Optimal Linear Combination of Facial Regions for Improving Identification Performance

Kin-Chung Wong, *Member, IEEE,* Wei-Yang Lin, *Member, IEEE,*
Yu Hen Hu, *Fellow, IEEE,* Nigel Boston, *Member, IEEE,*
and Xueqin Zhang, *Member, IEEE*

***Abstract*— This paper presents a novel 3D multi-regions face recognition algorithm that consists of new geometric summation invariant features, and an optimal linear feature fusion method. A summation invariant, which captures local characteristics of a facial surface, is extracted from multiple sub-regions of a 3D range image as the discriminative features. Similarity scores between two range images are calculated from the selected sub-regions. A novel fusion method based on linear discriminant analysis (LDA) is developed to maximize the verification rate by weighted combination of these similarity scores. Experiments on the FRGC (Face Recognition Grand Challenge) V2.0 dataset show that this new algorithm improves the recognition performance significantly in the presence of facial expressions.**

***Index Terms*— Face recognition, 3D faces, Face Recognition Grand Challenge (FRGC), information fusion.**

W. Lin is with the Department of Computer Science and Information Engineering, National Chung Cheng University, Min-Hsiung, Chia-Yi, Taiwan (e-mail:wylin@cs.ccu.edu.tw).

K. Wong, Y. H. Hu and N. Boston are with the Department of Electrical and Computer Engineering , University of Wisconsin-Madison, Madison, WI 53706-1691 (e-mail:kinchungwong@gmail.com, hu@engr.wisc.edu, boston@engr.wisc.edu).

X. Zhang is with the the Department of Electron and Communication Engineering, East China University of Science and Technology(ECUST), Shanghai, China (e-mail:zxq@ecust.edu.cn).

I. INTRODUCTION

Human face recognition has received unprecedented interest in recent years [1], [2]. However, rigorous tests with real-world data such as FERET, FRVT have revealed many shortcomings of existing approaches [3], [4]. In particular, for large scale, real world situations, current systems still cannot deliver the performance needed for practical applications.

A majority of current face recognition approaches make use of 2D frontal facial texture

features which may be sensitive to lighting, pose, distance, age (temporal) variations, and can easily be altered through simple make-up efforts. On the other hand, if 3D facial surfaces are available, one may exploit features that are invariant to appearance variations. For example, the facial surface around cheek bones or the nose would remain unchanged under varying lighting conditions, are less likely to change due to aging, and are seldom covered with hairs. Hence, in this research, we focus on exploiting *invariant* features extracted from 3D facial surfaces.

With 3D facial surfaces, the illumination variation problem can be mostly alleviated; and the pose variations can be normalized successfully [5] if the location of facial features, such as nose tip, corners of eyes, ..., etc, are given. Unfortunately, one issue remains due to the non-rigid nature of human face. So far, performing face recognition in the presence of facial expressions is still one of the most challenging problems in pattern recognition. One simple method for handling this issue is to use the facial regions which will not change significantly with varying expressions. However, if we discard too much facial data, the recognition performance will decrease accordingly. So, it is important to choose a sub-region or several sub-regions which will maintain recognition performance and also be robust in the presence of expression variations. This observation prompts researchers to explore face recognition methods that make use of multiple sub-regions [6], [7]. However, these earlier results focus only on 2D texture facial images, and do not discuss the tolerance of their approaches to the changes in facial expressions.

In this work, we propose a novel multi-regions face recognition algorithm for 3D face recognition. Specifically, we identify multiple sub-regions over a given range facial image, and extract summation invariant features from each sub-region. For each sub-region, and the corresponding summation invariant feature, a matching score is calculated. Then, a linear fusion method is developed to combine the matching scores of individual regions to arrive at a final matching score. Earlier, in [8], [9] we presented the derivation of the summation invariants and reported preliminary results of applications of this new type of invariant feature to single-region face recognition problems. In this paper, the key innovation is the development of the multi-regions 3D face recognition method that incorporates novel feature fusion and feature selection methods to significantly enhance the performance.

In pattern classification literatures, it is well known that combining multiple classifiers often leads to superior performance than that achieved by any individual classifier [10]–[12].

There are two major categories of classifier fusion architecture: *stacked generalization* [13], also known as the *committee machine* [14]–[17], or the *ensemble method* [18]–[23]; versus mixture of experts [24]–[26].

Previously, in [27], optimal data fusion is presented under the constraint of a fixed k out of n weighted threshold fusion architecture. In [28], a hierarchical model is used and Bayesian Gibbs sampling method is used to design the fusion rule. Data fusion has also been studied in the context of combining multiple classifiers. In [29], three types of classifier combination methods, namely, *averaged Bayes classifiers*, *voting principals*, and *Dempster-Shafer fuzzy combinations* have been reviewed. Some experiments have been conducted but no conclusive comparison results are available. In [30], the accuracy of individual classifiers are estimated, and classifiers are selected dynamically based on which classifier will yield best performance in specific local region. Ji and Ma [31] proposed to use a structure consisting of randomly generated linear local classifiers with a voting fusion mechanism to perform pattern classification tasks. Petrakos *et al.* [32] discussed the effect of correlations between classifiers and their impacts on fusion performance. The behavior knowledge space (BKS) method [33] is a non-parametric decision fusion method that uses a look-up-table (LUT) to implement the decision fusion classifier. For the given set of training samples, the BKS method is guaranteed to offer the best decision fusion results. However, with a finite number of training samples, the BKS method does not necessarily yield the best generalization results.

