Exploiting Global and Local Decisions for Multimodal Biometrics Verification

Kar-Ann Toh, *Senior Member, IEEE*, Xudong Jiang, *Member, IEEE*, and Wei-Yun Yau, *Member, IEEE*

***Abstract*—In this paper, we address the multimodal biometric decision fusion problem. By exploring into the user-specific approach for learning and threshold setting, four possible paradigms for learning and decision making are investigated. Since each user requires a decision hyperplane specific to him in order to achieve good verification accuracy, those tedious iterative training methods like the neural network approach would not be suitable. We propose to use a model that requires only a single training step for this application. The four global and local learning and decision paradigms are then explored to observe their decision capability. Besides the proposal of a relevant receiver operating characteristic performance for the local decision, extensive experiments were conducted to observe the verification performance for fusion of two and three biometrics.**

***Index Terms*—Biometrics, decisions fusion and multivariate polynomials, pattern classification, pattern recognition.**

I. Introduction

A. Background

DUE to inherent properties in each biometric and external manufacturing constraints in the sensing technologies, to date, no single biometric method can warrant 100% authentication accuracy by itself. This problem can be alleviated by combining multiple biometric methods without resorting to the use of a conventional password system. The importance of multimodal biometrics thus need not be overemphasized.

The biometric verification problem can be considered to be a classification problem wherein a decision is made upon whether or not a claimed identity is genuine with inference to some matching criteria. We thus treat the problem of combining multimodal biometrics as a classifier decision combination problem in this paper. Generally, the approaches for classifiers combination differ in terms of assumptions about classifier dependencies, type of classifier outputs, combining strategies, and combining procedures [1]. Two main types of combination can be identified: *classifier selection* and *classifier fusion*. The difference between these two types lies in whether the classifiers are assumed to be complementary or competitive. Classifier selection assumes that each classifier is a "local expert," whereas classifier fusion assumes that all classifiers' outputs formed an overall feature space for training (see, e.g., [1]). In this paper, our focus will be on classifier fusion, and our main effort will be on arriving at a fusion methodology that maximizes the accuracy of the combined decision.

According to the information adopted, three levels of combination can be identified ([2], [3]):

i) abstract level;
ii) rank level;
iii) measurement level.

At abstract level, the output information taken from each classifier is only a possible label for each pattern class, whereas at rank level, the output information taken from each classifier is a set of ordered possible labels that is ranked by decreasing confidence measure. At measurement level, the output information taken from each classifier is a set of possible labels with associated confidence measure. In this way, with the measurement outputs taken from each individual system, the decision is brought forward to the final output of the combined system. We will work at the measurement level to combine several biometric verification systems.

B. Multimodal Biometrics: State of the Art

For the classifier selection approach, a design scheme is proposed in [4] for classifier combination at abstract level. Prior to the classifier combination, a classifier selection scheme is proposed using the class separation statistics as the feature effectiveness criterion. An exhaustive search of all possible feature subsets was proposed to obtain the best feature subsets containing the individual biometric decisions. The final decision combination used the Neyman–Pearson rule, which was known to be optimal when one type of error was specified. Experimental results on the proposed method was reported to have either comparable or better receiver operating characteristics (ROC) performance with respect to the *sum* rule and always better ROC performances than the *product* rule. It was also reported to observe that independence among various classifiers is directly related to the improvement in performance of the combination.

Many other multimodal biometrics developments adopted the classifier fusion approach. In [5] and [6], the multimodal biometric fusion problem was treated as a classification problem. In [5], the k-nearest-neighbor (k-NN) classifier was applied to combine one vocal and two visual cues (frontal and profile faces) in an identity verification system. In [6], the classification problem was solved using the support vector machine (SVM). Several kernels namely, linear, polynomials, radial basis function (RBF), and multilayer perceptron (MLP), were experimented with in the SVM learning. The biometric experts come from the following modalities: vocal, frontal

Manuscript received October 31, 2003; revised February 28, 2004. The associate editor coordinating the review of this paper and approving it for publication was Dr. Anil K. Jain.

K.-A. Toh and W.-Y. Yau are with the Institute for Infocomm Research, Singapore 119613 (e-mail: katoh@i2r.a-star.edu.sg; katoh@ieee.org; wyyau@i2r.a-star.edu.sg).

X. Jiang is with the School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore (e-mail: exdjiang@ntu.edu.sg).

Digital Object Identifier 10.1109/TSP.2004.833862

face, and profile face. The training data set consists of 111 genuine users and 3996 imposters, and the test data set consists of 37 genuine users and 1332 imposters. Based on these rather small data sets, it was found that the linear kernel outperforms those nonlinear ones. The SVM method was compared with several conventional classification methods including the k-NN classifier, and the results show superior classification accuracy of SVM on test data for most cases. In [7], several binary classification schemes, including SVM, MLP, C4.5, and Fisher's linear discriminant Bayesian classifier, were evaluated to combine the face and speech data for person identity verification. It was reported that the SVM and Bayesian classifiers achieved the best results among the evaluated schemes. In [8], several clustering algorithms, including their variants, were used to combine several unimodal face and voice authentication means. It was reported that fuzzy clustering algorithms have better performance compared with classical k-means and other simple fusion algorithms like AND and OR rules. It was also reported that the median RBF provides a reliable technique for data fusion.

Another important approach is statistical based. In [9], a Bayesian supervisor [10] was proposed to combine the face and speech experts. It was demonstrated that a multimodal system was capable of improving the decision accuracy significantly. The Bayesian supervisor was reported to be more successful than a Mean supervisor in terms of multimodal decision making. The *sum* and *product* rules [11] for combining classifiers can be considered as statistical-based since many statistical conditions have been implicitly assumed. The *product* rule implicitly assumes independence among the decision experts, whereas the *sum* rule implicitly assumes the *a posteriori* probability computed by the decision expert's resemblance to the prior probabilities [4]. The following are two examples of how the product rule can be exploited. In [3], the face and fingerprint biometrics were integrated using a product-based composite imposter distribution. In [2], the speech and face biometrics were integrated using a weighted geometric average, where normalization of each input played an important role since the weighting emphasizes the classification power of the most reliable classifiers.

C. User-Specific Multimodal Biometrics

The idea of localized multimodal biometric learning and decision has probably first been seen in [12], wherein a user-specific scheme is proposed for the multibiometric parameters. In this method, a user-specific matching threshold can be computed using the cumulative histogram of imposter scores for each biometric for each user. The scores for each user-specific-threshold biometric can then be averaged to produce the final score label. Alternatively, each user can be assigned a specific weight for each biometric, and these weights can be estimated using an exhaustive search for minimal error rates (sum of false accept and false reject rates) over all users. Although only a moderately small database (50 users) has been tested in [12], the authors had demonstrated that such specific weights and thresholds for individual users can improve combined matching accuracy as compared with one that uses only common thresholds and equal weights. For both cases, the total number of parameters to be estimated is proportional to the number of biometrics used and the number of users. Indeed, the total number of weighting parameters to be estimated is given by (number_of_user × number_of_biometrics). When there is a large number of users in the system, then estimation of these weighting parameters by these exhaustive means for each biometric and each user becomes nontrivial. Moreover, when a new user is added into the system, an additional set of weight parameters has to be, again, exhaustively estimated and included into the system. It is also noted that a linear decision hyperplane has been adopted for this application, and the system cannot cater to more complex nonlinear decision hyperplanes.

D. Contributions and Organization

Noting that local learning and decision is not well-explored in the context of multimodal biometrics, in this paper, we explore into using a nonlinear decision separation hyperplane to improve the verification accuracy. We adopt a learning methodology that does not need exhaustive estimation of weighting parameters, as in [12], when a new user is added into the system. In addition to those *user-specific-thresholds-with-equal-weights* and *common-threshold-with-user-specific-weights* paradigms, as seen in [12], we explore a new learning and decision paradigm called *local-learning-with-local-decision* and study its performance. While [13] focused only on the conventional nonuser-specific learning and decision treatments similar to those in [4]–[10], contributions of this paper are related to the improvement of verification performance via the proposal of new learning and decision methodologies such as

i) a proposed new local learning scheme that does not require exhaustive learning of weighting parameter for each addition of a new user, as in [12];

ii) proposed new threshold setting schemes and performance measure for local decisions;

iii) origination of a new learning and decision paradigm called local-learning-with-local-decision;

iv) empirical evaluation of four learning and decision paradigms using three biometrics.

This unified evaluation of four possible learning and decision paradigms is important in multimodal biometrics research since no such information has been available.

The paper is organized as follows: In Section II, the problem of multimodal biometric decision fusion is defined before several possible problem treatments are stated. In Section III, the concept of local learning and decisions is introduced and illustrated. Several arising issues like learning from small sample data, learning of local decision hyperplanes, and setting of local thresholds are addressed in the same section. In Section IV, a decision model that requires only single-step training is introduced. This decision model is deemed to be suitable for both local and global decisions learning. Three biometrics are then introduced in Section V before experimental results are reported in Section VI. In Section VII, some concluding remarks are drawn.

II. Problem Definition and Preliminaries

A. *Problem of Multimodal Biometric Decisions Fusion*

Given (l, m, n, p) as positive integers, and consider the following sets of data obtained from comparing two identities: a training set $\mathcal{S}_{\text{train}} = \{\boldsymbol{r}_i \in \Re^p, s_i \in \Re\}, i = 1, \ldots, m$ and a test set $\mathcal{S}_{\text{test}} = \{\boldsymbol{r}_i \in \Re^p, s_i \in \Re\}, i = 1, \ldots, n$, where $\boldsymbol{r}$ and s denote the feature vector and the class inference, respectively.

Given a set of biometric authentication decisions $\mathcal{F} = \{\hat{f}_j(\boldsymbol{r})\}$, $j = 1\ldots, l$, where each of its elements $\hat{f}_j(\boldsymbol{r})$ approximates a true *decision function* $f(\boldsymbol{r}) = s$ (assuming it exists) that classifies the data given by $\{\mathcal{S}_{\text{train}}, \mathcal{S}_{\text{test}}\}$ either as *genuine-users* (class C^g) or *imposters* (class C^i). In this treatment, the output of each decision approximation function $\hat{f}_j(\boldsymbol{r})$ (individual modality of biometrics) is also called the *match-score* of the comparison or matching process.

