \dots, \Phi(\mathbf{x}_{N_i}^i)\}, \\ \Phi(X) &= \{\Phi(X_1), \Phi(X_2), \dots, \Phi(X_C)\}. \end{aligned} \quad (6)$$

Also, the mean of the mapped class $\Phi(X_i)$ and that of the mapped sample set $\Phi(X)$ are, respectively, given by

$$\mathbf{m}_i = \frac{1}{N_i} \sum_{j=1}^{N_i} \Phi(\mathbf{x}_j^i), \quad \mathbf{m} = \frac{1}{N} \sum_{i=1}^C \sum_{j=1}^{N_i} \Phi(\mathbf{x}_j^i). \quad (7)$$

In kernel feature space \mathbf{F} (let f be dimensionality of \mathbf{F}), the within-class scatter matrix \mathbf{S}_w^Φ and between-class scatter matrix \mathbf{S}_b^Φ are, respectively, defined as

$$\mathbf{S}_w^\Phi = \frac{1}{N} \sum_{i=1}^C \sum_{\mathbf{x} \in X_i} (\Phi(\mathbf{x}) - \mathbf{m}_i) (\Phi(\mathbf{x}) - \mathbf{m}_i)^T = \Phi_w \Phi_w^T, \quad (8)$$

$$\mathbf{S}_b^\Phi = \frac{1}{N} \sum_{i=1}^C N_i (\mathbf{m}_i - \mathbf{m}) (\mathbf{m}_i - \mathbf{m})^T = \Phi_b \Phi_b^T,$$

where

$$\Phi_w = [\phi_1^1, \dots, \phi_{N_1}^1, \phi_1^2, \dots, \phi_{N_2}^2, \dots, \phi_1^C, \dots, \phi_{N_C}^C]_{df \times N},$$

$$\phi_j^i = \frac{1}{\sqrt{N}} (\Phi(\mathbf{x}_j^i) - \mathbf{m}_i),$$

$$\Phi_b = [\varphi_1, \dots, \varphi_C]_{df \times C}, \quad \varphi_i = \sqrt{\frac{N_i}{N}} (\mathbf{m}_i - \mathbf{m}). \quad (9)$$

The kernel Fisher criterion is defined as

$$J^\Phi(\mathbf{W}) = \frac{\text{tr}(\mathbf{W}^T \mathbf{S}_b^\Phi \mathbf{W})}{\text{tr}(\mathbf{W}^T \mathbf{S}_w^\Phi \mathbf{W})}, \quad (10)$$

where $\mathbf{W} = \{\mathbf{w}_1, \dots, \mathbf{w}_q\}$ is a $f \times q$ ($f > q$) projection matrix. Kernel discriminant analysis is to find an optimal projection matrix $\mathbf{W}^* : \mathbf{R}^{df} \rightarrow \mathbf{R}^q$ in mapped feature space \mathbf{F} , such that $\mathbf{W}^* = \text{argmax}_{\mathbf{W}} J^\Phi(\mathbf{W})$.

3.2. Diagonalization Strategy. We use the same diagonalization strategy as KDDA [16] to deal with the small sample size (SSS) problem; that is, first diagonalizing \mathbf{S}_b^Φ to \mathbf{I} (identical matrix) and then diagonalizing \mathbf{S}_w^Φ to Λ_w , which is briefly expressed as follows.

3.2.1. Eigenanalysis of \mathbf{S}_b^Φ in the Feature Space. $\Phi_b^T \Phi_b$ can be expressed using the kernel matrix \mathbf{K} as follows:

$$\begin{aligned} \Phi_b^T \Phi_b &= \frac{1}{N} \mathbf{D} \cdot \left(\mathbf{A}_{NC}^T \cdot \mathbf{K} \cdot \mathbf{A}_{NC} - \frac{1}{N} \mathbf{A}_{NC}^T \cdot \mathbf{K} \cdot \mathbf{1}_{NC} \right. \\ &\quad \left. - \frac{1}{N} \mathbf{1}_{NC}^T \cdot \mathbf{K} \cdot \mathbf{A}_{NC} + \frac{1}{N^2} \mathbf{1}_{NC}^T \cdot \mathbf{K} \cdot \mathbf{1}_{NC} \right) \cdot \mathbf{D}, \end{aligned} \quad (11)$$

where $\mathbf{D} = \text{diag}(\sqrt{N_1}, \dots, \sqrt{N_C})$ ($C \times C$ diagonal matrix), $\mathbf{1}_{NC}$ is a $N \times C$ matrix with terms all equal to one, $\mathbf{A}_{NC} = \text{diag}(\mathbf{a}_{N_1}, \dots, \mathbf{a}_{N_C})$ is a $N \times C$ block diagonal matrix, and \mathbf{a}_{N_i} is a $N_i \times 1$ vector with all terms equal to $1/N_i$.

