Available online at www.sciencedirect.com
ScienceDirect
Pattern Recognition 41 (2008) 805–813
PATTERN RECOGNITION
THE JOURNAL OF THE PATTERN RECOGNITION SOCIETY
www.elsevier.com/locate/pr

Multimodal biometrics using geometry preserving projections

Tianhao Zhang[a,c], Xuelong Li[b], Dacheng Tao[c,*], Jie Yang[a]

[a]*Institute of Image Processing and Pattern Recognition, Shanghai Jiao Tong University, Shanghai 200240, PR China*
[b]*School of Computer Science and Information Systems, Birkbeck College, University of London, London WC1E 7HX, UK*
[c]*Biometrics Research Centre, Department of Computing, The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong*

Received 30 April 2007; accepted 26 June 2007

Abstract

Multimodal biometric system utilizes two or more individual modalities, e.g., face, gait, and fingerprint, to improve the recognition accuracy of conventional unimodal methods. However, existing multimodal biometric methods neglect interactions of different modalities during the subspace selection procedure, i.e., the underlying assumption is the independence of different modalities. In this paper, by breaking this assumption, we propose a Geometry Preserving Projections (GPP) approach for subspace selection, which is capable of discriminating different classes and preserving the intra-modal geometry of samples within an identical class. With GPP, we can project all raw biometric data from different identities and modalities onto a unified subspace, on which classification can be performed. Furthermore, the training stage is carried out once and we have a unified transformation matrix to project different modalities. Unlike existing multimodal biometric systems, the new system works well when some modalities are not available. Experimental results demonstrate the effectiveness of the proposed GPP for individual recognition tasks.

Keywords: Multimodal biometrics; Geometry preserving projections; Subspace selection

1. Introduction

Biometrics has drawn extensive attention during the past decades for its huge potentials in many applications e.g., information security [1], surveillance [2], and human computer interface [3]. The key issue of these applications is the identification of individuals by their physiological or behavioral characteristics, including face [4], fingerprint [5], palmprint [6], iris [7], speech [8], hand geometry [9], and gait [10]. Based on one of modalities mentioned above, one can construct the unimodal biometrics.

However, recent researches revealed that: in biometrics recognition tasks, utilizing a single source suffers from various problems such as poor robustness, small sample size problem [11], and spoof attacks [12]. To alleviate some of these problems, a number of multimodal methods, which utilize two or more individual modalities, have been developed. Hong et al. [13] developed a prototype multimodal biometric system, which integrates faces and fingerprints at the identification stage. The decision-fusion scheme improves performance by integrating multiple cues with different confidence measures. Ribaric and Fratric [14] presented a multimodal biometric system based on features extracted from fingerprint and palmprint data. Fusion is applied at the matching-score level after extracting the features by using PCA projection. Jain et al. [15] studied the performance of different score normalization techniques and fusion rules in the context of a multimodal biometric system based on face, fingerprint and hand-geometry modalities. A reduced multivariate polynomial model was introduced by Toh et al. [16] to overcome the tedious recursive learning problem in multimodal biometrics in order to achieve good decision accuracy. Gao et al. [17] investigated the fusion at the feature level to combine multiple views of face and palmprint data in personal identification. Multiview line segment Hausdorff distance has also been proposed to integrate the homogeneous lines from different sensors. Jin et al. [11] combined the face images and palmprint images vertically after applying Gabor

* Corresponding author. Tel.: +852 2766 7266.
E-mail address: dacheng.tao@gmail.com (D. Tao).

0031-3203/$30.00
doi:10.1016/j.patcog.2007.06.035

transform on these data. Yao et al. [18] designed a distance-based separability weighting strategy to conduct feature level fusion after using Gabor-based image preprocessing and PCA techniques. Sim et al. [19] proposed a continuous multimodal biometric system, which combines face and fingerprint modalities at the score level. In here, holistic fusion is utilized by exploiting Hidden Markov Model in a Bayesian framework.

Most of the above methods are designed by essentially information fusion. Their models are basically in three categories [13,20]: fusion at *feature* level [11,17,18], fusion at *match score* level [14,15,19] and fusion at *decision* level [13,16]. Fusion at feature level is implemented by concatenating two or more features to form a long vector. This category of modes is not practical because the features of the various modalities may not be compatible, for example, face images normally have larger sizes than those of finger imagers. Moreover, in this type of fusion modes, the recognition system does not work if one or more modalities of testing samples are not available. For fusion at match score level, matching scores are generated by different classifiers and then combined. Finally, for fusion at decision level, modes consolidate final decisions based on different modalities for a final decision. In real practice, fusion at match score level and decision level are usually employed since they are much more practical, but in these modes, the useful information has never been exploited for fusion before match and decision.

Recently, subspace methods, which select low-dimensional features to represent raw data, have been widely studied in biometrics researches [21–24]. It is shown that subspace selection is one of the most important steps for entire biometric systems. Nevertheless, as far as we know, most of the existing multimodal biometric methods seldom focus on the interaction of the different modalities in subspace selection. These methods which carry out fusion at feature level obtain the low-dimensional features integrated the information of all modalities. However, they can not get the low-dimensional representations if one or more modalities are missing. The methods whose fusions are at match score level and decision level never consider the fusion in the subspace selection as they may extract the low-dimensional features and yield the different maps, respectively, for each modality.

In this paper, we propose a new system for multimodal biometric recognition. In the developed system, *Geometry Preserving Projections* (GPP), as a new subspace selection approach (especially for the multimodal problem), is developed. GPP is a linear discriminant algorithm, and it effectively preserves local information by capturing the intra-modal geometry. To overcome the non-linear problems in real world and further improve the performance, its non-linear variant, *kernel GPP*(KGPP), is also developed by applying the kernel trick [25,26]. By using GPP, all the raw biometric data from various modalities can be trained to obtain a unified subspace, and classification can be performed in the obtained subspace. In the given system, the training stage is carried out only once and we have the unified transformation matrix for all the multimodal data.

The rest of this paper is organized as follows: Section 2 introduces the proposed system and the GPP algorithm. The classification in the subspace will be described in Section 3. Then, in Section 4, the experiments results are reported to evaluate our algorithm. Finally, the conclusions are given in Section 5. For convenience, Table 1 lists the important notations used in the rest of the paper.