Given a set of raw biometric data (before matching) containing N identities, each with M samples: $\{\boldsymbol{r}_{k,l} \in \Re^p\}$, $k = 1, \ldots, N$, and $l = 1, \ldots, M$. Assuming $i = 1, \ldots, N$ and $j = 1, \ldots, M$, the match-scores for genuine users are generated from *intra-identity matching* among the M samples for each identity, i.e., $s = \hat{f}(\boldsymbol{r}_{i,j}, \boldsymbol{r}_{k,l}) = \hat{f}(\boldsymbol{r})$, $i = k$, $j \neq l$ ($\{\boldsymbol{r}_{i,j}, \boldsymbol{r}_{k,l}\} = \{\boldsymbol{r}\}$ in our notation). On the other hand, the imposter match-scores are generated from *inter-identity matching* across the N users: $s = \hat{f}(\boldsymbol{r}_{i,j}, \boldsymbol{r}_{k,l}) = \hat{f}(\boldsymbol{r})$, $i \neq k$.

Given outputs of some $\mathcal{F}$ and using the known class labels in training set $\mathcal{S}_{\text{train}}$, our problem of *multimodal biometric decision fusion* is to find the best possible authentication decision according to s using this set of $\mathcal{F}$. The set $\mathcal{S}_{\text{test}}$, which has not been used in training, will be used to test the classification performance.

The terms *user-specific*, *personalized*, and *local* for learning and decision will be used interchangeably in our presentation. The terms *user* and *identity* will also be used interchangeably for convenience.

B. *Problem Treatments*

The above problem can be treated as a *pattern classification* problem consisting of two stages of processing: *learning* and *decision making*. In biometric problems, apart from globally classifying the data set into two labels, namely, the *genuine-users* and the *imposters*, a local level of labeling can be imposed on each individual user. The learning and decision making processes can then be applied to this local level of user labeling to exploit possible improvement in decision accuracy as compared with the usual global-learning-and-decision case. The following learning paradigms are thus possible from this consideration:

a) Learn *globally*, decide *globally* (GG): This is *one learning function* and *one decision threshold* for all users.

b) Learn *globally*, decide *locally* (GL): This is *one learning function* and *multiple decision thresholds* (each individual user corresponds to a decision threshold).

c) Learn *locally*, decide *globally* (LG): This is *multiple learning functions* (each individual user corresponds to a decision function) and *one decision threshold* for all users.

d) Learn *locally*, decide *locally* (LL): This is *multiple learning functions* (each individual user corresponds to a decision function) and *multiple decision thresholds* (each individual user corresponds to a decision threshold).

GG is the most commonly adopted approach to solving the multimodal biometric decisions fusion problem (see, e.g., [2], [3], [7], [13], and [14]). The main idea of this approach is to treat all match-scores from genuine users as one single class that is to be differentiated from those match-scores obtained from imposters that formed the other class. In a more general sense, this decision fusion problem can be treated as a classification problem, separating the class consisting of genuine-user match-scores and the class consisting of imposter match-scores.

GL and LG are relatively new, and thus far, only one article from the multimodal biometrics literature [12] has mentioned them, and their potential has yet to be fully exploited since only simple weights and thresholds have been incorporated.

LL has not been covered in any multimodal biometrics literature to the best of our knowledge. Their usability and robustness are thus not known. In this approach, the genuine-user scores from each individual user are to be separated from the scores from respective imposters, and each user forms his/her own personalized decision hyperplane. In other words, the problem is broken down into N subproblems (suppose there are N users) for the individual users, where each performs his/her own learning and classification of match-scores for the genuine-users from those respective imposters.

In this paper, we will compare the positive and negative aspects of the above four paradigms empirically using three biometrics, namely, fingerprint, speech, and hand-geometry. We will skip discussion on GG as it is quite well known. As many basics come from LL, we will begin with an illustration on LL in Section III before the follow-up issues on local learning and local decision.

III. Learning and Decisions

A. *Illustrative Example*

We begin with an illustrative example. Every individual biometric method has a certain matching characteristic for each identity. When combining different biometric modalities from the same person, each individual identity thus displays a certain trait of the relative matching characteristics (e.g., match-scores) among the modalities. For instance, consider the match-scores of three genuine users and their respective imposters, as shown in Fig. 1(a) (for reasons of simplicity, we do not segregate imposters with respect to each user as the distribution of these subsets of scores is not apparently distinguishable).

When a person has persistent wet fingerprints [see, e.g., user-1 in Fig. 1 and some of his image samples, as shown in Fig. 3(a) and (b)], then his/her fingerprint match-scores (high scores for good matchings) for verifying the same fingers would exhibit a certain distribution, possibly low and scattered due to different degree of wetness recorded during each query. If this person speaks quite clearly such that a low matching error measure is always obtained when he uses the speaker verification (low scores for good matchings) system, then the combination of fingerprint and voice-based verification systems can incorporate the information on the relatively low scores due to fingerprints (low but above a certain recognition threshold)

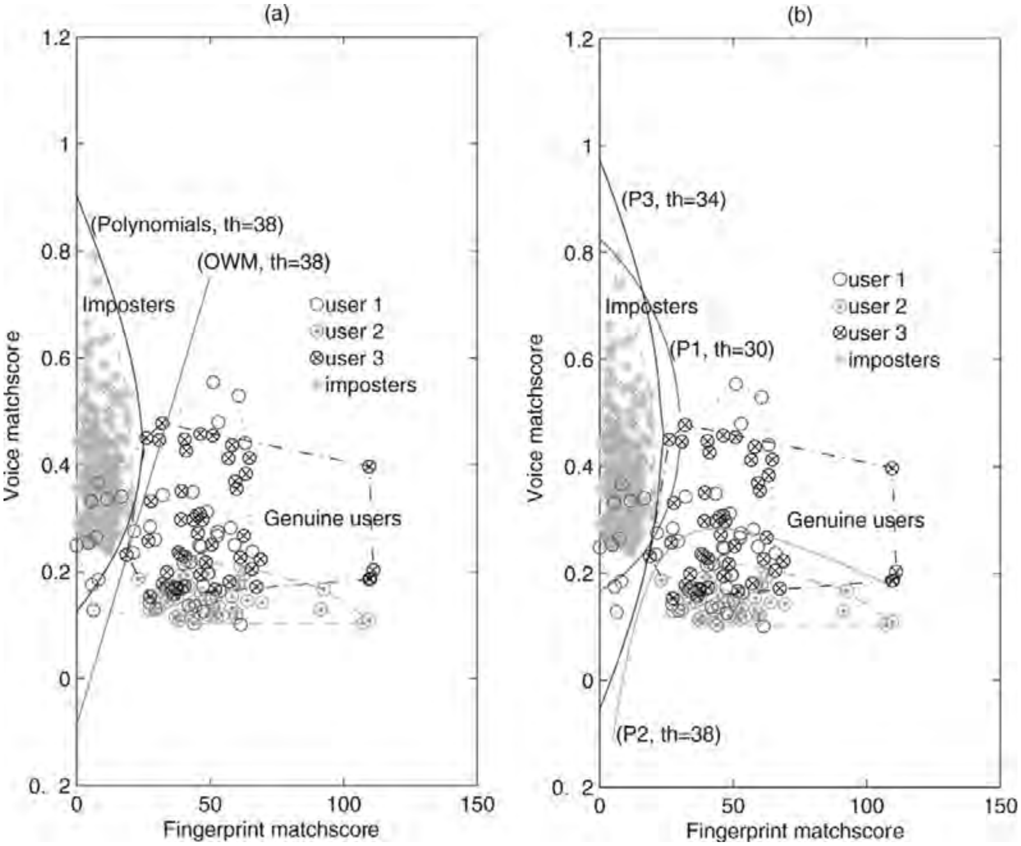

Fig. 1. Distribution of genuine and imposter scores for three different users with (a) nonpersonalized (global) decisions and (b) personalized (local) decisions.

and the information on the relatively low scores due to voice. Similarly, for user-2 and user-3 in Fig. 1(a), their genuine match-scores exhibit different clustering.

Instead of a linear decision hyperplane, we can learn a nonlinear decision hyperplane specific to the individual user for a better description of these individual users. The quest for a nonlinear decision hyperplane can be gathered from Fig. 1(b), where we adopt a personalized polynomial (P1-P3) for each user. A full bivariate polynomial model of order 2 was used for all three users for this illustration. The resulting separation curves at three different indicated threshold levels are shown as (P1, $th = 30$), (P2, $th = 38$), and (P3, $th = 34$), respectively, for the three individual users. It can be seen from this figure that both genuine-user scores from user-2 and user-3 are linearly separable from the imposter scores, and hence, a nonlinear separation curve may not be needed. However, for user-1, a nonlinear decision curve is preferred over a linear one since it provides better classification accuracy. It is noted that in this example of three users, the personalized (local) decision, as shown in Fig. 1(b), as compared with the nonpersonalized (global) decision, as shown in Fig. 1(a), has resulted in a reduction of number of false rejects by two for user-1. In both cases of personalized and nonpersonalized decisions, a nonlinear decision is preferred over a linear one [OWM in Fig. 1(a) is the linear optimal weighting method, as seen in [15]].

Having illustrated the idea of localized learning and decisions, we are ready to look into several issues that arise from this localized formulation. The first issue is the learning data available, i.e., we may not have the two generally large groups of imposter and genuine-user scores as in the nonpersonalized case. The second issue is the learning methodology to be used for this data structure. The third issue is the threshold settings under such a learning framework. We will follow up in Sections III-B–D for a more detailed discussion about these issues and show how we can exploit the merits of such personalized formulation.

B. Learning from Small Sample Data

In the nonpersonalized case, usually, two large sets of data are available to learn the decision hyperplane. These two data sets are the genuine-user scores and the imposter scores. The genuine-user scores are obtained from *intra-matchings* among multiple biometric samples taken from the same identity, whereas the imposter scores are obtained from *inter-matchings* among the biometric samples taken from different identities. Learning in this nonpersonalized case is thus aimed at forming a decision hyperplane based on these two aggregate clusters of match-scores taken from every person where the total data size for each class is relatively large (usually more than a few hundred match-scores for the genuine-user class in the set C^g and more than a few thousand match-scores for the imposter class in the set C^i in a typical application).