Let λ_i and \mathbf{e}_i ($i = 1, \dots, C$) be the i th largest eigenvalue and corresponding eigenvector of $\Phi_b^T \Phi_b$. Let $r (\leq C-1)$ be the rank of $\mathbf{S}_b^\Phi (= \Phi_b \Phi_b^T)$ (also the rank of $\Phi_b^T \Phi_b$). Denote $\mathbf{E}_r = (\mathbf{e}_1, \dots, \mathbf{e}_r)$ and $\mathbf{V} = (\mathbf{v}_1, \dots, \mathbf{v}_r) = \Phi_b \mathbf{E}_r$. It can be derived that $\mathbf{V}^T \mathbf{S}_b^\Phi \mathbf{V} = \Lambda_b$, with $\Lambda_b = \text{diag}(\lambda_1^2, \dots, \lambda_r^2)$, a nonsingular diagonal matrix. Let $\mathbf{U} = \mathbf{V} \Lambda_b^{-1/2}$. Then $\mathbf{U}^T \mathbf{S}_b^\Phi \mathbf{U} = \mathbf{I}$.

3.2.2. Eigenanalysis of \mathbf{S}_w^Φ in the Feature Space. Based on the analysis in Section 3.2.1, it can be seen that

$$\mathbf{U}^T \mathbf{S}_w^\Phi \mathbf{U} = (\mathbf{E}_r \Lambda_b^{-1/2})^T (\Phi_b^T \mathbf{S}_w^\Phi \Phi_b) (\mathbf{E}_r \Lambda_b^{-1/2}), \quad (12)$$

where $\Phi_b^T \mathbf{S}_w^\Phi \Phi_b$ can be expressed using \mathbf{K} , with similar details to that seen in [16].

Let \mathbf{z}_j be the eigenvector of $\mathbf{U}^T \mathbf{S}_w^\Phi \mathbf{U}$ corresponding to the j th smallest eigenvalue λ'_j , $j = 1, \dots, r$. Denote $\mathbf{Z} = (\mathbf{z}_1, \dots, \mathbf{z}_r)$. Defining $\mathbf{Y} = \mathbf{U} \mathbf{Z}$, it can be derived that $\mathbf{Y}^T \mathbf{S}_w^\Phi \mathbf{Y} = \Lambda_w$, with $\Lambda_w = \text{diag}(\lambda'_1, \dots, \lambda'_r)$.

Based on the derivation presented in Sections 3.2.1 and 3.2.2, an optimal projection matrix for kernel discriminant analysis is obtained as

$$\mathbf{W}^* = \mathbf{Y} \Lambda_w^{-1/2} = \Phi_b \mathbf{E}_r \Lambda_b^{-1/2} \mathbf{Z} \Lambda_w^{-1/2}. \quad (13)$$

Certainly, as the nonlinear mapping Φ is implicitly defined by the kernel function (or matrix), Φ_b (defined by (9)) remains unknown, and \mathbf{W}^* can not be evaluated. The real meaning of (13) is obtaining matrix $\mathbf{E}_r \Lambda_b^{-1/2} \mathbf{Z} \Lambda_w^{-1/2}$, which can be computed from the kernel matrix \mathbf{K} . This is the core result of diagonalization process.

3.3. Optimization Criterion and Objective. We adopt the maximum margin criterion (MMC) [26] as the objective function to optimize the kernel parameter σ for each specific module (module of forehead, eyes, nose, or lips):

$$F(\mathbf{W}, \sigma) = \text{tr}(\mathbf{W}^T \mathbf{S}_b^\Phi \mathbf{W}) - \text{tr}(\mathbf{W}^T \mathbf{S}_w^\Phi \mathbf{W}), \quad (14)$$

where \mathbf{W} is a projection matrix, and σ is the parameter for the Gaussian RBF kernel as in (4).