Table 1
Important notations used in the paper

X	Training set	θ	Scaling factor
Y	Dimension-reduced training set	k	Number of neighbors
P	Testing set	N_C	Number of classes
Q	Dimension-reduced testing set	m	Dimension of raw data
N	Number of the training data	d	Reduced dimension
N_M	Number of modalities in training sample	U	Transformation matrix
N_T	Number of modalities in testing sample	I	An identity matrix
N_l	Number of lth modal data in training set	$\mathbb{R}^m$	a m-dimensional Euclidean space

2. Geometry preserving projections

In this section, GPP is proposed as a new subspace selection algorithm, which can be used to extract features from multimodal biometric data. To begin with, we describe a new multimodal biometric recognition system, in which the developed GPP is employed.

2.1. The proposed multimodal biometric recognition system

In this multimodal biometric recognition system, all raw biometric data from various modalities are mixed and sorted by individuals, i.e., an individual is a class. Note that each datum is normalized to the same size and scanned into a long vector by writing its values in a fixed order. Denote the multimodal biometric data set $X = [\vec{x}_1, \ldots, \vec{x}_N] \in \mathbb{R}^{m\times N}$, and each datum $\vec{x}_i$ belongs to one of the N_C classes, each of which has two or more modalities. The number of all the modalities is N_M.

The data X is the training data set, on which the GPP algorithm is developed. The system obtains a transformation matrix $U \in \mathbb{R}^{m\times d}$ that maps the set X of N points to the set $Y = [\vec{y}_1, \ldots, \vec{y}_N] \in \mathbb{R}^{d\times N}$, such that $Y = U^{\mathrm{T}}X$, where $d < m$. Given the testing set P, we can project it to the subspace via the transformation matrix U as $Q = U^{\mathrm{T}}P$. Finally, we classify the dimension-reduced testing data Q in the subspace by matching with the corresponding training data Y. Fig. 1 shows the block diagram of the proposed system for multimodal biometrics.

2.2. The geometry preserving projections algorithm

Specifically, in the low-dimensional subspace, we aim to map the raw biometric data from different individuals as far apart as possible, while preserving the geometric properties of the data which belong to the same modality of each class. Fig. 2 shows the main idea of the proposed GPP algorithm. Three modalities, face, palmprint and gait from two sample classes (individuals) are involved in Fig. 2. As shown in Fig. 2 (a), the

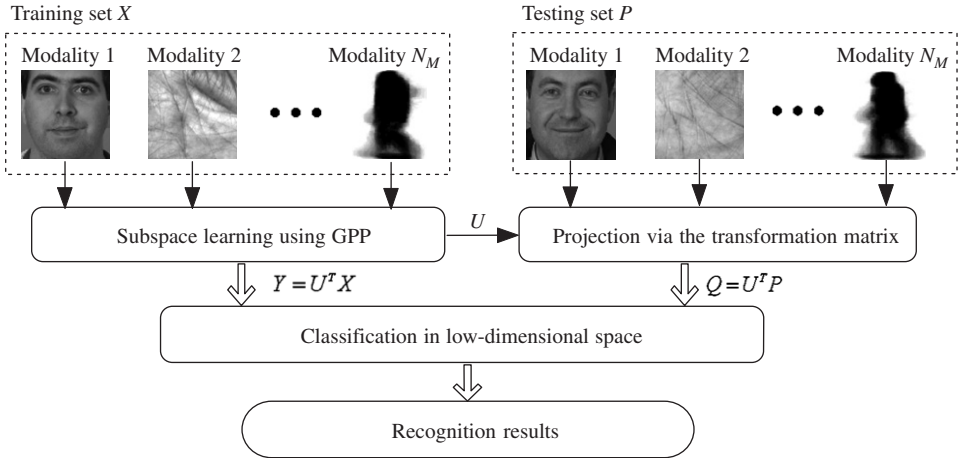

Fig. 1. The block diagram of the proposed system.

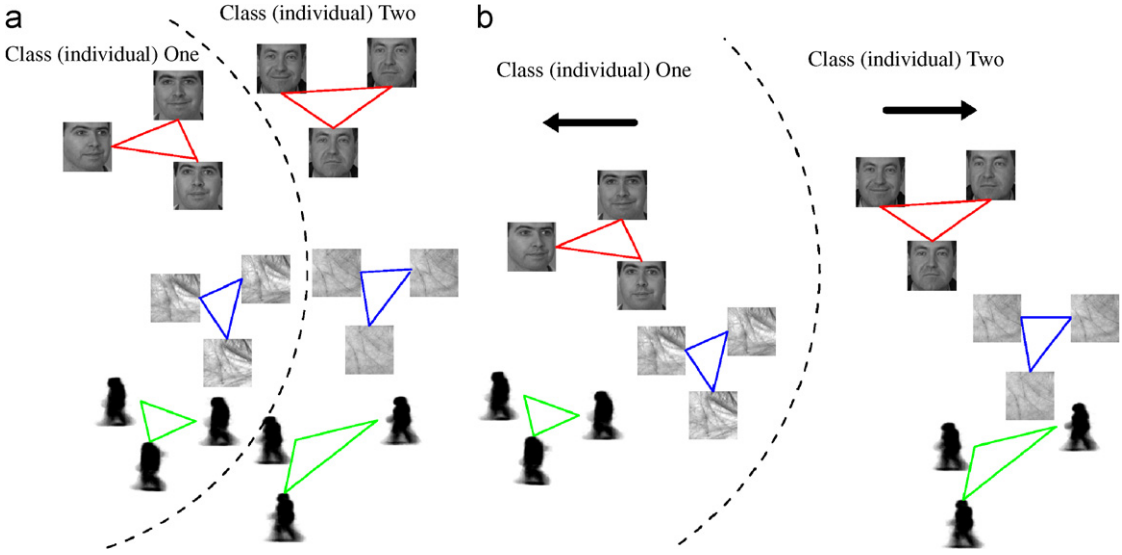

Fig. 2. Illustration of the GPP algorithm: (a) three pairs of two-classes in the high-dimensional space; (b) expected result in the low-dimensional space.

biometric data of the same modalities are close together in the high-dimensional space. Fig. 2 (b) gives the expected result of the algorithm. It shows that the data from different classes are separated while the intra-modal data from same class remain nearby as before.