In contrast to the above nonpersonalized case, the personalized case may not enjoy the luxury of such a large data size. Under a typical application scenario, a user may have only a few sample biometric data enrolled. The size of genuine-user scores is thus limited to the intra-matchings among these few enrolled samples from the same person. As for the imposter scores, the data is obtained from inter-matching the selected user across all other users. Differing from the nonpersonalized case above, these imposter scores corresponding to a selected user do not include imposter scores that are unrelated to him (i.e., the imposter scores are generated using only other users). The size of imposter scores in the personalized case is thus smaller than that in the nonpersonalized case, although it may not be as small as those for genuine-users. The main problem here is that this unbalance of different classes may affect the density-based training. Another possible problem is that the small number of genuine-user score samples may not be representative enough for possible large variations during query applications.

To circumvent these problems, rather than reducing the already limited imposter samples, we propose to include additional "noisy" samples into the genuine-user class for each user in order for all available data to be exploited. From our empirical studies, a few percent of Gaussian noise (with respect to the original genuine-user match-scores) centered about each original genuine-user scores would be useful. The total number of genuine-user samples including "noisy" and original "nonnoisy" points could be chosen to approach a certain fraction of the number of imposter samples. We will provide more details in the experiment section.

The robust parameter estimation by different kinds of noise injection is out of the scope of this work. Therefore, we use a simple way to do noise injection. To find out more on the effect of different kinds of noise injection on the generalization improvement, see [16] and references within.

C. Learning of Local Decision Hyperplanes

Learning the personalized parameters for each user in a medium/large database calls for an efficient and accurate learning methodology as well as a small parameter set for each user. This renders many iterative learning methods that are not suitable since they take up a lot of effort during the learning stage.

In machine learning, a plausible way to handle this problem is to use SVMs to learn the nonlinear discrimination hyperplane, which optimizes the class boundary separation rather than the usual error objective. However, for nonseparable classes, the use of SVMs requires special treatment, which is nontrivial. Moreover, training of SVMs for the nonlinear discriminant hyperplane, which usually results in a constrained quadratic formulation, requires an iterative learning procedure that does not guarantee global optimal solution when the formulation is nonlinear.

The feedforward neural network provides good approximation and classification accuracies as it has been shown to be a universal approximator (see, e.g., [17] and [18]). However, training of the network requires an iterative process whereby its initialization for good accuracy is usually a trial-and-error game [19]. The RBF network (RBFN) (e.g., [20]) has been widely used for approximation due to its structural simplicity. Typically, training of the RBFN involves selection of the hidden-layer neuron centers, choosing the width parameters, and estimation of the weights that connect the hidden and the output layers. Although the weights can be estimated using the linear least squares algorithm, once the centers and width parameters are fixed, selection of these centers and width parameters remains a nontrivial task.

In view of these problems, we introduce a decision model that does not require iterative learning and, at the same time, provides good classification capability. Since the formulation and derivation is rather involved, we will present this decision model in a separate section (see Section IV).

D. Setting of Local Thresholds

As mentioned in the problem definition section, there are two ways in which the thresholds can be set: many local thresholds and a global threshold. Since setting of global threshold can be based directly on the ROC for require recognition rates, we will pay attention to the local thresholds setting.

The ROC curve is not directly obtained in a straightforward manner as in the case of the local thresholds setting. In this work, we propose an approach to obtain a comparable ROC performance for user-specific threshold settings and, hence, a measure to select these user-specific thresholds.

Let $\mathcal{A}_k$ be the set containing $k = 1, \ldots, N$ users with normalized genuine scores $\{a_{ki} \rightarrow 1\}_{i=1,\ldots,m}(a_{ki} = \hat{f}(\boldsymbol{r}_{kj}, \boldsymbol{r}_{kl}), j \neq l)$, and let $\mathcal{B}_k$ be the set containing the corresponding normalized imposter scores $\{b_{kj} \rightarrow 0\}_{j=1,\ldots,n}(b_{kj} = \hat{f}(\boldsymbol{r}_{kj}, \boldsymbol{r}_{il}), k \neq i)$. m and n are, respectively, the total number of intra/inter matching among/across the N users.

Define midpoint$_k$ as the average match-score of $\min(a_{ki})$ and $\max(b_{kj})$ for the kth user:

$$\text{midpoint}_k = \frac{\min(a_{ki}) + \max(b_{kj})}{2}$$
$$i = 1, \ldots, m; \quad j = 1, \ldots, n. \tag{1}$$

In addition, define the following minimum and maximum quantities:

$$\text{minpoint}_k = \min(a_{ki}), \quad i = 1, \ldots, m \tag{2}$$
$$\text{maxpoint}_k = \max(b_{kj}), \quad j = 1, \ldots, n. \tag{3}$$

Based on available training data, we then have the following baseline local threshold settings $\forall\, k = 1, \ldots, N$:

L1: $th_k = \text{midpoint}_k$
L2: if $(\text{minpoint}_k < \text{maxpoint}_k)$ then $(th_k = \text{minpoint}_k)$ else $(th_k = \text{midpoint}_k)$
L3: if $(\text{minpoint}_k < \text{maxpoint}_k)$ then $(th_k = \text{maxpoint}_k)$ else $(th_k = \text{midpoint}_k)$
L4: if $(\text{minpoint}_k < \text{maxpoint}_k)$ then $(th_k = \text{maxpoint}_k)$ else $(th_k = \text{minpoint}_k)$
L5: if $(\text{minpoint}_k < \text{maxpoint}_k)$ then $(th_k = \text{minpoint}_k)$ else $(th_k = \text{maxpoint}_k)$
L6: $th_k = \text{mean}(a_{ki}), i = 1, \ldots, m$
L7: $th_k = \text{mean}(b_{kj}), j = 1, \ldots, n$
L8: $th_k = \text{median}(a_{ki}), i = 1, \ldots, m$
L9: $th_k = \text{median}(b_{kj}), j = 1, \ldots, n$.

Considering all users, for each point given by *L1* through *L9*, this acts as a reference at one point on the ROC curve. These thresholds th_k are then varied concurrently for every identity with similar equal value of *incremental* and *decremented* steps such that a full range of ROC performance can be obtained.

The final ROC is obtained by summing the error rates across all identities for each threshold step.

Threshold settings using baselines *L1–L3* basically used the midpoint_k, $k \in \{1, \ldots, N\}$ as reference for separable cases (separation between the scores of genuine-users and imposters) and used the boundary points (minpoint_k or maxpoint_k) for nonseparable cases. Baselines *L4–L5* used only the boundary points (minpoint_k or maxpoint_k) as reference for separable cases wherein emphasis is placed reversely for nonseparable cases. The main purpose of these "reversing" of thresholds under the overlapping regions is to strike a balance between the false alarm rate (FAR) and the false reject rate (FRR). For baselines *L6–L9*, some statistical aspects of distribution (i.e., mean and median) for both genuine-users and imposters were considered.

Bearing in mind that we only have training data for the above threshold settings, we have no "optimal" means on which to base the selection of the best combination such that the test performance is also good. It is therefore the aim of this work to evaluate empirically which of the above baseline threshold settings is suitable for practical use. In an application scenario, the best baseline threshold setting can be used in a cross-validation test to choose an appropriate operating point.

IV. Decision Model

In this section, we first recall the multivariate polynomials decision model before introducing a reduced model for decisions fusion. This is because the least squares solution form for parameters estimate in the multivariate polynomials decision model can be applied in a straightforward manner in subsequent development with minor modification. To simplify the expression as well as to avoid possible confusion, the notation for individual biometric classifiers $\hat{f}_j(\boldsymbol{r})$, $j = 1, \ldots, l$ to be combined will be replaced by x_j, $j = 1, \ldots, l$ as polynomial inputs. For example, in the bivariate case, x_1 represents the match-score of biometric-1, and x_2 represents the match-score of biometric-2, and a weight parameter α is attached to each polynomial expansion term, which will be described below.

A. Multivariate Polynomial Model

The general multivariate polynomial model can be expressed as

$$g(\boldsymbol{\alpha}, \boldsymbol{x}) = \sum_{i}^{K} \alpha_i x_1^{n_1} x_2^{n_2} \cdots x_l^{n_l} \tag{4}$$

where the summation is taken over all non-negative integers $n_1, n_2, \ldots, n_l$ for which $n_1 + n_2 + \cdots + n_l \leq r$ with r being the order of approximation. $\boldsymbol{\alpha} = [\alpha_1, \ldots, \alpha_K]^T$ is the parameter vector to be estimated, and $\boldsymbol{x}$ denotes the regressors vector $[x_1, \ldots, x_l]^T$. K is the total number of terms in $g(\boldsymbol{\alpha}, \boldsymbol{x})$.

Without loss of generality, consider a second-order bivariate polynomial model ($r = 2$ and $l = 2$) given by

$$g(\boldsymbol{\alpha}, \boldsymbol{x}) = \boldsymbol{\alpha}^T p(\boldsymbol{x}) \tag{5}$$

where $\boldsymbol{\alpha} = [\alpha_1\ \alpha_2\ \alpha_3\ \alpha_4\ \alpha_5\ \alpha_6]^T$, and $p(\boldsymbol{x}) = [1\ x_1\ x_2\ x_1^2\ x_1 x_2\ x_2^2]^T$.

Given m training data points with $m > K$ and using the regularized least squares error minimization objective given by

$$s(\boldsymbol{\alpha}, \boldsymbol{x}) = \sum_{i=1}^{m} [y_i - g(\boldsymbol{\alpha}, \boldsymbol{x}_i)]^2 + b\|\boldsymbol{\alpha}\|_2^2 \tag{6}$$

the parameter vector $\boldsymbol{\alpha}$ can be estimated from

$$\boldsymbol{\alpha} = (\mathbf{P}^T\mathbf{P} + b\mathbf{I})^{-1}\mathbf{P}^T\mathbf{y} \tag{7}$$

where $\mathbf{P} \in \Re^{m \times K}$, $\mathbf{y} \in \Re^{m \times 1}$ and $\mathbf{I}$ is a $(K \times K)$ identity matrix. $\|\cdot\|_2$ denotes the l_2-norm, and b is a regularization constant. $\mathbf{P} \in \Re^{m \times K}$ denotes the Jacobian matrix of $p(\boldsymbol{x})$:

$$\mathbf{P} = \begin{bmatrix} 1 & x_{1,1} & x_{2,1} & x_{1,1}^2 & x_{1,1}x_{2,1} & x_{2,1}^2 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_{1,m} & x_{2,m} & x_{1,m}^2 & x_{1,m}x_{2,m} & x_{2,m}^2 \end{bmatrix} \tag{8}$$

and $\mathbf{y} = [y_1, \ldots, y_m]^T$ is the known inference vector from training data. To avoid a large difference among the output values that may cause numerical ill-conditioning, a normalized output $y \in [0, 1]$ will be used. In (8), the first and second subscripts of the matrix elements $x_{j,k} (j = 1, 2,\ k = 1, \ldots, m)$ indicate the dimension of polynomial input and the the number of instances, respectively. The estimated parameter can then be used in the model given by (5) for the prediction of future class labels.