Based on the result of (13) in Section 3.2, the optimal projection matrix $\mathbf{W}^* = \Phi_b \mathbf{E}_r \Lambda_b^{-1/2} \mathbf{Z} \Lambda_w^{-1/2}$. Denoting $\mathbf{G} = \mathbf{E}_r \Lambda_b^{-1/2} \mathbf{Z} \Lambda_w^{-1/2}$, which can be computed from the kernel matrix \mathbf{K} . Then the objective function (14) can be reformulated as

$$\begin{aligned} F(\sigma) &= \text{tr}(\mathbf{W}^{*T} \mathbf{S}_b^\Phi \mathbf{W}^* - \mathbf{W}^{*T} \mathbf{S}_w^\Phi \mathbf{W}^*) \\ &= \text{tr}(\mathbf{G}^T \Phi_b^T \Phi_b \Phi_b^T \Phi_b \mathbf{G} - \mathbf{G}^T \Phi_b^T \Phi_w \Phi_w^T \Phi_b \mathbf{G}) \\ &= \text{tr}(\mathbf{G}^T \mathbf{P} \mathbf{P}^T \mathbf{G} - \mathbf{G}^T \mathbf{Q} \mathbf{Q}^T \mathbf{G}), \end{aligned} \quad (15)$$

where $\mathbf{P} = \Phi_b^T \Phi_b$ and $\mathbf{Q} = \Phi_b^T \Phi_w$ can be expressed in terms of the kernel matrix \mathbf{K} as follows:

$$\begin{aligned} \mathbf{P} &= \frac{1}{N} \mathbf{D} \cdot \left(\mathbf{A}_{NC}^T \cdot \mathbf{K} \cdot \mathbf{A}_{NC} - \frac{1}{N} \mathbf{A}_{NC}^T \cdot \mathbf{K} \cdot \mathbf{1}_{NC} \right. \\ &\quad \left. - \frac{1}{N} \mathbf{1}_{NC}^T \cdot \mathbf{K} \cdot \mathbf{A}_{NC} + \frac{1}{N^2} \mathbf{1}_{NC}^T \cdot \mathbf{K} \cdot \mathbf{1}_{NC} \right) \cdot \mathbf{D}, \end{aligned} \quad (16)$$

with \mathbf{D} , \mathbf{A}_{NC} , and $\mathbf{1}_{NC}$ defined the same as in (11);

$$\begin{aligned} \mathbf{Q} &= \frac{1}{N} \mathbf{D} \cdot \left(\mathbf{A}_{NC}^T \cdot \mathbf{K} - \mathbf{A}_{NC}^T \cdot \mathbf{K} \cdot \mathbf{H}_{NN} \right. \\ &\quad \left. - \frac{1}{N} \mathbf{1}_{NC}^T \cdot \mathbf{K} + \frac{1}{N} \mathbf{1}_{NC}^T \cdot \mathbf{K} \cdot \mathbf{H}_{NN} \right), \end{aligned} \quad (17)$$

where $\mathbf{H}_{NN} = \text{diag}(\mathbf{h}_{N_1}, \dots, \mathbf{h}_{N_C})$ is a $N \times N$ block diagonal matrix and \mathbf{h}_{N_i} is a $N_i \times N_i$ matrix with all terms equal to $1/N_i$.

Therefore, regarding \mathbf{G} as constant, the objective function $F(\sigma)$ is an explicit function of the kernel parameter σ . To maximize the objective function with σ_0 as the initial value of the parameter, an iterative procedure based on Newton's method is developed in our method to update the kernel parameter, as shown in the following section.

3.4. Solving the Optimization Problem. Assume that \mathbf{G} is constant. To obtain the extremum of the objective function $F(\sigma)$, we need to differentiate $F(\sigma)$ with respect to σ .

For the sake of clarity, we denote the differentiation of the kernel matrix \mathbf{K} with respect to σ as $d\mathbf{K}_\sigma$, which is an $N \times N$ matrix expressed as follows:

$$d\mathbf{K}_\sigma = \left[\frac{dk(\mathbf{x}_j^i, \mathbf{x}_h^l)}{d\sigma} \right]_{\substack{i=1, \dots, C, j=1, \dots, N_i \\ l=1, \dots, C, h=1, \dots, N_i}}. \quad (18)$$