In the low-dimensional space, we expect that the different class points can be mapped as far as possible. Consider that the inter-class margin can better characterize the separability of different classes than the inter-class variance [21,27]. Hence, we only map the neighbor different class points far apart

$$\max \sum_{i=1}^{N}\sum_{j=1}^{N} \|U^{\mathrm{T}}\vec{x}_i - U^{\mathrm{T}}\vec{x}_j\|^2 s_{ij}, \tag{1}$$

where $s_{ij} = 1$ if $\vec{x}_i$ and $\vec{x}_j$ are from the different classes, and $\vec{x}_i$ is one of k nearest neighbors of $\vec{x}_j$ or $\vec{x}_j$ is one of k nearest neighbors of $\vec{x}_i$, otherwise 0. Eq. (1) reduces to

$$\begin{aligned}\max &\sum_{i=1}^{N}\sum_{j=1}^{N} \mathrm{tr}\{(U^{\mathrm{T}}\vec{x}_i - U^{\mathrm{T}}\vec{x}_j)(U^{\mathrm{T}}\vec{x}_i - U^{\mathrm{T}}\vec{x}_j)^{\mathrm{T}}\}s_{ij}\\ &= \max\ 2\,\mathrm{tr}(U^{\mathrm{T}}XDX^{\mathrm{T}}U - U^{\mathrm{T}}XSX^{\mathrm{T}}U)\\ &= \max\ \mathrm{tr}(U^{\mathrm{T}}XLX^{\mathrm{T}}U),\end{aligned} \tag{2}$$

where tr() denotes the trace operator, $L = D - S$, and D is a diagonal matrix with $d_{ii} = \sum_j s_{ij}$.

At the same time, we expect that in the subspace, the geometric configurations of various modalities of each class can be preserved. So, we characterize the intra-modal geometry of each class by linear coefficients which can reconstruct the given data point by other points from the same modal. The fashion of representation is similar to that of LLE [28]. LLE regards each data point and its nearest neighbors as the local patch. Here, since the modal information is available, we treat each modality of each class as the locality. We can obtain the coefficient

matrix W by minimizing the reconstruction error

$$E = \sum_{i=1}^{N} \left\| \vec{x}_i - \sum_j w_{ij} \vec{x}_j \right\|^2, \tag{3}$$

where $w_{ij} = 0$ unless $\vec{x}_i$ and $\vec{x}_j$ are from the same modality of each class, and the rows of W sum to 1: $\sum_j w_{ij} = 1$. With these constraints mentioned above, the coefficient matrix W can be computed in closed form. We give a demonstration on how to obtain the coefficients of an arbitrary point $\vec{x}$ in the following. Denote $\vec{x}$'s n same-class intra-modal points $\vec{\eta}_1, \cdots, \vec{\eta}_n$ and the corresponding coefficients $w_1, \ldots, w_n$, the reconstruction error can be converted to:

$$\begin{aligned} \xi &= \left\| \vec{x} - \sum_{j=1}^{n} w_j \vec{\eta}_j \right\|^2 = \left\| \sum_{j=1}^{n} w_j (\vec{x} - \vec{\eta}_j) \right\|^2 \\ &= \sum_{j=1}^{n} \sum_{k=1}^{n} w_j w_k G_{jk}, \end{aligned} \tag{4}$$

where, $G_{jk} = (\vec{x} - \vec{\eta}_j)^{\mathrm{T}} (\vec{x} - \vec{\eta}_k)$, called local Gram matrix. By solving the least squares problem with the constraint $\sum_{j=1}^{n} w_j = 1$, the optimal coefficients are given

$$w_j = \frac{\sum_{k=1}^{n} G_{jk}^{-1}}{\sum_{p=1}^{n} \sum_{q=1}^{n} G_{pq}^{-1}}. \tag{5}$$

To faithfully preserve the intra-modal geometric properties of each class, we assume that the coefficients which reconstruct the input x_i in the high-dimensional space will also reconstruct the output y_i from the corresponding intra-modal points in the low-dimensional space. Hence, for all the points, we need to minimize the cost function as follows:

$$\min \sum_{i=1}^{N} \left\| U^{\mathrm{T}} \vec{x}_i - \sum_j w_{ij} U^{\mathrm{T}} \vec{x}_j \right\|^2. \tag{6}$$

Eq. (6) reduces to

$$\begin{aligned} &\min \sum_{i=1}^{N} \mathrm{tr} \left\{ \left(U^{\mathrm{T}} \vec{x}_i - \sum_j w_{ij} U^{\mathrm{T}} \vec{x}_j \right) \left(U^{\mathrm{T}} \vec{x}_i - \sum_j w_{ij} U^{\mathrm{T}} \vec{x}_j \right)^{\mathrm{T}} \right\} \\ &= \min \mathrm{tr} \{ U^{\mathrm{T}} X (I - W)^{\mathrm{T}} (I - W) X^{\mathrm{T}} U \} \\ &= \min \mathrm{tr} (U^{\mathrm{T}} X M X^{\mathrm{T}} U), \end{aligned} \tag{7}$$

where $M = (I - W)^{\mathrm{T}} (I - W)$ and I is an identity matrix. Combining Eqs. (1) and (6) together, we can write the final objective function which needs to be solved as follows:

$$\begin{aligned} \min \Bigg(&\sum_{i=1}^{N} \left\| U^{\mathrm{T}} \vec{x}_i - \sum_j w_{ij} U^{\mathrm{T}} \vec{x}_j \right\|^2 \\ &- \theta \sum_{i=1}^{N} \sum_{j=1}^{N} \| U^{\mathrm{T}} \vec{x}_i - U^{\mathrm{T}} \vec{x}_j \|^2 s_{ij} \Bigg), \end{aligned} \tag{8}$$

where, θ is a scaling factor with the range of 0–1.

According Eqs. (2) and (7), Eq. (8) can reduce to

$$\min \mathrm{tr} \{ U^{\mathrm{T}} X (M - \theta L) X^{\mathrm{T}} U \}. \tag{9}$$

To uniquely determine U, we impose the constraint $U^{\mathrm{T}} U = I$, that is, the columns of U are orthonormal. Now, the objective function has the final form:

$$\begin{cases} \arg \min\limits_{U} \mathrm{tr} \{ U^{\mathrm{T}} X (M - \theta L) X^{\mathrm{T}} U \} \\ \text{s.t. } U^{\mathrm{T}} U = I \end{cases}. \tag{10}$$

Obviously, the above optimization problem can be converted to solving a standard eigenvalue problem:

$$X (M - \theta L) X^{\mathrm{T}} \vec{\beta} = \lambda \vec{\beta}. \tag{11}$$

Let the column vectors $\vec{\beta}_1, \vec{\beta}_2 \ldots, \vec{\beta}_d$ be the solutions of Eq. (11), ordered according to the eigenvalues, $\lambda_1 < \lambda_2 < \ldots < \lambda_d$. Thus, the optimal transformation matrix U is given

$$U = [\vec{\beta}_1, \vec{\beta}_2 \ldots, \vec{\beta}_d]. \tag{12}$$

The proposed algorithm successfully avoids the singular problem since it never computes the inverse of one matrix, different from the algorithms which lead to the generalized eigenvalue problem and are thereby troubled with the problem of singularity.