B. Reduced Polynomial Model

The above multivariate polynomial expansion provides a natural platform for spanning a wide variety of combination of product and power terms. However, for a full interaction multivariate polynomial model, as shown in (4), the number of terms or parameters becomes very large for high-dimensional and high order problems. Multivariate polynomial expansion becomes impractical due to this explosive number of product terms. In view of this problem, we introduce a reduced model whose number of parameters do not increase exponentially, yet preserve the necessary variety of combinations for the desired approximation and classification capabilities.

In [13], a reduced polynomial model is proposed to combine two biometrics. The model has been shown to have good verification accuracy as compared with several conventional decision fusion methods. Starting from a *multinomial* model and based on Taylor's first-order approximation with appropriate omittance and addition of certain nonlinear terms (see [13] for details), a reduced multivariate polynomial regressor model (RM) is obtained as

$$\begin{aligned} g_{\mathrm{RM}}(\boldsymbol{\alpha}, \boldsymbol{x}) = {}& \alpha_0 + \sum_{k=1}^{r}\sum_{j=1}^{l} \alpha_{kj} x_j^k \\ & + \sum_{j=1}^{r} \alpha_{rl+j}(x_1 + x_2 + \cdots + x_l)^j \\ & + \sum_{j=2}^{r} \left(\boldsymbol{\alpha}_j^T \cdot \boldsymbol{x}\right)(x_1 + x_2 + \cdots + x_l)^{j-1}. \end{aligned} \tag{9}$$

The number of terms in this model can be expressed as $K = 1 + r + l(2r - 1)$. The plots for the number of terms over model

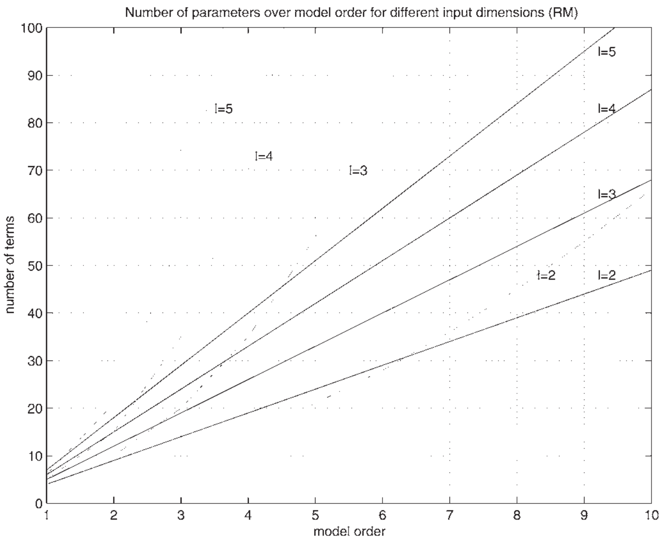

Fig. 2. Number of terms plotted over model order for different input dimensions (dashed line: full multivariate polynomials, continuous line: RM).

order for each input dimension of the RM is shown in Fig. 2. For comparison purposes, the plot includes the exponential curve for those corresponding full multivariate polynomial models.

V. Fingerprint, Speech, and Hand Biometrics

A. Fingerprint Verification

In general, an automatic fingerprint identification or verification (see, e.g., [21]–[24]) system consists of three main processing stages, namely, *image acquisition*, *feature extraction*, and *matching*. In image acquisition, query and template database images are acquired through various input devices. Development over the years has seen through use of devices that mechanically scan the ink based fingerprints into the computer system to invention of devices that directly capture the fingerprints using sophisticated solid-state sensors. With fingerprint images that could be distorted or contaminated with noise, the automated system seeks to *extract* characteristic *features* that are discriminating for different fingers and yet invariant with respect to image orientation for the same fingers. The final stage of fingerprint identification is to search and verify matching image pairs.

Our representation for the fingerprint consists of a global structure and a local structure. The global structure consists of positional and directional information of ridge endings and ridge bifurcations. The local structure consists of relative information of each detected minutia with other neighboring minutiae. Fingerprint verification is then performed by comparing the minutia information between two templates [25]. Fig. 3 shows four samples of fingerprint images with detected minutiae and area-of-interest segmentation. See [25] and [26] for details-of-minutia detection and matching.

B. Speaker Verification

Speaker verification seeks to determine whether an unknown voice matches the known voice of a speaker with a known identity. It is a subset of the more general problem of speaker recognition that includes the task of speaker identification (see, e.g., [27]). Operation of the above systems can either be in fixed-text mode or in free-text mode. In fixed-text mode, a predetermined text is required to be recited for reliable comparison, whereas in free-text mode, speech utterances of unrestricted text can be accepted. The fixed-text system provides a more precise and reliable comparison between two utterances of the same text than that of the free-text system since it works under a better controlled environment. The fixed-text system is thus primarily used in access control applications, and the free-text systems are more for surveillance and other applications [27].

In this application, the fixed-text mode and the template matching method is adopted for speaker verification. Comparison of two utterances is performed by aligning the two templates at corresponding points in time. To cater to the difference in duration of the two utterances, the dynamic time warping (DTW) method is adopted when minimizing a distance metric between two feature sets extracted from the speech data. Fig. 4 shows some samples of voice data uttering the word “zero.” See [28] for more details about the system (see also [27] and [29] for similar matching designs).

C. Hand-Geometry Verification

The hand geometry is considered to achieve medium security as compared with fingerprint technology [30]. However, it has several advantages over the use of fingerprints, namely, low computational cost, lack of relation to criminal records, and no imaging problems due to hand’s wetness. These features will be exploited to complement the high accuracy feature of fingerprint systems in this fusion development for a robust and yet highly secure system. The problem of having a finger being too wet or too dry for a fingerprint image capture will be resolved without compromising the high security requirement.

A light box was used to flush the background of the captured images such that a sharp edge could be obtained for the hand geometry. Segmentation on this back-lit image became simple as the contrast was high. A pair of hand shape contours can be compared using the contour string-match method [31] or based on the so-called “handcrafted features” [32]. Typical handcrafted features include the length and the width of each finger, aspect ratio of the palm or fingers, thickness of the hand, finger perimeter and areas, and so on [32].

In our current application, we use the width and length information. First, the hand contour is analyzed, and dominant points are located [33]. These points are further identified as finger tips and valleys based on the convex or concave curvature of the contour. The principal axis of each finger is then found using a set of equal separated grid points starting from the respective finger tips. The widths are measured perpendicular to the axes at the grid points. The features used were similar to those in [32], except that a fixed interval was used for the width measurements with a total of 15 to 30 width features being collected for each hand image, depending on the finger length. The length is found using the fingertip and its neighboring valley information. These features of each finger from both the query image and the template image are compared separately. Their absolute matching differences are summed up and normalized as the matching score. Fig. 5 shows a sample captured hand image and

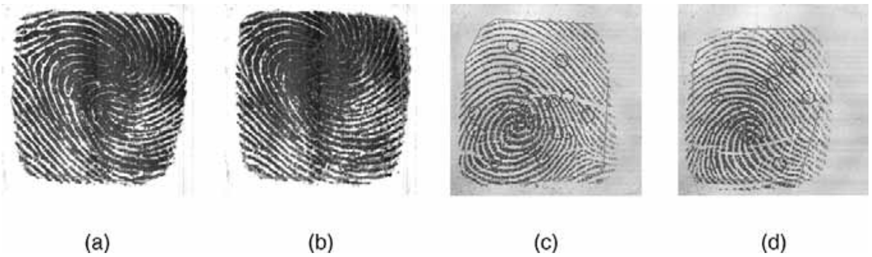

Fig. 3. Fingerprint image samples. (a), (b) Wet fingers and (c), (d) normal fingers.

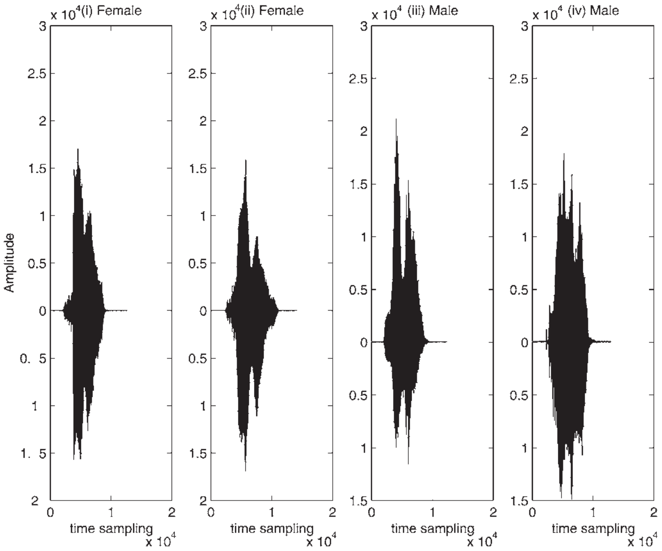

Fig. 4. Voice samples.

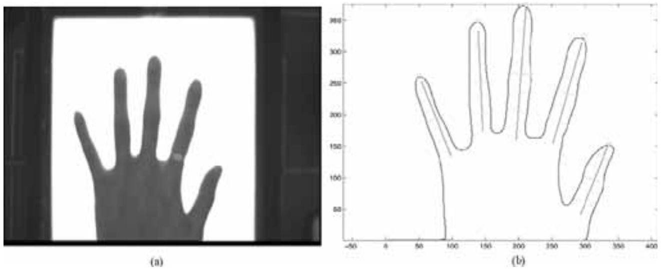

Fig. 5. (a) Hand image sample. (b) Extracted hand geometry.

its extracted hand-geometry including width/length features in our application.