Each element in matrix $d\mathbf{K}_\sigma$ is the differentiation of the corresponding element in the kernel matrix \mathbf{K} with respect to σ and can be formulated as

$$\begin{aligned} \frac{dk(\mathbf{x}_j^i, \mathbf{x}_h^l)}{d\sigma} &= \frac{d}{d\sigma} \exp\left(-\frac{\|\mathbf{x}_j^i - \mathbf{x}_h^l\|^2}{\sigma^2}\right) \\ &= \frac{2\|\mathbf{x}_j^i - \mathbf{x}_h^l\|^2}{\sigma^3} \exp\left(-\frac{\|\mathbf{x}_j^i - \mathbf{x}_h^l\|^2}{\sigma^2}\right). \end{aligned} \quad (19)$$

From (16) and (17), matrices \mathbf{P} and \mathbf{Q} can be differentiated with respect to σ as follows:

$$\begin{aligned} \frac{d\mathbf{P}}{d\sigma} &= \frac{1}{N} \mathbf{D} \cdot \left(\mathbf{A}_{NC}^T \cdot d\mathbf{K}_\sigma \cdot \mathbf{A}_{NC} - \frac{1}{N} \mathbf{A}_{NC}^T \cdot d\mathbf{K}_\sigma \cdot \mathbf{1}_{NC} \right. \\ &\quad \left. - \frac{1}{N} \mathbf{1}_{NC}^T \cdot d\mathbf{K}_\sigma \cdot \mathbf{A}_{NC} + \frac{1}{N^2} \mathbf{1}_{NC}^T \cdot d\mathbf{K}_\sigma \cdot \mathbf{1}_{NC} \right) \\ &\quad \cdot \mathbf{D}, \end{aligned}$$



FIGURE 3: Images of one person from the FERET database.

$$\begin{aligned} \frac{d\mathbf{Q}}{d\sigma} = \frac{1}{N} \mathbf{D} \cdot \left(\mathbf{A}_{NC}^T \cdot d\mathbf{K}_\sigma - \mathbf{A}_{NC}^T \cdot d\mathbf{K}_\sigma \cdot \mathbf{H}_{NN} \right. \\ \left. - \frac{1}{N} \mathbf{1}_{NC}^T \cdot d\mathbf{K}_\sigma + \frac{1}{N} \mathbf{1}_{NC}^T \cdot d\mathbf{K}_\sigma \cdot \mathbf{H}_{NN} \right), \end{aligned} \quad (20)$$

where $d\mathbf{P}/d\sigma$ is a $C \times C$ matrix and $d\mathbf{Q}/d\sigma$ is a $C \times N$ matrix.

Then, the derivative of F with respect to σ can be formulated as

$$\begin{aligned} F'(\sigma) = \text{tr} \left(\left(\mathbf{G}^T \frac{d\mathbf{P}}{d\sigma} \mathbf{P}^T \mathbf{G} + \mathbf{G}^T \mathbf{P} \frac{d\mathbf{P}^T}{d\sigma} \mathbf{G} \right) \right. \\ \left. - \left(\mathbf{G}^T \frac{d\mathbf{Q}}{d\sigma} \mathbf{Q}^T \mathbf{G} + \mathbf{G}^T \mathbf{Q} \frac{d\mathbf{Q}^T}{d\sigma} \mathbf{G} \right) \right). \end{aligned} \quad (21)$$

Thus the derivative of F with respect to σ can be expressed in terms of matrix \mathbf{K} and matrix $d\mathbf{K}_\sigma$. To achieve the maximum of $F(\sigma)$, we set the derivative to zero:

$$F'(\sigma) = 0. \quad (22)$$

We use Newton's method to solve (22) with the initial value of the kernel parameter

$$\sigma_0 = \sqrt{\frac{1}{N^2} \sum_{i=1}^N \sum_{j=1}^N \|\mathbf{x}_i - \mathbf{x}_j\|^2}; \quad (23)$$

that is, σ_0^2 is the average squared distance of all the samples in the given module. And the iteration formula is

$$\sigma_{k+1} = \sigma_k - \frac{F(\sigma_k)}{F'(\sigma_k)}. \quad (24)$$

$k = 0, 1, 2, \dots$

3.5. Identification Using Multiple Kernels. For each of the four modules (forehead, eyes, nose, and lips), the optimization of the Gaussian kernel parameter runs and finds the optimal value as above separately. Then the four submodules of a testing image are fed into the corresponding kernel discriminant classifiers to compute the membership degree of the image belonging to the every subject according to (1). Finally the image is assigned to the subject which shows the greatest membership degree of the image.