In our work, the 3D face recognition experiments are conducted by using *Face Recognition Grand Challenge* (FRGC) version 2.0 data set under the BEE (Biometric Experimental Environment) protocol. The experimental results are stored in a 4007 by 4007 matrix, called the similarity matrix. The $(i, j)^{th}$ entry of this similarity matrix represents a measure of how close the feature vector extracted from the i^{th} gallery 3D face range image is to the j^{th} probe 3D face range image. A similarity score is often a real number between 0 and 1 after normalization. With ten sub-regions and hence ten feature vectors extracted from one face image, the objective of fusion is to combine the ten corresponding similarity scores into a single similarity score. Toward this goal, we propose an linearly optimal fusion method based on the classical linear discriminant analysis (LDA) method to estimate the set of optimal weights. We compare our results with that of other decision fusion methods using linear support vector machines and found the performance is comparable while the LDA based approach requires much less computation.

The rest of our paper is organized as follows: In section II, we give a brief review of previous works on 3D face recognition and illustrate the face recognition technique being employed. Section III presents the experimental setup and results of three fusion schemes, namely sum rule, Linear discriminant Analysis and Linear Support Vector Machine. Finally, we make some conclusive remarks in section IV.

II. 3D FACE RECOGNITION

The majority of face recognition research focuses on using intensity images of the face. However, 3D images of the face have several advantages over the intensity-based features. In particular, 3D data provide a better representation for describing properties of the face in areas such as the cheeks, forehead, and chin, and are illumination invariant. In the following sections, we will briefly review the previous work on 3D face recognition and then illustrate our 3D face recognition system.

A. Previous Work in 3D Face Recognition

With the advance in 3D sensor technologies, face recognition based on 3D information has become an active research area in the recent years. 3D shape information has some inherent advantages over 2D texture information, such as invariance to illumination changes. A comprehensive survey on 3D face recognition can be found in [34]. The early work of applying invariant functions on 3D face recognition was done over a decade ago. At that time, people began with the geometrical properties introduced in differential geometry, such as principal curvatures, Gaussian curvature, ..., etc. Cartoux *et al.* [35] proposed a face recognition algorithm based on Gaussian curvature of the face surface. Their approach yields a 100% verification rate on a small dataset (5 subjects and 18 range images). Lee and Milios [36] presented an algorithm for establishing a correspondence between features of two faces. Facial features are obtained by a segmentation of the range image based on the sign of the mean and Gaussian curvatures at each point. Gordon [37] identifies the nose region, ridge and valley lines by using mean and Gaussian curvatures. Tanaka *et al.* [38] also perform curvature-based segmentation and represent the face surface using an Extended Gaussian Image (EGI). Basically, these approaches use the invariant functions, *e.g.* Gaussian curvature is invariant under Euclidean transformations, to extract information from the face surface and then perform classification based on extracted information. Although different methods have been used to address recognition based on invariant features, most of these invariant features belong to the category called *differential invariants* which rely on derivative operations. The

numerical computation of differential invariants is not reliable because of the quantization error in discretized data. This fundamental issue limits their potential to achieve high recognition performance in a large dataset.

More recently, Medioni *et al.* [39] perform 3D face recognition using iterative closest point (ICP) matching. Hesher *et al.* [40] perform principal component analysis (PCA) of range images. Similarly, Chang *et al.* [5] use a PCA-based method separately on 2D texture and 3D range images. The proposed multi-modal 2D+3D face recognition scheme can achieve significant improvement over those using single modality. Further investigation into the 3D-PCA approach has been reported by Heseltine *et al.* [41]. A 3D face recognition algorithm based on fitting a deformable model to an input range image is proposed by Kakadiaris *et al.* [42]. They perform experiments on the FRGC v1.0 dataset and report a 97% verification rate measured at a 0.1% false accept rate (FAR). In our previous work [9], we introduce a novel family of geometrically invariant features, called *summation invariants*, for 3D face recognition. Our algorithm achieves the verification rate of 97.2% at 0.1% FAR on the FRGC v1.0 dataset.

Expression is a relatively unaddressed issue in face recognition literature. Chang *et al.* [43] tackle the issue of expression changes by utilizing three different regions around a nose, as being relatively rigid areas across different facial expressions. Passalis *et al.* [44] propose a fully automatic 3D face recognition algorithm based on the elastically adapted deformable model framework. Their algorithm can fit an Annotated Face Model (AFM) to an input 3D facial surface in the presence of facial expressions. A 3D face recognition approach based on geometric invariants was introduced by Bronstein *et al.* [45]. The key idea of their algorithm is to approximate facial expressions as isometric transformations, *i.e.* length preserving transformations. Their experiments are performed on a dataset containing 220 faces of 30 subjects and 1.9% equal error rate is reported. Their system can distinguish between two identical twins based on 3D images of their faces.