VI. Experiments

A. Data Sets

Using the data from the above three biometrics, we perform two sets of experiments: One combines only two biometrics (fingerprint and speech data), and the other one uses all three biometrics. Our purpose is to observe the possible effect due to the different number of modality on the verification performance. The fingerprint images were collected using Veridicom's iTouch sensor, and the voice data were taken from TIDIGIT database. The images for hand geometry were collected using the mentioned light box setup with a sharp hand image edge.

In the following experiments, each data set corresponding to fingerprint verification, speaker verification, and hand-geometry verification consists of 96 identities, with each identity containing ten samples. For training and test purposes, each of these biometric data sets are randomly partitioned into two equal sets consisting of $\mathcal{S}_{\rm train}$ and $\mathcal{S}_{\rm test}$, each with 96×5 samples. The genuine-user and the imposter match scores are generated from these two sets by intra-identity and inter-identity matching among the image/voice samples for each biometric. A total of 960 ($96 \times 5 \times 4/2$) sample match scores is thus available for the genuine-user class in each training and test set for each biometric. As for the imposter scores, there are 228 000 ($96 \times 95 \times 5 \times 5$) sample match scores for the 96 identities. Since all three biometrics have the same number of genuine-user and imposter samples, an arbitrary one-to-one identity correspondence was assumed among the three biometric data sets. For the user-specific case, we have 10 (960/96) genuine-user samples and 2375 (228 000/96) imposter samples corresponding to each user for each biometric.

B. Preprocessing and Settings

Depending on individual implementation, the matching output ranges for different modalities may differ significantly. For such cases, numerical sensitivity may be affected, and hence, a score normalization should be performed between the outputs of different modalities. In our experiments, the match scores for all biometrics are normalized using the largest output value from $\{\mathcal{S}_{\rm train}, \mathcal{S}_{\rm test}\}$ to within the interval $[0, 1]$ before performing data fusion.

Fig. 6(a)–(f) shows the matching performances for the training and test sets, respectively, for individual fingerprint, speaker, and hand-geometry verifications, before multimodal fusion. The plots show the original nonnormalized match scores.

From the match-score distribution plots shown in Fig. 6(a), (c), and (e), the verification performance depends much on the overlap between the genuine-user and the imposter classes. For the case of fingerprint data, this overlap in match scores depends on the minimum value of the genuine-user scores and the maximum value of the imposter scores. For speech and hand-geometry data, since match scores correspond to matching errors, this overlap is due to the crossing of the maximum value of the genuine-user score and the minimum value of the imposter scores. For convenience, we will call these maximum and minimum values as *boundary match-scores*.

The boundary match-scores distribution corresponding to the 96 individual identities for both genuine-user and imposter classes for each biometric are shown in Fig. 7. We will observe

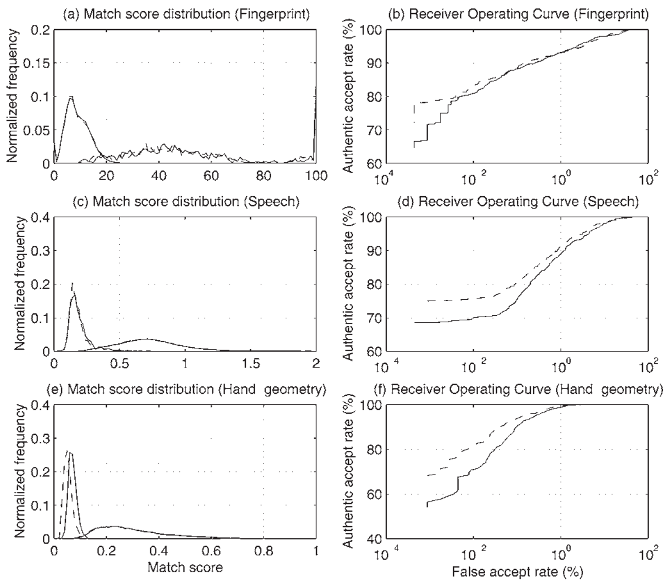

Fig. 6. Matching performance for fingerprint, speech, and hand-geometry verification systems: Training (dashed lines) and test (continuous lines) sets.

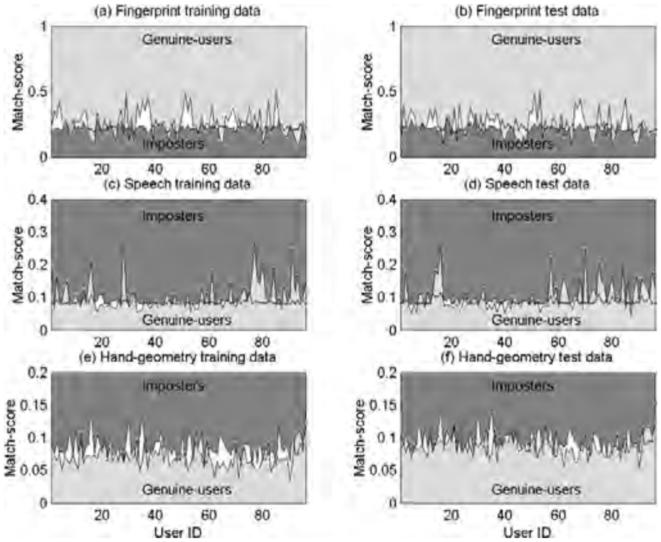

Fig. 7. Boundary match-scores distribution for fingerprint, speech, and hand biometrics.

in the following experiments how such a decision boundary varies with different learning settings. In all the following experiments, we set $r = 3$ and $b = 10^{-4}$ for the reduced model since we found this setting produces good training and test results from our empirical studies.

As mentioned in Section III-B, we always face the problem of an unbalanced proportion of data for different classes in biometric applications. The ratio of the number of genuine users over the number of imposters is 1:237.5 in our experiment. For the training data, 3% noise with respect to the largest magnitude of the match scores was added to the ten samples of genuine-user scores, and the total number of genuine-user samples, including the original ones, added up to 100 samples. This reduces the ratio to 1:23.75.

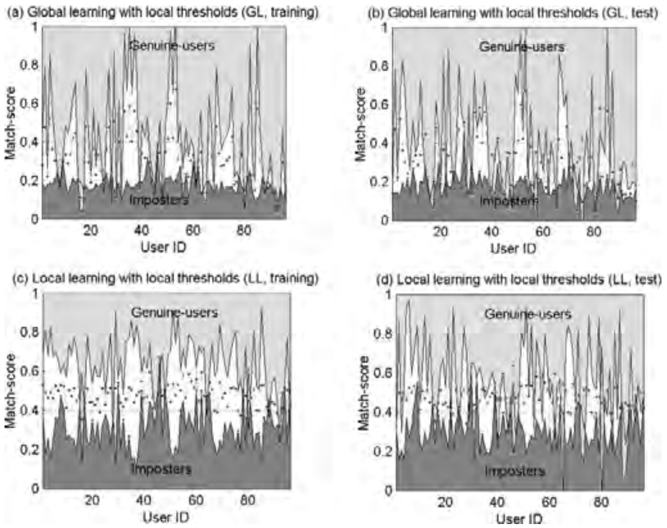

Fig. 8. (a), (b) Boundary match-scores distribution for global learning and (c), (d) local learning of two biometrics.

C. Combining Two Biometrics: Fingerprint and Speech

1) Global Learning with Global Decision (GG): In this experiment, the commonly used GG paradigm is applied for fingerprint and speaker verification decision fusion. Fig. 8(a) and (b) shows the boundaries of the globally learned combined system for both the genuine and imposter classes. It is clear from this figure that the class separation has been widened in the combined system as compared with those in the single biometric, as shown in Fig. 7.

The GG paradigm has been employed in most applications. In [13], the GG paradigm was compared with several commonly used decision fusion methods, namely, SVM, neural networks, Naive–Bayes, and OWM. The authors showed that the RM model in (9) has either superior or comparable training and test ROC performances over these commonly used decision fusion methods. In this experiment, we enlarge the data set especially for the imposter class for user-specific applications. The ROC performance to combine the fingerprint and speech biometrics is shown in Fig. 10 (curve labeled as GG). As many comparative studies with those conventional methods has been shown in [13], and to remain focussed on this work, we will compare GG learning (using learning model RM) only with the other three learning and decision paradigms (GL, LG, and LL). We will move on to describe the results for these three paradigms before summarizing the comparative results at the end of this section.

2) Global Learning with Local Decision (GL): The local threshold settings (*L1–L9*) described in Section III-D are applied for ROC performance evaluation. Fig. 9(a) and (b) shows the ROC performances for these nine settings of GL on two biometrics (labeled as GL1–GL9 here). Here, we see that the best overall training ROC performances go to GL4, GL3, and GL1 [see Fig. 9(a)]. However, for the test performance as seen in Fig. 9(b), the best results goes to GL9, GL7, and GL5, which have closely comparable ROC curves. The sharp drop in performance for GL4 from training to test can be understood from the globally learned boundary match score plots, as shown in Fig. 8(a). Here, the test boundary match scores for the genuine

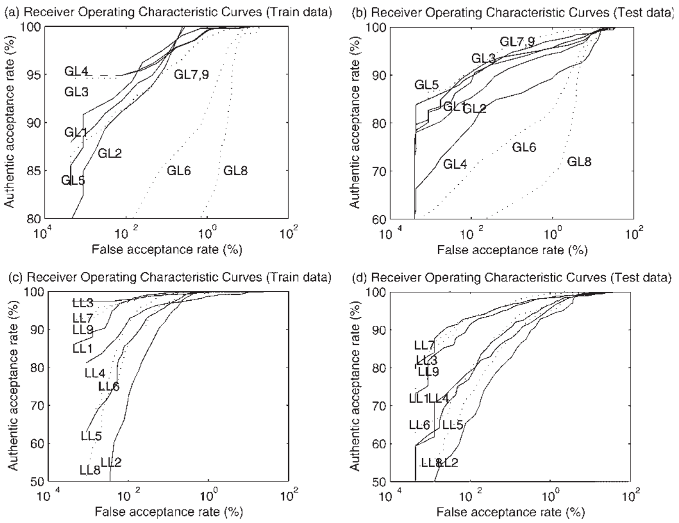

Fig. 9. ROC curves for nine settings of (a), (b) GL on two biometrics, (c), (d) LL on two biometrics. L1–L5: continuous lines. L6–L9: dotted lines.