Note that, the raw data may be high-dimensional and thereby the eigen-decomposition on $X (M - \theta L) X^{\mathrm{T}}$ would be computationally expensive. So, the step of PCA projection is recommended for compressing the raw data. Moreover, the preprocessing using PCA can reduce the noise which may degrade the recognition rates. We can choose the optimal value for the reduced dimension in the PCA step, without the discriminant information loss. Thus, we perform GPP in the PCA-projected subspace and the ultimate transformation matrix is as follows:

$$U = U_{PCA} U_{GPP}. \tag{13}$$

2.3. The kernel extension

GPP is essentially a linear algorithm. It always has limitations under the situations when the biometric data are highly non-linear. In that non-linear case, the measurement using the Euclidean distance fails to give the true neighbor relationship for discriminant and fails to detect the intra-modal geometric structure. In this subsection, we attempt to overcome this non-linear problem by using the kernel trick [25,26], which produces the algorithm: *kernel GPP* (KGPP).

We first map the raw data $X = [\vec{x}_1, \cdots, \vec{x}_N] \in \mathbb{R}^{m \times N}$ into some high-dimensional feature space $\mathcal{F}$ via a non-linear mapping $\phi: \mathbb{R}^m \to \mathcal{F}$. So, we have

$$\phi(X) = [\phi(\vec{x}_1), \ldots, \phi(\vec{x}_N)]. \tag{14}$$

Assume that in the feature space, there is the transformation matrix U_{KGPP} which projects $\phi(X)$ to the 1-dimensional space for simplifying the discussion, such that $U_{KGPP}^{\mathrm{T}} \phi(X) = [U_{KGPP}^{\mathrm{T}} \phi(\vec{x}_1), \cdots, U_{KGPP}^{\mathrm{T}} \phi(\vec{x}_N)]$. According to the analysis in Section 2.2, we have the new objective function:

$$\min \mathrm{tr} \{ U_{KGPP}^{\mathrm{T}} \phi(X) (M - \theta L) \phi(X)^{\mathrm{T}} U_{KGPP} \}. \tag{15}$$

Because the transformation matrix U_{KGPP} should lie in the span of $\phi(\vec{x}_1), \ldots, \phi(\vec{x}_N)$, there exist coefficients $\vec{\alpha} = [\alpha_1, \ldots \alpha_N]^{\mathrm{T}}$, such that

$$U_{KGPP} = \sum_{i=1}^{N} \alpha_i \phi(\vec{x}_i) = \phi(X)\vec{\alpha}. \tag{16}$$

Combining Eqs. (15) and (16), we get

$$\min \mathrm{tr}\{\vec{\alpha}^{\mathrm{T}} K(M - \theta L)K\vec{\alpha}\}, \tag{17}$$

where K is a kernel matrix with

$$K_{ij} = k(\vec{x}_i, \vec{x}_j) = \phi(\vec{x}_i) \cdot \phi(\vec{x}_j). \tag{18}$$

Considering the constraint $U_{KGPP}^{\mathrm{T}} U_{KGPP} = I$, we have the final objective function for KGPP

$$\begin{cases} \arg\min\limits_{\vec{\alpha}} \mathrm{tr}\{\vec{\alpha}^{\mathrm{T}} K\,(M - \theta L)\,K\vec{\alpha}\} \\ \text{s.t. } \vec{\alpha}^{\mathrm{T}} K\vec{\alpha} = 1 \end{cases}. \tag{19}$$

The above minimization problem can be converted to solving a generalized eigenvalue problem as follows:

$$K(M - \theta L)K\vec{\alpha} = \lambda K\vec{\alpha}. \tag{20}$$

The eigenvector associated with the smallest eigenvalues are the solution.

If a testing sample p comes, we have the projection

$$q = U_{KGPP}^{\mathrm{T}} \phi(p) = \vec{\alpha}^{\mathrm{T}} \phi(X)^{\mathrm{T}} \phi(p) = \sum_{i=1}^{N} \alpha_i k(x_i, p). \tag{21}$$

3. Classification in the subspace

By using the GPP algorithm, the transformation matrix U is yielded, and then the training data can be projected into the subspace: $[\vec{y}_1, \ldots, \vec{y}_N] = U^{\mathrm{T}}[\vec{x}_1, \ldots, \vec{x}_N]$. As introduced in Section 2.1, the testing data P are projected onto $Q = U^{\mathrm{T}} P$. The dimension-reduced testing data Q and the dimension-reduced training data Y have the same space where we can perform the classification.

Given a testing individual constituted by the multimodal features $\vec{q}_1, \ldots \vec{q}_{N_T}$, where N_T is the number of the modalities of the individual and $1 \leqslant N_T \leqslant N_M$, note that, the number is not constrained to N_M, that is, it is permitted that some modalities of testing samples are not available. For each modality $\vec{q}_j$, we measure the Euclidean distance between the given feature and all the dimension-reduced training samples $\vec{y}_1, \ldots, \vec{y}_N$. So we have the distance vector $\vec{D}_j$ for modality $\vec{q}_j$:

$$\vec{D}_j = [d(\vec{q}_j, \vec{y}_1), \ldots, d(\vec{q}_j, \vec{y}_N)], \tag{22}$$

where $j = 1, \ldots, N_T$. Since the training samples $\vec{y}_1, \ldots, \vec{y}_N$ come from the different modalities, it is necessary to normalize the distance measurement for each modality. Denote N_l the number of data in lth modality in the training set. We have the normalization

$$\tilde{d}(\vec{q}_j, \vec{y}_i) = \frac{d(\vec{q}_j, \vec{y}_i)}{\sqrt{d^2(\vec{q}_j, \vec{y}_i) + \sum_{k=1}^{N_l - 1} d^2(\vec{q}_j, \vec{y}_i^k)}}, \tag{23}$$

where, assume $\vec{y}_i^k$ and $\vec{y}_i$ come for the same lth modality. Rewrite $\vec{D}_j$ as

$$\vec{D}_j = [\tilde{d}(\vec{q}_j, \vec{y}_1), \cdots, \tilde{d}(\vec{q}_j, \vec{y}_N)]. \tag{24}$$

Considering all the modalities of the testing individual, we have the following decision,

$$\vec{q}_1, \cdots \vec{q}_{N_T} \in label(\vec{y}_i), \tag{25}$$

if $\tilde{d}(\vec{q}_j, \vec{y}_i) = \min[\min \vec{D}_1, \ldots, \min \vec{D}_{N_T}]$, for $j = 1, \ldots, N_T$ and $i = 1, \ldots, N$.