4. Simulation

To evaluate the performance of our multiple kernel framework for distributed wireless face recognition, we have made



FIGURE 4: Some images of one person from the CMU PIE face database.

experimental comparisons with KDDA based on Gaussian RBF kernel, in terms of recognition accuracy. Images are from two face databases, namely, the FERET and the CMU PIE databases.

In our experiments, for the weight value in the fusion criterion (1), we set $\alpha_1 = 0.1$, $\alpha_2 = 0.3$, $\alpha_3 = 0.4$, and $\alpha_4 = 0.2$, meaning that we give larger weights to the nose module and the eyes module.

4.1. Face Image Datasets. From the FERET database [27], we select 72 people, with 6 frontal-view images for each individual. Face image variations in these 432 images include illumination, facial expression, wearing glasses, and aging. All the images are aligned by the centers of the eyes and the mouth and then normalized with a resolution of 92×112 . The pixel value of each image is normalized between 0 and 1. The original images with resolution 92×112 are reduced to wavelet feature faces with resolution 49×59 after 1-level Daubechies-4 (Db4) wavelet decomposition. Images from one individual are shown in Figure 3.

In the CMU PIE face database [28], there are a total of 68 people, and each person has 13 pose variations ranged from the full right profile image to the full left profile image and 43 different lighting conditions, 21 flashes with ambient light on or off. In our experiments, for each person, we select 56 images including 13 poses with neutral expression and 43 different lighting conditions in the frontal view. For all frontal-view images, we apply alignment based on two-eye center and nose center points, and no alignment is applied on the other images with poses. All the segmented images are rescaled to the resolution of 92×112 , and then reduced to wavelet feature faces with resolution 49×59 after 1-level Daubechies-4 (Db4) wavelet decomposition. Some images of one person are shown in Figure 4.

4.2. Recognition Results. This section reports the recognition results of the proposed multiple kernel framework and KDDA with a single Gaussian RBF kernel on the FERET and the CMU PIE datasets. For KDDA, the parameter of Gaussian RBF kernel is optimized via grid search. For each subject in the FERET dataset, we randomly select n ($n = 2$ to 5) out

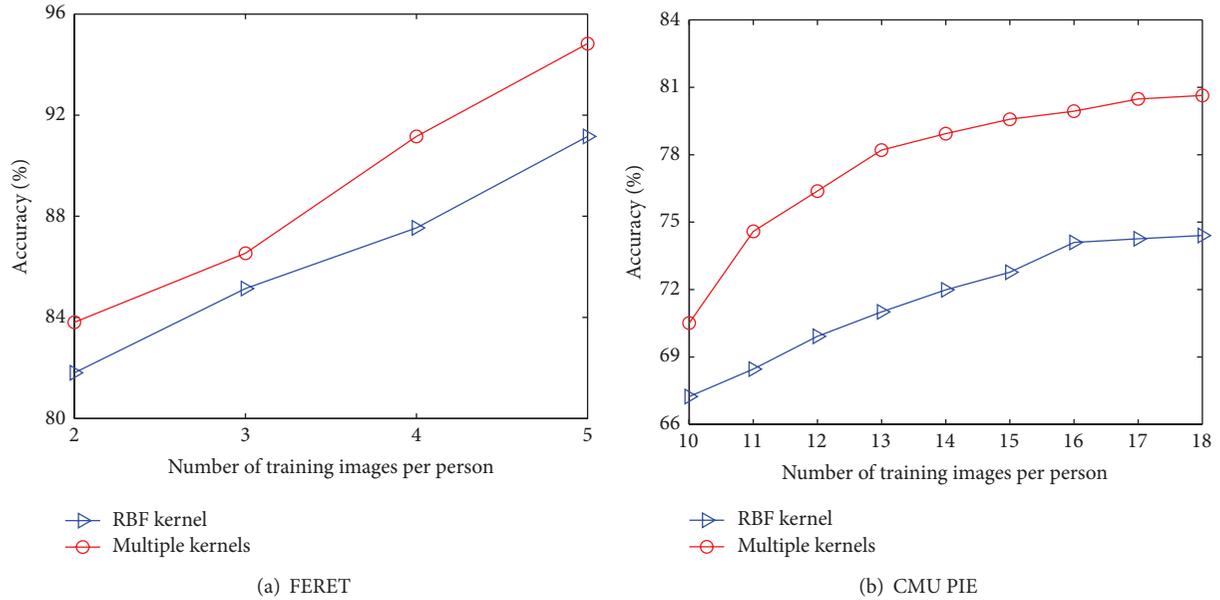


FIGURE 5: Comparison of accuracies obtained by multiple kernel framework and KDDA with a single kernel.