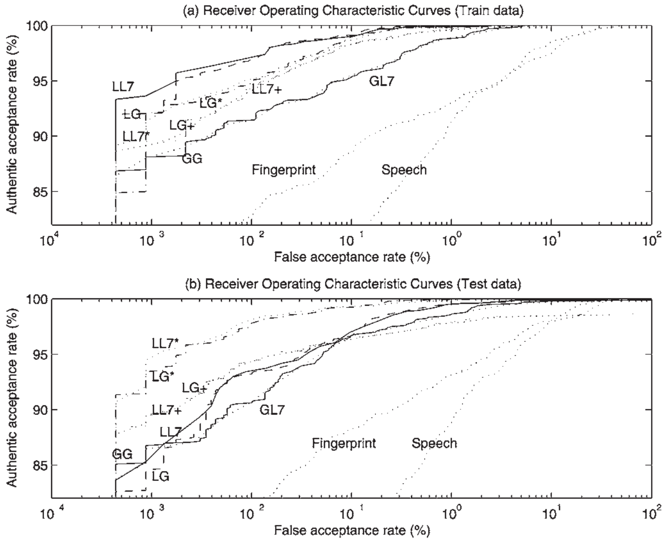

Fig. 10. ROC curves for GG (continuous), GL7 (dotted), LG (dashed), LG* (dashed-dotted), LG+ (dashed), LL7 (continuous), LL7* (dotted), and LL7+(dotted) from two biometrics (fingerprint and speech).

users are seen to vary significantly from corresponding training samples, and GL4 capitalizes its thresholds on these much varied decision boundary for the genuine users. Conversely, GL5 has its baseline set at the less-varied decision boundary for the imposters, and therefore, it has better test performance.

The fact that GL7 and GL9 have better ROC results than GL6 and GL8 shows that the imposters have relatively stable match score distribution than that of genuine users across the training and test data sets.

One of the local thresholds (*L1*) based on the above globally trained data is plotted in Fig. 8(a) and (b) as a dotted line between the boundaries of the genuine and imposter scores. One of the best test results (GL7) from the nine settings is plotted in Fig. 10 along side with other paradigms for comparison.

3) Local Learning with Global Decision (LG): The local learning method, adopting the RM model from Section IV-B, is applied in this experiment. As the model order was chosen as $r = 3$ for every identity, we have 14 weight parameters in RM for each identity. This corresponds to a total of 96×14 weight parameters in the local learning system as compared with only 14 weight parameters in the previous global learning system. Having learned locally, a global threshold is set for all identities. The local learning boundary match score distribution is shown in Fig. 8(c) and (d). It is seen from this figure that the decision boundary for imposters is more varied than it was in GG. However, the decision boundaries between the genuine users and imposters are widen for many identities. The ROC plot for LG is shown in Fig. 10 along side with other learning paradigms for comparison.

4) Local Learning with Local Decision (LL): Both the local learning method, using the RM model from Section IV-B, and the local threshold settings (*L1–L9*) described in Section III-D are applied in this experiment. Fig. 9(c) and (d) show the ROC plots for the nine different threshold settings (labeled as LL1–LL9). It is seen from this figure that LL7, LL3, and LL9 have the best performances for both training and test data among the nine different threshold settings. This is followed by LL1, LL4, LL6, LL5, LL8, and, last, LL2 considering both training and test data. The relative better performances of LL3 over LL2 and LL4 over LL5 (for some regions only in the test data) indicate that the boundary imposter scores have relatively stable distribution over the overlapping regions. Similar to those GL cases, the relatively stable imposter match score distribution is also reflected in LL7 and LL9 over LL6 and LL8. One of the best results (LL7) is plotted in Fig. 10 as a comparison with other learning paradigms.

5) Local Learning Using kNN Partitioned Data (LG+, LL7+, LG and LL7*):* In order to observe the effect of adding noisy samples to the genuine-user scores under the condition of well-represented training and test data sets, two additional experiments are conducted for the two local learning paradigms (LG and LL7). Here, we partition the data into training and test sets by clustering. The 20 genuine scores for each user are classified into ten clusters by a simple k-means clustering method. The nearest sample to each cluster center is chosen as the training sample, and the remaining ten samples constitute the test set for a user. The obtained training set usually should be more representative than that obtained from random selection. The first experiment is the original partitioning using k-means without addition of noisy samples (labeled with a "+" attached to the learning paradigm acronym) and the second experiment with addition of noisy samples (a total of 100 genuine-user samples as in previous cases, labeled with a "*" attached to the learning paradigm acronym). The results for LG+, LL7+, LG*, and LL7* are plotted in Fig. 10 for comparison.

With this partition of the training and test sets, the experiments show that LG(+,*) and LL7(+,*) consistently outperform GG and GL7 for both the training and test data. Although in practice we cannot guarantee to have the representative training data, this experiment shows that LL will outperform GL, even for a small training data set if it is representative. It is also seen that test performance for the case using enlarged noisy samples (LG* and LL7*) is significantly (and consistently) better than that of the case using the original genuine-user scores (LG+ and LL7+). This shows that noisy samples can improve the situation even for well represented data.

TABLE I
COMBINING TWO AND THREE BIOMETRICS: PERCENTAGE ERROR RATES

	Two Biometrics						Three Biometrics					
	Training (%)			Test (%)			Training (%)			Test (%)		
Methods	FRR	FAR	EER	FRR	FAR	EER	FRR	FAR	EER	FRR	FAR	EER
RM-GG	0	5.2912	1.1109	0	100	1.2910	0	0.2316	0.1317	0	0.5127	0.1838
RM-GG	17.8125	0		24.0625	0		5.3125	0		8.0208	0	
RM-GL	0	8.8544	1.0257	0	> 35.3789	1.2874	0	0.3096	0.1368	0	0.9632	0.2182
RM-GL	18.5417	0		24.7917	0		6.5625	0		8.5417	0	
RM-LG	0	1.9939	0.2327	0	99.7947	0.6509	0	0.0101	0.0523	0	0.0702	0.0854
RM-LG	12.2917	0		28.5417	0		0.5208	0		6.6667	0	
RM-LL	0	2.3719	0.2904	0	> 36.7439	0.7394	0	0.0425	0.1044	0	0.8811	0.2165
RM-LL	14.1667	0		27.2917	0		0.2083	0		4.5833	0	

D. *Summary of Results for Combining Two Biometrics*

Table I shows the results for *Equal Error Rate* (EER), *False Reject Rate at zero False Accept Rate* ($\mathrm{FRR}_{\mathrm{zeroFAR}}$), and *False Accept Rate at zero False Reject Rate* ($\mathrm{FAR}_{\mathrm{zeroFRR}}$). For *local decisions*, we see that from GG to GL and from LG to LL, there is an improvement of $\mathrm{FAR}_{\mathrm{zeroFRR}}$ for the test results *from about 100% to about 30+%*. As for $\mathrm{FRR}_{\mathrm{zeroFAR}}$, no significant trend is observed. For test EER, no significant trend is seen from the local threshold settings (GL and LL) to those global thresholds (GG and LG). For *local learning*, it is observed that the EER makes *about 50% improvement* when comparing GG with LG and LL.

To summarize, based on the few operating point of EER, $\mathrm{FRR}_{\mathrm{zeroFAR}}$, and $\mathrm{FAR}_{\mathrm{zeroFRR}}$, we do not observe a significant trend for local decisions, except for $\mathrm{FAR}_{\mathrm{zeroFRR}}$, which shows significant improvement from from about 100% to about 30+%. However, the ROC plots revealed that there could be 2%–4% improvement using local threshold at regions of low FAR using LG and LL as compared with GG. For local learning, significant EER improvement of 50% is observed, as compared with that of global learning.

E. *Combining Three Biometrics: Fingerprint, Speech, and Hand-Geometry*

In this experiment, we combine the verification decisions from all the three biometrics described. Similar to previous experiments on two biometrics, the four learning and decision paradigms are compared.

1) Global Learning with Global Decision (GG): Fig. 11(a) and (b) shows the match score boundaries for the globally trained combined system. It can be seen from this figure that the gaps for these decision boundaries have been widen as compared with those using two biometrics. This has resulted in occurrence of more separable cases (separate between genuine-users and imposters), hence, better ROC performance than the two biometrics case. For comparison purposes, the ROC results for global decision (GG) are presented in Fig. 13 along side other paradigms for training and test data.

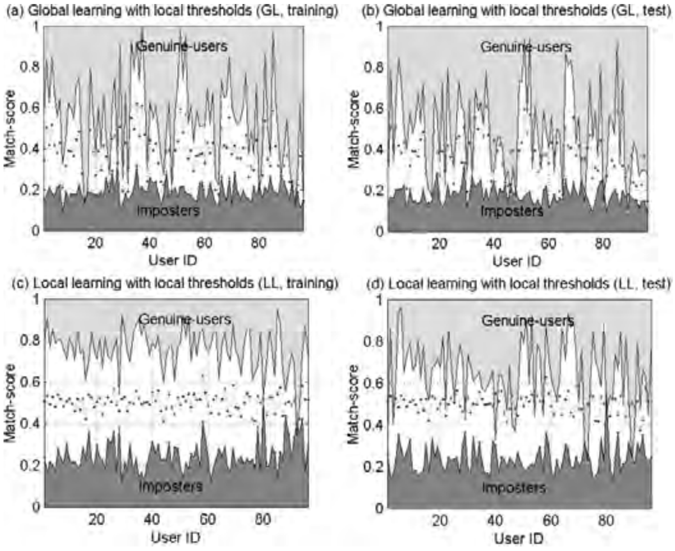

Fig. 11. (a), (b) Boundary match-scores distribution for global learning and (c), (d) local learning of three biometrics.