4. Experiments

In experiments, three modalities, namely face, palmprint and gait, are used. Among them, face and gait are contactless while palmprint is a contacting biometric resource. Our system can be extended to include more modalities.

For face modality, we select two databases: YALE [29] and FERET [30]. The YALE database contains 15 subjects and each subject has 11 face images with varying facial expressions and configurations: center-light, with glasses, happy, left-light, without glasses, normal, right-light, sad, sleepy, surprised, and wink. In comparison to YALE, FERET is a rather larger database. It contains 13,539 face images of 1565 subjects taken during different photo sessions with variations in size, pose, illumination, facial expression and even age.

The palmprint [31,6] database is provided by the Hong Kong Polytechnic University (HKPU). This database contains 7752 palmprint images from 386 individuals. Each participant contributes about 20 images.

For gait, we adopt the USF HumanID outdoor gait database [32] of version 2.1. The database was built for vision-based gait recognition, and it is widely used. It consists of 1870 sequences from 122 subjects. For each of the subjects, there are the following covariates: change in viewpoint (left or right), change in shoe type (A or B), change in walking surface (grass or concrete), change in carrying condition (briefcase or no briefcase), and elapsed time (May or November) between sequences being compared. All the images of the database are divided into one gallery set and 12 probe sets A-L.

Using the databases briefly introduced above, we built two multimodal data sets, the YALE-HKPU-USF data set and the FERET-HKPU-USF data set, for biometric recognition. In the experiments, GPP along with its kernel extension KGPP are implemented in comparison with baseline, PCA [22], LDA [23], LPP [24], and MFA [21]. Among these algorithms, baseline performs the recognition directly in the raw data without dimensionality reduction; PCA and LDA are the traditional dimension reduction methods; LPP and MFA are the recently proposed manifold learning methods. Gaussian kernels are used in KGPP. To better show the system's advantage, the unimodal biometrics by using baseline, PCA, LDA, LPP and MFA are also implemented for each modality. For all the unimodal tests, the Nearest Neighbor classifier is employed in the last recognition stage.

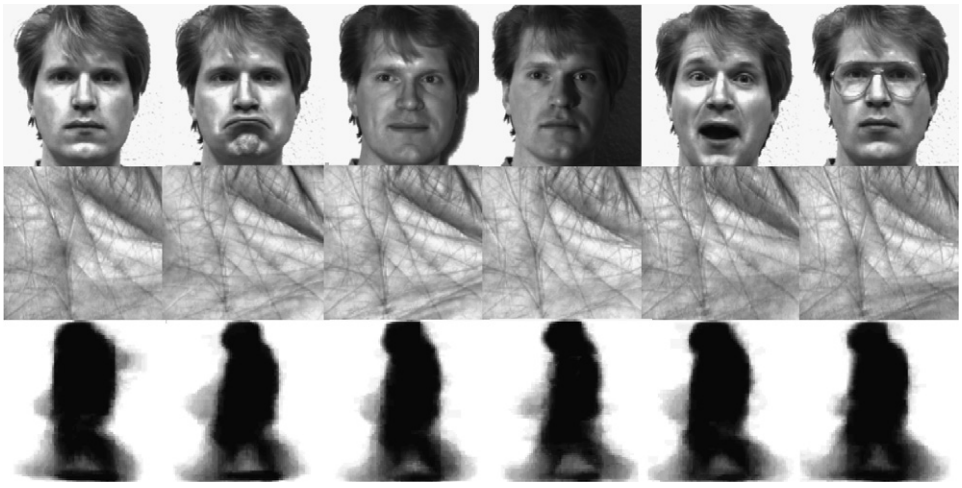

Fig. 3. Multimodal images from the YALE-HKPU-USF data set.

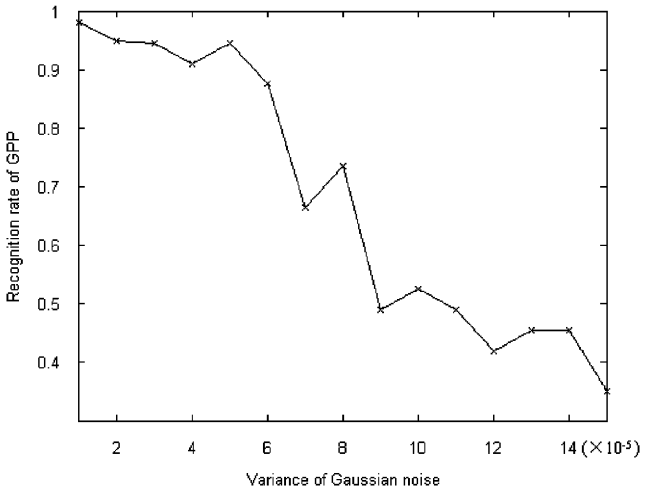

Fig. 4. Recognition rate of GPP versus the variance of the Gaussian on the YALE-HKPU-USF data set.

Table 2
Best recognition results (%) on the YALE-HKPU-USF data set

Modalities	Methods	Recognition rate (%)
Unimodal biometrics	Baseline	55.33(1600)
	PCA	62.67(44)
YALE	LDA	76.67 (8)
	LPP	63.33 (40)
	MFA	77.33(11)
HKPU	Baseline	63.33(1600)
	PCA	66.67 (58)
	LDA	42.67 (14)
	LPP	69.33 (56)
	MFA	73.33 (29)
USF	Baseline	66.73 (1600)
	PCA	68.67 (53)
	LDA	79.33 (20)
	LPP	67.33 (54)
	MFA	80.67 (17)
Multimodal biometrics	Baseline	65.56 (1600)
	PCA	72.89 (46)
	LDA	70.44 (14)
	LPP	64.44 (27)
	MFA	74.67 (84)
	GPP	87.56 (69)
	KGPP	90.22 (67)

4.1. YALE-HKPU-USF

We randomly chose 15 subjects to form the YALE-HKPU-USF data set. Each subject has 6 images from YALE, 6 images from HKPU and 6 images from USF, respectively. Therefore, the YALE-HKPU-USF data set contains $(6+6+6)\times 15 = 270$ images in total. For the USF modality, the images are selected from the gallery set and probe set A. All the images are cropped and normalized to the size of 40×40 pixels whose values are within the scope of 0 to 1. Assume that the subjects of same order number from different modalities originate from the identical individual. Fig. 3 shows the multimodal images of one sample subject (individual).