TABLE 1: Performance comparison between multiple kernel framework and KDDA with a single kernel.

Type of kernel	FERET ($n = 4$)	CMU PIE ($n = 14$)
RBF kernel		
Mean	87.54%	71.99%
Std	0.042	0.013
Multi kernels		
Mean	91.16%	78.96%
Std	0.011	0.006

The bold value means the higher average accuracy of the two methods.

of 6 images for training, with the rest for testing. In the CMU PIE dataset, the number of randomly selected training images is ranged from 10 to 18 out of 56 for each individual, while the rest are testing images. The average recognition accuracies over 10 runs on the FERET and CMU PIE datasets are shown in Figures 5(a) and 5(b), respectively.

Table 1 shows the average and standard deviation of the accuracies for FERET ($n = 4$: 4 images per subject for training with the rest for testing) and CMU-PIE ($n = 14$: 14 images per subject for training with the rest for testing), respectively.

From the results in Figure 5 and Table 1, it can be seen that the proposed multiple kernel framework can achieve higher accuracies than KDDA with an optimized parameter.

5. Conclusion

In this paper, on the assumption that multiple kernels can characterize geometrical structures of the original data from multiple views which can complement to improve recognition performance, we integrate multiple kernel discriminant analysis with face recognition in wireless sensor networks and propose a module based distributed wireless face recognition

system. For each module, we separately perform an iterative scheme based on Newton's method for kernel parameter optimization, by maximizing the margin maximization criterion. The multiple kernel framework and the optimization procedure yield high recognition accuracy on the FERET and CMU PIE face database, compared with a single kernel.

Conflict of Interests

The authors declare that they have no conflict of interests regarding the publication of this paper.

Acknowledgments

This work is partially supported by the Basic and Frontier Technology Research Project of Henan Province in China under Grant no. 122300410321 and Science and Technology Development Project of Henan Province under Grant no. 132102210186.

References

- [1] W. Zhao, R. Chellappa, P. J. Phillips, and A. Rosenfeld, "Face recognition: a literature survey," *ACM Computing Surveys*, vol. 35, no. 4, pp. 399–458, 2003.
- [2] X. Zhang and Y. Gao, "Face recognition across pose: a review," *Pattern Recognition*, vol. 42, no. 11, pp. 2876–2896, 2009.
- [3] Q.-M. Lin, J.-W. Yang, N. Ye, R.-C. Wang, and B. Zhang, "Face recognition in mobile wireless sensor networks," *International Journal of Distributed Sensor Networks*, vol. 2013, Article ID 890737, 7 pages, 2013.
- [4] I. F. Akyildiz, W. Su, Y. Sankarasubramaniam, and E. Cayirci, "Wireless sensor networks: a survey," *Computer Networks*, vol. 38, no. 4, pp. 393–422, 2002.