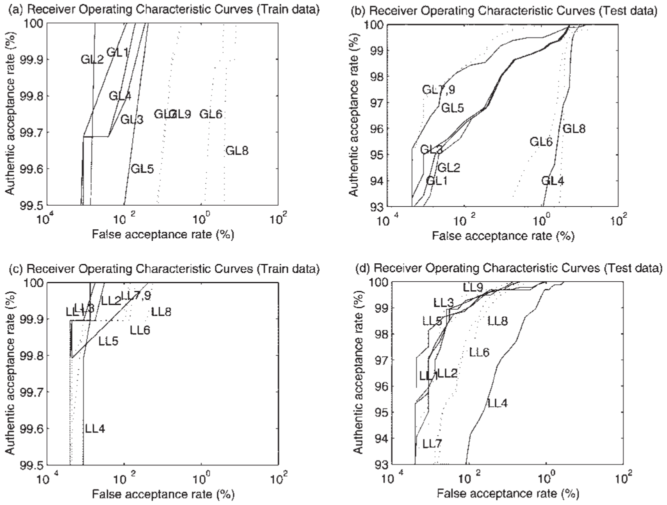

Fig. 12. ROC curves for nine settings of (a), (b) GL on three biometrics and (c), (d) LL on three biometrics. L1–L5: continuous lines. L6–L9: dotted lines.

2) Global Learning with Local Decision (GL): One of the local threshold settings (*L1*) is shown as a dotted line in between the decision boundaries in Fig. 11(a) and (b). Fig. 12(a) and (b) shows the ROC for nine local threshold settings *L1–L9* (labeled as GL1–GL9) described in Section III-D. It is seen from this figure that GL5 has relatively poor ROC performance in training

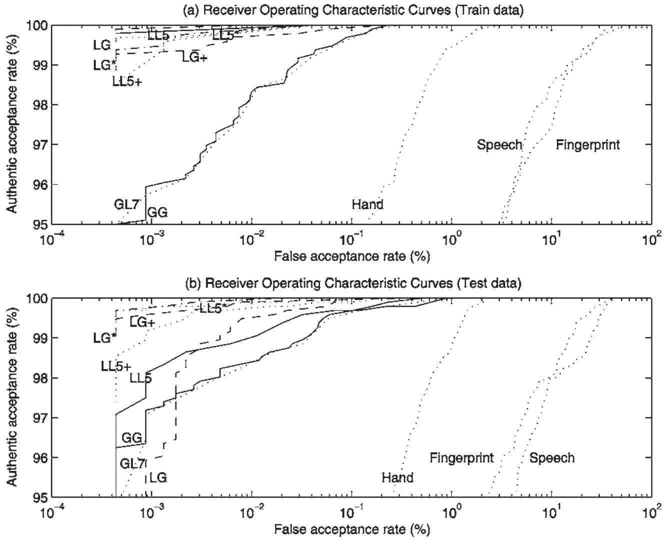

Fig. 13. ROC curves for GG (continuous), GL7 (dotted), LG (dashed), LG* (dashed-dotted), LG+ (dashed), LL5 (continuous), LL5* (dotted), and LL5+(dotted) from three biometrics (fingerprint, speech, and hand-geometry).

but among the best ROC performances in test (comparable to GL7 and GL9). The local threshold settings GL1–GL3 have comparable test ROC performances. Similar to the two biometrics case, GL4, GL6, and GL8 have the worst test ROC performance. One of the best results (GL7) is plotted in Fig. 13 for comparison with other paradigms.

3) Local Learning with Global Decision (LG): The boundary match score distribution for the locally learned system combining three biometrics is shown in Fig. 11(c) and (d). It is seen from this figure that the number of overlapping regions has been reduced as compared with that of global learning case [Fig. 11(a) and (b)]. Obviously, the gaps in between the decision boundaries are widened as compared with that in the two-biometrics case. The globally thresholded ROC performance for this LG paradigm is plotted in Fig. 13 along side with other learning and decision paradigms for comparison.

4) Local Learning with Local Decision (LL): One of the local decisions (*L1*) for the locally learned combined system is shown as a dotted line in Fig. 11(c) and (d). The ROC performances for the nine threshold settings are shown in Fig. 12(c) and (d). It is seen from this figure that LL9, LL3, LL5, LL1, LL2, and LL7 have comparable ROC performance over many operating points. LL5 excels at the low FAR region, and this curve is plotted in Fig. 13 for comparison. Similar to GL above, the LL4, LL6, and LL8 have the worst test ROC performance.

5) Local Learning Using kNN *Partitioned Data (LG+, LL5+, LG***, and LL5***):* Similar to two biometrics, two additional sets of experiments for noisy and non-noisy cases using k-means partitioning are conducted for three biometrics. The results for LG+, LL5+, LG*, and LL5* in Fig. 13 show similar trends of performances to those in two biometrics: local learning over global learning (LL5+, *, and LG+, * over GG and GL7) and noisy over non-noisy (LL5* over LL5+ and LG* over LG+).

6) Summary of Results for Combining Three Biometrics: Table I shows the results for EER, $\mathrm{FRR}_{\mathrm{zeroFAR}}$ and $\mathrm{FAR}_{\mathrm{zeroFRR}}$. For local decisions, we see that from GG to GL and from LG to LL, there is no improvement of $\mathrm{FAR}_{\mathrm{zeroFRR}}$ for the test results, as compared with those in the two-biometrics case. As for $\mathrm{FRR}_{\mathrm{zeroFAR}}$, no significant trend is observed. For test EER, the local threshold settings (GL and LL) are seen to deteriorate from those global thresholds (GG and LG). For local learning, as compared with global learning (LG comparing with GG), a 50% EER improvement is observed. However, different from that in the two-biometrics case, LL does not show improvement of EER here. The reason for the deterioration of LL is due to the low performance of LL5 at this region.

Summarizing the performance plots from Fig. 13, we have LL5 and LG showing significant improvement of about 1% over large FAR regions, where LG has a cross-over of performance at a low FAR region. This 1% improvement is significant because the performance is obtained near perfection (100%).

7) Comments: Although it is obvious that the verification accuracy for all compared fusion methods improves significantly from the addition of the hand-geometry biometric into the fingerprint and speaker verification systems, several observations can be made from Figs. 10 and 13 and Table I.

The experiments using k-means partitioning to generate the training and test data show that LL will outperform GL, even for a small training data set if it is representative. Moreover, the addition of noisy samples to enlarge the genuine-user scores can improve the situation. In short, an overdetermined LL system (i.e., size of training data is larger than the size of parameter vector to be estimated) with representative training data would produce good results. A practical scenario to acquire representative data is to enroll multiple samples for each user, including his highest and lowest possible genuine scores.

Regarding the random data partitioning for both two-biometrics and three-biometrics cases, the LL paradigm shows the best overall ROC performance among the compared paradigms. The LG paradigm has ROC performance that is comparable or better, except for low FAR regions, where the ROC crossed over below that of GG. It is further seen from Table I that *local learning* can achieve about 50% of EER performance improvement for both the two- and three-biometrics cases. This shows that *local learning can have significant improvement on verification accuracy for multimodal biometrics.* However, these performance improvements may not be globally true for the entire ROC as seen from those EER, $\mathrm{FRR}_{\mathrm{zeroFAR}}$, and $\mathrm{FAR}_{\mathrm{zeroFRR}}$ indicators from Table I.

Table II summarizes the overall performances of the nine local threshold settings based on the ROC plots for both the two- and three-biometrics cases. Here, we see that the baseline *L9* is consistently found within the top three performers for all four test cases (GL and LL for two and three biometrics). The baselines *L7* and *L5* are also found to have good performances. It is seen from this table that baseline *L3* is only suitable for the LL paradigm. The baselines *L8*, *L6*, and *L4* are seen to have consistently poor performances for most experiments.

Based on the above observations, we can conclude that with a relatively small amount of learning data (especially for genuine users), *local learning* can provide significant verification accuracy improvement over the conventional decision fusion, which is based on *global learning and global decision* (GG). In addition, the *local decision* (threshold) can have accuracy

TABLE II
SUMMARY OF OVERALL ROC ACCURACY FOR COMBINING TWO- AND THREE-BIOMETRICS USING LOCAL DECISIONS

		Overall ROC Accuracy		
	Methods	Top-3	Medium-3	Bottom-3
Two-Biometrics	RM-GL	**GL9**, GL7, GL5	GL3, GL1, GL2	GL4, GL6, **GL8**
	RM-LL	LL7, LL3, **LL9**	LL1, LL4, LL6	LL5, **LL8**, LL2
Three-Biometrics	RM-GL	**GL9**, GL7, GL5	GL3, GL1, GL2	GL6, GL4, **GL8**
	RM-LL	LL5, **LL9**, LL3	LL1, LL2, LL7	LL6, **LL8**, LL4

improvement when appropriate threshold settings are selected for each user. From the experiments, we found that *L9*, *L7*, and *L5* are good candidates for use in GL and LL paradigms, and *L3* is only suitable for use in the LL paradigm. To combine two biometrics, we can expect about a 2% to 4% improvement in ROC performance over the conventional GG paradigm at an operating range above 85%. To combine three biometrics, we can expect about a 1% improvement of ROC performance over the conventional GG paradigm at an operating range above 95%. For both cases, the EER may be expected to have about a 50% performance improvement.

VII. CONCLUSION

In this paper, we proposed to treat the multimodal biometric decision fusion problem as a two stage problem: learning and decision. Based on this treatment, four learning and decision paradigms were identified with one being new in the literature of multimodal biometrics. Possible issues arising from such formulations were addressed. A reduced multivariate polynomial model was introduced to overcome the tedious recursive learning problem, as seen in neural network training. The learning model is advantageous over the exhaustive weights estimation method proposed by Jain and Ross since it requires only a single learning step, and the solution is least-squares optimal. Moreover, the model can accommodate learning of nonlinear decision hyperplanes. Several baselines for local threshold settings and an ROC performance measure were proposed for local decision making. The four learning and decision paradigms were investigated, adopting the reduced polynomial model for biometric decision fusion. Experiments on fingerprint, speech, and hand geometry biometric data showed that local learning alone can improve verification equal error rates of about 50%. The local decision can have accuracy improvement when appropriate threshold settings were selected for each user. Considering the overall ROC performance, the new LL paradigm was found to be the best among the four learning and decision paradigms studied. We thereby conclude that application of local learning and decision can significantly improve the learning accuracy for high security applications. As current investigation is only limited to verification accuracy, we will extend our future research for possible identification applications.

ACKNOWLEDGMENT

The authors would like to thank Prof. A. K. Jain and the reviewers for their very constructive and insightful comments to improve the paper. The hand geometry data set generated by W. Xiong is also acknowledged.