For each modality of one individual, we randomly select 4 images for training and the rest 2 images for testing respectively, that is, we have 12 images for training and 6 images for testing for one individual. We repeat this operation 5 times and obtain 5 random subsets for computing the average results. Since the images of HKPU and USF are captured and processed specially for lab experiments, the recognition rates of most algorithms on these data are quite high. In real world applications, it is hard to obtain such high-quality images. To simulate the case in reality, we added some Gaussian white noise to the images from HKPU and USF to tune down the corresponding recognition rates on them. To show the Gaussian noise's effect on the final performance, Fig. 4 gives the relationship between recognition rate of GPP and the variance of the Gaussian noise when the mean of the noise is fixed to zero and the dimension of the subspace is fixed to 69. The final parameters of the Gaussian noise are set zero mean and 6×10^{-5} variance. Table 2 gives the best recognition rates along with the corre-

sponding reduced dimensions of the employed algorithms in both unimodal and multimodal tests. Fig. 5 shows the average recognition rates versus subspace dimensions in multimodal tests. As shown, GPP and KGPP yield the best recognition rates than the other methods not only in multimodal but also unimodal biometrics.

4.2. FERIET-HKPU-USF

Similar to YALE-HKPU-USF, for the FERET-HKPU-USF data set, we randomly choose 100 subjects each of which has six images from FERET, HKPU, and USF, respectively. For USF, the images are still selected from the gallery set and probe set A. The formed multimodal data set totally contains 1800 images. All the images are cropped and normalized to the size of 40×40 pixels whose values are between 0 and 1. We also have the similar assumption as YALE-HKPU-USF. Fig. 6 shows one individual's images which come from three modalities.

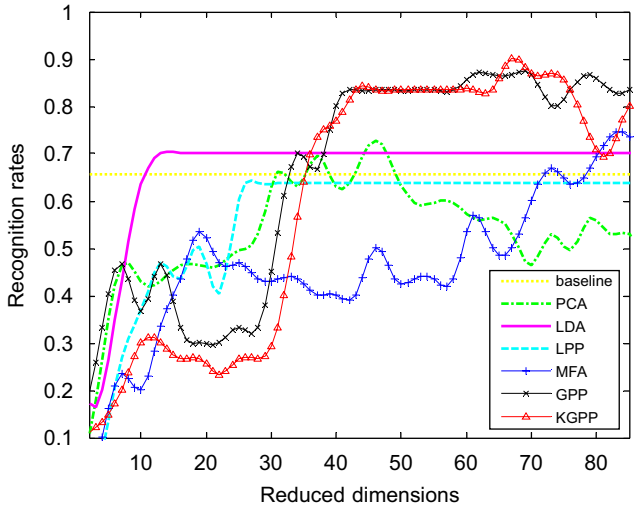

Fig. 5. Recognition rates versus subspace dimensions on the YALE-HKPU-USF data set.

Three images per modality (nine images per individual) are randomly selected for training and the rest three images per modality (nine images per individual) are used for testing. Some Gaussian noise has been added to the images of the HKPU modality to adjust the recognition rates. The parameters of the noise are set zero mean and 2×10^{-5} variance. All the tests are repeated over five random splits independently and the average recognition results are calculated. Table 3 and Fig. 7 give the experimental results. It is clear that we can draw the similar conclusions as before.

4.3. The case of modalities missing

In this subsection, tests are carried out in the case that some modalities are not available. The YALE-HKPU-USF data set are used to conduct the experiment on. The strategy used is similar as that in Section 4.1. While, the difference is that: in this set of experiments, we randomly removed some modalities of the testing samples. The number of the missing modalities for all the testing samples is set from 1 to 30. Fig. 8 plots the corresponding recognition rate of GPP when the dimension of the subspace is fixed to 69. Experiments show promising results that when the number of missing modalities increases, the performance of GPP degrades slowly.

4.4. Discussions

It is necessary to highlight some views in the experiments.

The multimodal biometric data intrinsically follow the multimodal Gaussian distribution. GPP is specially developed to tackle the multimodal problems. It attempts to obtain a more discriminant subspace while succeeds in describing and representing each Gaussian modality by using the linear reconstruction coefficients. This is the main reason that GPP has the best recognition rates in multimodal biometrics.

MFA is also an effective method to solve the multimodal Gaussian problems because of its consideration of both class and neighbor information [21]. MFA can be seen as an

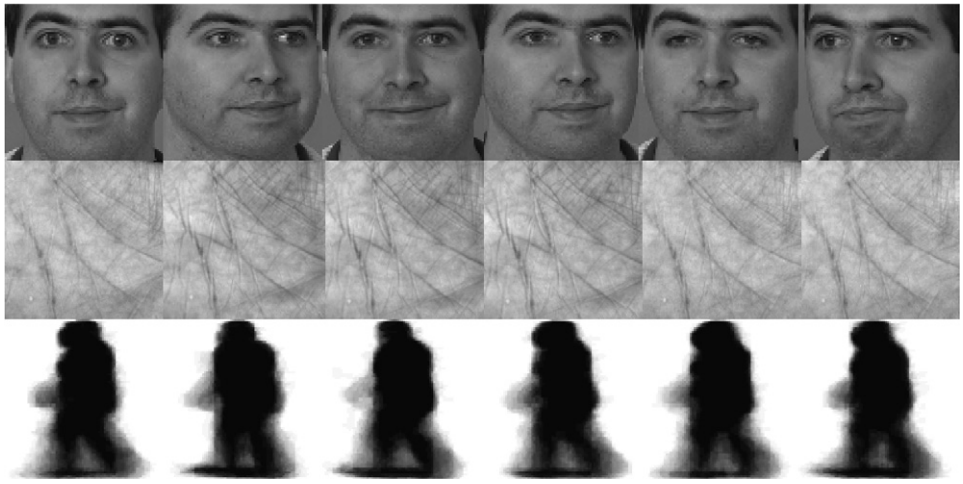

Fig. 6. Multimodal images from the FERET-HKPU-USF data set.

Table 3
Best recognition results (%) on the FERET-HKPU-HUMI data set

Modalities	Methods	Recognition rate (%)
Unimodal biometrics	Baseline	39.07(1600)
	PCA	39.07(295)
FERET	LDA	55.13(34)
	LPP	42.93(205)
	MFA	49.73(126)
	Baseline	87.93(1600)
	PCA	88.13(294)
HKPU	LDA	81.93 (94)
	LPP	65.67 (251)
	MFA	88.33 (60)
	Baseline	75.47 (1600)
	PCA	75.60(189)
HUMI	LDA	83.53(88)
	LPP	63.60 (215)
	MFA	85.87 (84)
	Baseline	73.89 (1600)
Multimodal biometrics	PCA	74.78 (400)
	LDA	75.00 (80)
	LPP	55.44 (380)
	MFA	79.67(375)
	GPP	93.11 (216)
	KGPP	93.67 (224)

extension of LPP which does not resort to the class information. MFA is unsupervised since it seeks the multimodal distributions in the biometric data by using the k-nearest neighbors search. While, GPP is supervised as it exploits the modal information. In addition, GPP gives more precise representations on the local (intra-modal) structure by the linear coefficients than MFA which uses the simple joint relationship of the neighbor points. These can explain why GPP is superior to MFA for multimodal biometrics.