- [5] I. Kim, J. Shim, J. Schlessman, and W. Wolf, "Remote wireless face recognition employing zigbee," in *Proceedings of the ACM SenSys Workshop on Distributed Smart Cameras (DSC '06)*, Boulder, Colo, USA, 2006.
- [6] R. Muraleedharan, Y. Yan, and L. A. Osadciw, "Increased efficiency of face recognition system using wireless sensor network," *Systemics, Cybernetics and Informatics*, vol. 4, no. 1, pp. 38–46, 2006.
- [7] C. C. Chang and H. Aghajan, "Collaborative face orientation detection in wireless image sensor networks," in *Proceedings of the ACM SenSys Workshop on Distributed Smart Cameras (DSC '06)*, Boulder, Colo, USA, 2006.
- [8] N. Zaeri, F. Mokhtarian, and A. Cherri, "Efficient face recognition for wireless surveillance systems," in *Proceedings of the 9th IASTED International Conference on Computer Graphics and Imaging (CGIM '07)*, E. Gobbetti, Ed., pp. 132–137, OACTA Press, Innsbruck, Austria, February 2007.
- [9] M. Turk and A. Pentland, "Eigenfaces for recognition," *Journal of Cognitive Neuroscience*, vol. 3, no. 1, pp. 71–86, 1991.
- [10] P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman, "Eigenfaces versus fisherfaces: recognition using class specific linear projection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19, no. 7, pp. 711–720, 1997.
- [11] A. Ruiz and P. E. López-de-Teruel, "Nonlinear kernel-based statistical pattern analysis," *IEEE Transactions on Neural Networks*, vol. 12, no. 1, pp. 16–32, 2001.
- [12] V. N. Vapnik, *The Nature of Statistical Learning Theory*, Springer, New York, NY, USA, 1995.
- [13] B. Schölkopf, A. Smola, and K. Müller, "Nonlinear component analysis as a kernel eigenvalue problem," Tech. Rep. 44, MPI für Biologische Kybernetik, Tübingen, Germany, 1996.
- [14] S. Mika, G. Rätsch, J. Weston, B. Schölkopf, and K. R. Müller, "Fisher discriminant analysis with kernels," in *Proceedings of the IEEE Signal Processing Society Workshop, Neural Networks for Signal Processing IX*, pp. 41–48, Madison, Wis, USA, 1999.
- [15] G. Baudat and F. Anouar, "Generalized discriminant analysis using a kernel approach," *Neural Computation*, vol. 12, no. 10, pp. 2385–2404, 2000.
- [16] J. Lu, K. N. Plataniotis, and A. N. Venetsanopoulos, "Face recognition using kernel direct discriminant analysis algorithms," *IEEE Transactions on Neural Networks*, vol. 14, no. 1, pp. 117–126, 2003.
- [17] O. Chapelle, V. Vapnik, O. Bousquet, and S. Mukherjee, "Choosing multiple parameters for support vector machines," *Machine Learning*, vol. 46, no. 1–3, pp. 131–159, 2002.
- [18] S. Sonnenburg, G. Rätsch, and C. Schäfer, "A general and efficient multiple kernel learning algorithm," in *Proceedings of the Neural Information Processing Systems*, 2005.
- [19] Z. Wang, S. Chen, and T. Sun, "MultiK-MHKS: a novel multiple kernel learning algorithm," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 30, no. 2, pp. 348–353, 2008.
- [20] J. Bi, T. Zhang, and K. P. Bennett, "Column-generation boosting methods for mixture of kernels," in *Proceedings of the 10th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, (KDD '04)*, pp. 521–526, New York, NY, USA, August 2004.
- [21] G. R. G. Lanckriet, N. Cristianini, P. Bartlett, L. El Ghaoui, and M. I. Jordan, "Learning the kernel matrix with semidefinite programming," *Journal of Machine Learning Research*, vol. 5, pp. 27–72, 2004.
- [22] F. R. Bach, G. R. G. Lanckriet, and M. I. Jordan, "Multiple kernel learning, conic duality, and the SMO algorithm," in *Proceedings of the 21th International Conference on Machine Learning, (ICML '04)*, pp. 41–48, New York, NY, USA, July 2004.
- [23] K. P. Bennett, M. Momma, and M. J. Embrechts, "MARK: a boosting algorithm for heterogeneous kernel models," in *Proceedings of the 8th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, (KDD '02)*, pp. 24–31, New York, NY, USA, July 2002.
- [24] X.-Z. Liu and G.-C. Feng, "Multiple kernel learning in fisher discriminant analysis for face recognition," *International Journal of Advanced Robotic Systems*, vol. 10, article 142, 2013.
- [25] J. Mercer, *Functions of Positive and Negative Type and Their Connection with the Theory of Integral Equations*, vol. 209 of *Philosophical Transactions of the Royal Society of London Series A*, The Royal Society, 1909.
- [26] H. Li, T. Jiang, and K. Zhang, "Efficient and robust feature extraction by maximum margin criterion," in *Advances in Neural Information Processing Systems 16*, S. Thrun, L. Saul, and B. Schölkopf, Eds., pp. 157–1165, The MIT Press, Cambridge, Mass, USA, 2004.
- [27] P. Jonathon Phillips, H. Moon, S. A. Rizvi, and P. J. Rauss, "The FERET evaluation methodology for face-recognition algorithms," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 10, pp. 1090–1104, 2000.
- [28] T. Sim, S. Baker, and M. Bsat, "The CMU pose, illumination, and expression (PIE) database," in *Proceedings of the 5th IEEE International Conference on Automatic Face and Gesture Recognition*, Washington, DC, USA, May 2002.