REFERENCES

[1] L. I. Kuncheva, J. C. Bezdek, and R. Duin, “Decision templates for multiple classifier design: An experimental comparison,” *Pattern Recogn.*, vol. 34, no. 2, pp. 299–314, 2001.

[2] R. Brunelli and D. Falavigna, “Personal identification using multiple cues,” *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 17, pp. 955–966, Oct. 1995.

[3] L. Hong and A. Jain, “Integrating faces and fingerprints for person identification,” *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 20, pp. 1295–1307, Dec. 1998.

[4] S. Prabhakar and A. K. Jain, “Decision-level fusion in biometric verification,” Michigan State Univ., East Lansing, MI, Tech. Rep. MSU-CSE-00-24, 2000.

[5] P. Verlinde and G. Chollet, “Combining vocal and visual cues in an identity verification system using k-NN based classifiers,” in *Proc. IEEE 2nd Int. Workshop Multimedia Signal Process.*, 1998, pp. 59–64.

[6] B. Gutschoven and P. Verlinde, “Multi-modal identity verification using support vector machines (SVM),” in *Proc. 3rd Int. Conf. Inform. Fusion*, vol. 2, 2000, pp. ThB3:3–8.

[7] S. Ben-Yacoub, Y. Abdeljaoued, and E. Mayoraz, “Fusion of face and speech data for person identity verification,” *IEEE Trans. Neural Networks*, vol. 10, pp. 1065–1074, Sept. 1999.

[8] V. Chatzis, A. G. Bors, and I. Pitas, “Multimodal decision-level fusion for person authentication,” *IEEE Trans. Syst., Man, Cybern. A*, vol. 29, pp. 674–680, Nov. 1999.

[9] J. Bigun, B. Duc, F. Smeraldi, S. Fischer, and A. Makarov, “Multi-modal person authentication,” in *Proc. Face Recognition: From Theory to Applicat.*, 1997, pp. 1–25.

[10] E. S. Bigun, J. Bigun, B. Duc, and S. Fischer, “Expert conciliation for multi modal person authentication systems by bayesian statistics,” in *Proc. Audio- Video-Based Biometric Person Authent.*, 1997, pp. 311–318.

[11] J. Kittler, M. Hatef, R. P. W. Duin, and J. Matas, “On combining classifiers,” *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 20, pp. 226–239, Mar. 1998.

[12] A. K. Jain and A. Ross, “Learning user-specific parameters in a multibiometric system,” in *Proc. IEEE Int. Conf. Image Process.*, Rochester, NY, 2002, pp. 57–60.

[13] K.-A. Toh, W.-Y. Yau, and X. Jiang, “A reduced multivariate polynomial model for multimodal biometrics and classifiers fusion,” IEEE Trans. Circuits Syst. Video Technol. (Special Issue on Image- and Video-Based Biometrics), vol. 14, no. 2, pp. 224–233, Feb. 2004, to be published.

[14] S. Prabhakar and A. K. Jain, “Decision-level fusion in fingerprint verification,” *Pattern Recogn.*, vol. 35, no. 4, pp. 861–874, 2002.

[15] K.-A. Toh, W. Xiong, W.-Y. Yau, and X. Jiang, “Combining fingerprint and hand-geometry verification decisions,” in *Proc. Fourth Int. Conf. Audio- Video-Based Biometric Person Authentication*. Guildford, U.K.: Springer, June 2003, pp. 688–696 (AVBPA, LNCS 2688).

[16] M. Skurichina, S. Raudys, and R. P. W. Duin, “K-nearest neighbors directed noise injection in multilayer perceptron training,” *IEEE Trans. Neural Networks*, vol. 11, pp. 504–511, Marf. 2000.

[17] G. Cybenko, “Approximations by superpositions of a sigmoidal function,” *Math. Cont. Signal Syst.*, vol. 2, pp. 303–314, 1989.

[18] K. Hornik, M. Stinchcombe, and H. White, “Multi-layer feedforward networks are universal approximators,” *Neural Networks*, vol. 2, no. 5, pp. 359–366, 1989.

[19] K.-A. Toh, “Deterministic global optimization for FNN training,” *IEEE Trans. Systems, Man Cybern. B*, vol. 33, pp. 977–983, Dec. 2003.

[20] F. Schwenker, H. A. Kestler, and G. Palm, “Radial-basis-function netowrks: Learning and applications,” in *Proc. 4th Int. Conf. Knowledge-Based Intelligent Eng. Syst. Allied Technol.*, Brighton, U.K., 2000, pp. 33–43.

[21] A. Jain, L. Hong, and R. Bolle, "On-line fingerprint verification," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 19, pp. 302–313, Apr. 1997.

[22] A. K. Jain, L. Hong, S. Pankanti, and R. Bolle, "An identity-authentication system using fingerprints," *Proc. IEEE*, pp. 1365–1388, Sept. 1997.

[23] N. K. Ratha, K. Karu, S. Chen, and A. K. Jain, "A real-time matching system for large fingerprint databases," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 18, pp. 799–812, Aug. 1996.

[24] U. Halici, L. C. Jain, and A. Erol *et al.*, "Introduction to fingerprint recognition," in *Intelligent Biometric Techniques in Fingerprint and Face Recognition*, L. C. Jain *et al.*, Eds. Boca Raton, FL: CRC, 1999, pp. 3–34.

[25] X. Jiang and W. Y. Yau, "Fingerprint minutiae matching based on the local and global structures," in *Proc. 15th Int. Conf. Pattern Recogn.*, vol. 2, 2000, pp. 1042–1045.

[26] X. Jiang, W. Y. Yau, and W. Ser, "Detecting the fingerprint minutiae by adaptive tracing the gray-level ridge," *Pattern Recogn.*, vol. 34, no. 5, pp. 999–1013, 2001.

[27] J. M. Naik, "Speaker verification: A tutorial," *IEEE Commun. Mag.*, pp. 42–48, Jan. 1990.

[28] C. Li and R. Venkateswarlu, "High accuracy connected digits recognition system with less computation," in *Proc. 6th World Multiconf. Systemics, Cybern., Informatics*, Orlando, FL, July 2002.

[29] L. Rabiner and B.-H Juang, *Fundamentals of Speech Recognition*. Englewood Cliffs, NJ: Prentice-Hall, 1993.

[30] R. Sanchez-Reillo, C. Sanchez-Avila, and A. Gonzalez-Marcos, "Biometric identification through hand geometry measurements," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 22, pp. 1168–1171, Oct. 2000.

[31] A. Jain and N. Duta, "Deformable matching of hand shapes for verification," in *Proc. Int. Conf. Image Processing*, Kobe, Japan, Oct. 1999.

[32] A. Jain, A. Ross, and S. Pankanti, "A prototype hand geometry-based verification system," in *Proc. 2nd Int. Conf. Audio- Video-Based Biometric Person Authentication*, Washington, DC, March 1999, pp. 166–171.

[33] N. Ansari and E. J. Delp, "On detecting dominant points," *Pattern Recogn.*, vol. 24, no. 5, pp. 441–451, 1991.

Kar-Ann Toh (M'92–SM'03) received the Ph.D. degree from Nanyang Technological University (NTU), Singapore, in 1999.

Prior to his postdoctoral appointments at research centers in NTU from 1998 to 2002, he worked for two years in the aerospace industry. Currently, he is with the Institute for Infocomm Research, Singapore. His research interests include biometrics and decision fusion, pattern classification, optimization, and neural networks. He has made several PCT filings related to biometric applications and has actively published his works in above areas of interest.

Dr. Toh has served as a reviewer for several international journals including the IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS I, and IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS B.

Xudong Jiang (M'87) received the B.Eng. and M. Eng. degrees from the University of Electronic Science and Technology of China, Chengdu, China, in 1983 and 1986, respectively, and the Ph.D. degree from the University of German Federal Armed Forces, Hamburg, Germany, in 1997, all in electrical and electronic engineering.

From 1986 to 1993, he was a Teaching Assistant and then a Lecturer at the University of Electronic Science and Technology of China. From 1993 to 1997, he was with the University of German Federal Armed Forces as a scientific assistant. From 1998 to 2002, he was with the Centre for Signal Processing, Nanyang Technological University, Singapore, first as a Research Fellow and then as a Senior Research Fellow, where he developed a fingerprint verification algorithm that achieved the fastest and the second most accurate fingerprint verification in the International Fingerprint Verification Competition (FVC2000). From 2002 to 2004, he was a Lead Scientist and Head of the Biometrics Laboratory at the Institute for Infocomm Research, Singapore. Currently, he is an Assistant Professor with the School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore. His research interest includes pattern recognition, neural networks, image processing, computer vision, biometrics, adaptive signal processing, and spectral analysis.

Dr. Jiang received two Science and Technology Awards from the Ministry for Electronic Industry of China and was a recipient of the German Konrad-Adenauer Foundation young scientist scholarship.

Wei-Yun Yau (S'90–M'98) received the B. Eng. degree with Honors in electrical engineering from the National University of Singapore in 1992 and the M. Eng. degree in biomedical image processing in 1995 and the Ph.D. degree computer vision in 1999, both from the Nanyang Technological University, Singapore.

From 1997 to 2002, he was a Research Engineer and then Program Manager at the Centre for Signal Processing, Singapore, leading the research and development effort in the area of biometric processing, where his team won the top three positions in both speed and accuracy in the international Fingerprint Verification Competition 2000 (FVC2000). He initiated and served as the Program Director of the Biometrics Enabled Mobile Commerce (BEAM) Consortium from 2001 to 2002. In addition, he actively participates in both national and international biometric standard activities. Currently, he is with the Institute for Infocomm Research, Singapore, as a Department Manager, leading the research and development effort in the area of human computer interaction. His research interest includes biomedical engineering, biometrics, computer vision, and intelligent systems.

Dr. Yau was a recipient of the Kuok Foundation Undergraduate Scholarship, and his undergraduate project won the top prize in the Electronic Engineering/Telecommunications category of the Technology Fair 1992. Currently, he is the Chair of the Biometrics Pro-tem Technical Committee, Singapore. He also received the TEC Innovator Award in 2002 and the Tan Kah Kee Young Inventors' Award in 2003 (Merit).