In the experiments, we also have the interesting observation that LPP performs well in the unimodal tests while its performance in multimodal tests is not satisfactory. This may stem from the fact that, in the multimodal biometric data, the intra-modal similarity is always larger than the intra-class similarity. For example, the Euclidean distance between two points from same modalities but different classes is usually shorter than that between two points from different modalities but same classes. The k-nearest neighbors search of LPP may be misled without restriction of doing that in each class.

5. Conclusions

In the paper, a new system is proposed for multimodal biometrics. In the given system, a new method named *Geometry Preserving Projections* (GPP) has been designed for the subspace selection on the multimodal biometric data, which intrinsically follow the multimodal Gaussian distribution.

As a discriminant algorithm, GPP optimization scheme focuses on the inter-class margin points. These points possess more discriminant abilities than the global inter-class points. Meanwhile, GPP preserves the intra-modal geometric structure by using linear reconstruction coefficients. A unified transformation matrix can be learnt by GPP for the raw biometric data formed by various modalities. GPP makes the system more flexible since it can also work well when one or more modalities of the testing samples are not available. *Kernel GPP* (KGPP) has been proposed to overcome the non-linear problem and further improve the ability of GPP. Experimental results have shown the effectiveness of the proposed system and algorithm upon large data sets.

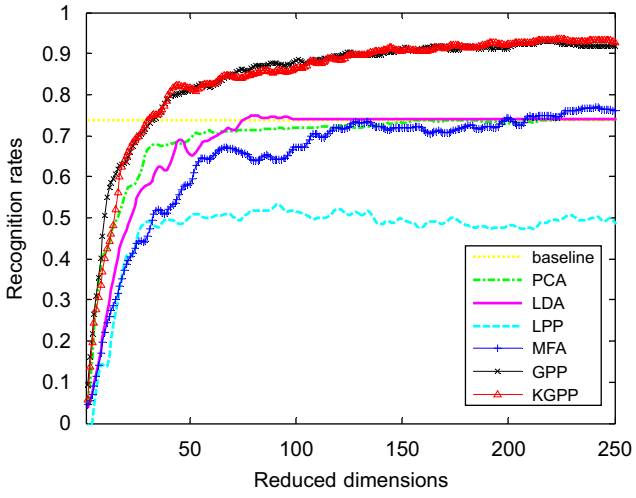

Fig. 8. Recognition rates versus subspace dimensions on the FERET-HKPU-USF data set.

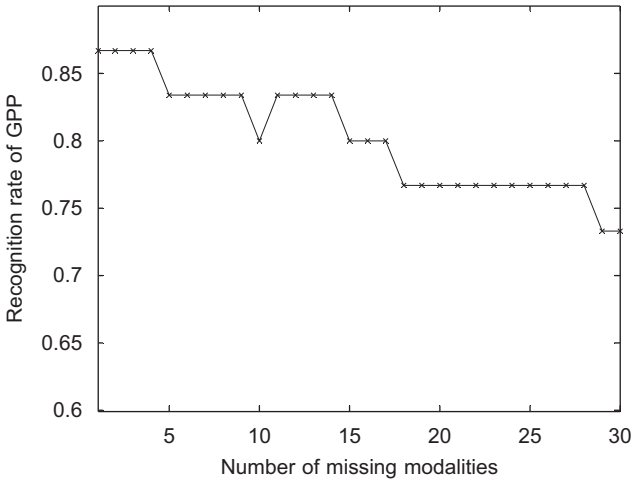

Fig. 9. Recognition rate of GPP versus the number of missing modalities on the YALE-HKPU-USF data set.

Acknowledgments

The authors would like to thank anonymous reviewers for their constructive comments on the first version of this paper. The research was supported by the Internal Competitive Research Grants of the Department of Computing with the Hong Kong Polytechnic University (under project number A-PH42), National Science Foundation of China (No. 60675023) and China 863 High Tech. Plan (No. 2007AA01Z164).

References

[1] A.K. Jain, A. Ross, S. Pankanti, Biometrics: a tool for information security, IEEE Trans. Inf. Forensics Security 1 (2) (2006) 125–143.

[2] W. Hu, T. Tan, L. Wang, S.J. Maybank, A survey on visual surveillance of object motion and behaviors, IEEE Trans. Syst. Man Cybern. Part C. 34 (3) (2004) 334–352.

[3] M. Betke, J. Gips, P. Fleming, The camera mouse: visual tracking of body features to provide computer access for people with severe disabilities, IEEE Trans. Neural Syst. Rehab. Eng. 10 (1) (2002) 1–10.

[4] L.D. Harmon, M.K. Khan, R. Lasch, P.F. Raming, Machine identification of human faces, Pattern Recognition 13 (2) (1981) 97–110.

[5] A.K. Jain, L. Hong, S. Pankanti, R. Bolle, An identity authentication system using fingerprints, Proc. IEEE 85 (9) (1997) 1365–1388.

[6] D. Zhang, W.K. Kong, J. You, M. Wong, On-line palmprint identification, IEEE Trans. Pattern Anal. Mach. Intell. 25 (9) (2003) 1041–1050.

[7] J.G. Daugman, High confidence visual recognition of persons by a test of statistical independence, IEEE Trans. Pattern Anal. Mach. Intell. 15 (11) (1993) 1148–1161.

[8] J. Campbell, Speaker recognition: a tutorial, Proc. IEEE 85 (9) (1997) 1437–1462.

[9] A.K. Jain, A. Ross, S. Pankanti, A prototype hand geometry-based verification system, in: Proceedings of the Second International Conference on Audio- and Video-based Biometric Person Authentication, 1999, pp. 166–171.

[10] T. Moeslund, E. Granum, A survey of computer vision-based human motion capture, Comput. Vision Image Understanding 81 (3) (2001) 231–268.

[11] X. Jing, Y. Yao, D. Zhang, M. Li, Face and palmprint pixel level fusion and Kernel DCV-RBF classifier for small sample biometrics recognition, Pattern Recognition 40 (11) (2007) 3209–3224.

[12] A.K. Jain, A. Ross, Multibiometric systems, Commun. ACM 47 (1) (2004) 34–40.

[13] L. Hong, A.K. Jain, Integrating faces and fingerprints for personal identification, IEEE Trans. Pattern Anal. Mach. Intell. 20 (12) (1998) 1295–1307.

[14] S. Ribaric, I. Fratric, A biometric identification system based on eigenpalm and eigenfinger features, IEEE Trans. Pattern Anal. Mach. Intell. 27 (11) (2005) 1698–1709.

[15] A.K. Jain, K. Nandakumar, A. Ross, Score normalization in multimodal biometric systems, Pattern Recognition 38 (12) (2005) 2270–2285.

[16] K.A. Toh, X.D. Jiang, W.Y. Yau, Exploiting global and local decisions for multi-modal biometrics verification, IEEE Trans. Signal Process. 52 (10) (2004) 3059–3072.

[17] Y. Gao, M. Maggs, Feature-level fusion in personal identification, in: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2005, pp. 468–473.

[18] Y. Yao, X. Jing, H. Wong, Face and palmprint feature level fusion for single sample biometrics recognition, Neurocomputing 70 (7–9) (2007) 1582–1586.

[19] T. Sim, S. Zhang, R. Janakiraman, S. Kumar, Continuous verification using multimodal biometrics, IEEE Trans. Pattern Anal. Mach. Intell. 29 (4) (2007) 687–700.

[20] A. Ross, A.K. Jain, Multimodal biometrics: an overview, in: Proceedings of the 12th European Signal Processing Conference, 2004, pp. 1221–1224.

[21] S. Yan, D. Xu, B. Zhang, H. Zhang, Q. Yang, S. Lin, Graph embedding and extensions: a general framework for dimensionality reduction, IEEE Trans. Pattern Anal. Mach. Intell. 29 (1) (2007) 40–51.

[22] M. Turk, A. Pentland, Eigenfaces for recognition, Int. J. Cognitive Neurosci. 3 (1) (1991) 71–86.

[23] P.N. Belhumeur, J.P. Hespanha, D.J. Kriegman, Eigenfaces vs. Fisherfaces: recognition using class specific linear projection, IEEE Trans. Pattern Anal. Mach. Intell. 19 (1997) 711–720.

[24] X. He, S. Yan, Y. Hu, P. Niyogi, H. Zhang, Face recognition using Laplacianfaces, IEEE Trans. Pattern Anal. Mach. Intell. 27 (3) (2005) 328–340.

[25] V. Vapnik, Statistical Learning Theory, Wiley, New York, 1998.

[26] B. Schölkopf, A. Smola, K.R. Müller, Nonlinear component analysis as a kernel eigenvalue problem, Neural Comput. 10 (1999) 1299–1319.

[27] K. Fukunaga. Introduction to statistical pattern recognition, second ed. Academic Press, Boston, 1990.

[28] S.T. Roweis, L.K. Saul, Nonlinear dimensionality reduction by locally linear embedding, Science 290 (2000) 2323–2326.

[29] Available at ⟨http://cvc.yale.edu/projects/yalefaces/yalefaces.html⟩.

[30] A.P.J. Phillips, H. Moon, P.J. Rauss, S. Rizvi, The FERET evaluation methodology for face recognition algorithms, IEEE Trans. Pattern Anal. Mach. Intell. 22 (10) (2000) 1090–1104.

[31] Available at ⟨http://www4.comp.polyu.edu.hk/∼biometrics/⟩

[32] S. Sarkar, P. Phillips, Z. Liu, I. Vega, P. Grother, K. Bowyer, The HumanID gait challenge problem: data sets, and analysis, IEEE Trans. Pattern Anal. Mach. Intell. 27 (2) (2005) 162–177.

About the Author—TIANHAO ZHANG is currently a research assistant at the Hong Kong Polytechnic University and a Ph.D. candidate at the Institute of Image Processing and Pattern Recognition, Shanghai Jiao Tong University. He was born in Shandong, PR China, in November 1980. He received his Bachelor degree in Electrical Engineering the from Shandong University in 2002, and the M.Phil in Power Machinery and Engineering from the Chang'an University in 2005. His research interests include manifold learning, pattern recognition, and computer vision. He serves as a reviewer of Neurocomputing, International Journal of Computer Mathematics (IJCM) and a member of the Program Committee of the 2007 IEEE Pacific Rim Symposium on Image Video and Technology (PSIVT2007).

About the Author—XUELONG LI currently holds a permanent academic post at the University of London. His research covers several topics in vision, cognition, interaction, and their applications. He has published more than 60 papers in IEEE T-PAMI, T-CSVT, T-IP, T-KDE, T-MM, T-SMC, CVPR, ICDM, etc. He is an editor of two books and 10 journals. He is a guest co-editor of seven special issues. He serves as a PC co-chair of CW2008, a PC co-chair and an invited session co-chair of the 6th IEEE ICMLC, a co-chair of the 5th UKCI, a theme co-chair of the 2nd IEEE PSIVT, and a publicity co-chair of the 7th IEEE ICDM, the 4th ICIG, and the 2nd DMAMH. He serves as PC member for more than 60 conferences. He is a reviewer for over a hundred journals and conferences, including 10 IEEE transactions. He is a member of the IEEE SMC TC on CI, and a member of the IEEE SP TC on MLSP. He is a Senior Member of the IEEE.

About the Author—DACHENG TAO received the B.E. degree from the University of Science and Technology of China (USTC), the M.Phil degree from the Chinese University of Hong Kong (CUHK), and the Ph.D. from the University of London. He is currently an assistant professor at the Hong Kong Polytechnic University. His research interests include artificial intelligence, computer vision, data mining, information theory, machine learning, and visual surveillance. He published extensively at IEEE TPAMI, TKDE, TIP, TMM, TCSVT, CVPR, ICDM, ACM Multimedia, ACM KDD, etc. Previously he gained several Meritorious Awards from the Int'l Interdisciplinary Contest in Modeling, which is the highest level mathematical modeling contest in the world, organized by COMAP. He is an editor two books and six journals. He is an associate editor of the Neurocomputing (Elsevier). He co-chaired the Special Session on Information Security in the IEEE ICMLC and Workshop on Knowledge Discovery and Data Mining from Multimedia Data and Multimedia Applications in IEEE ICDM.

About the Author—Jie Yang was born in Shanghai, China, in August 1964. He received a Ph.D. in computer from the Department of Computer, University of Hamburg, Germany. Dr Yang is now the Professor and Vice-director of the Institute of Image Processing and Pattern Recognition, Shanghai Jiao Tong University. He is incharge of more than 20 national and ministry level scientific research projects in image processing, pattern recognition, data amalgamation, data mining, and artificial intelligence.