B. Summation Invariant

In 3D face recognition, there are two requirements for extracting features from facial surfaces. First, the extracted feature should be able to represent the local characteristics of a surface. Second, the computation of such a feature should be reliable in the presence of perturbation, such as sensor noise, quantization error, ..., etc. In differential geometry [46], many geometric features, such as principal curvatures, Gaussian curvature, ..., etc, can capture

local properties of a surface. Unfortunately, the curvatures rely on differentiation which will amplify the effect of noise. Hence, their numerical computation is neither accurate nor reliable.

Lin *et al.* [8], [9] introduce a novel family of geometrically invariant features called *summation invariants*. The summation invariants are obtained by the method of moving frames [47], [48], originally introduced by Élie Cartan, which is a powerful tool for constructing invariants under group actions. In our system, features are computed by using $\eta_{i,j}$, the Euclidean summation invariants of curves . Given a curve segment sampled at points (x_n, y_n), its Euclidean summation invariants are given by

$$\eta_{i,j} = \sum_{n=1}^{N} \bar{x}_n^i \bar{y}_n^j \tag{1}$$

where $\bar{x}, \bar{y}$ denote the x, y coordinates transformed by a moving frame and N denote the length of a curve. The first order and second order $\eta_{i,j}$ are explicitly derived by Lin *et al.* in [9]. Among them, $\eta_{1,1}$ (see Eq. (2)) yields the highest face recognition performance on FRGC v1.0 dataset.

$$\begin{aligned}
\eta_{1,1} \;=\; & P_{1,1}((x_1 - x_N)^2 - (y_1 - y_N)^2) \\
& + P_{1,0}(y_1^3 + 2x_1 x_N y_N - 2y_N x_1^2 \\
& + x_1^2 y_1 - 2y_1^2 y_N + y_1 y_N^2 - x_N^2 y_1) \\
& - P_{0,1}(x_1^3 + 2y_1 y_N x_N - 2x_N y_1^2 \\
& + y_1^2 x_1 - 2x_1^2 x_N + x_1 x_N^2 - y_N^2 x_1) \\
& + (P_{0,2} - P_{2,0})(x_1 - x_N)(y_1 - y_N) \\
& + N(x_N y_1 - x_1 y_N) \\
& (x_1(x_N - x_1) + y_1(y_N - y_1))
\end{aligned} \tag{2}$$

where

$$P_{i,j} = \sum_{n=1}^{N} x_n^i y_n^j \tag{3}$$

A high value of $\eta_{1,1}$ represents a concave shape while low value represents a convex shape, see Fig. 1. Note that the values of summation invariants $\eta_{i,j}$ do not depend on the choice of world coordinate system and, therefore, they are useful features for comparing facial surfaces with different poses.

C. 3D Face Recognition Using Single Region

The single-region algorithm, which is the building block of the proposed multi-regions face recognition system, consists of the following stages:

1) Use only the shape information in a 3D image: The 3D images provided by current 3D sensors usually contain two

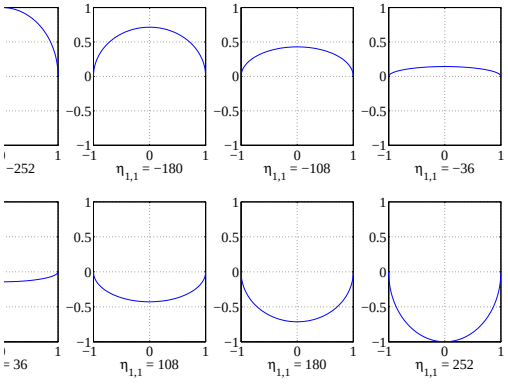


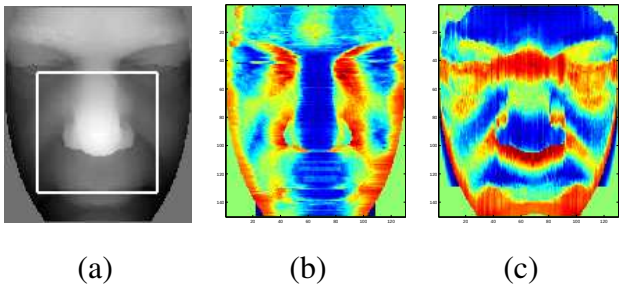
Fig. 1. Geometrical meaning of summation invariant $\eta_{1,1}$. The larger $\eta_{1,1}$, the smaller the area under the curve.

channels, namely texture and shape. The texture channel means the intensity images. In the shape channel, each pixel represents the depth value of the scene. In this work, we focus on the pure 3D face recognition technology, hence the 2D texture information is not utilized in our experiments.

2) Arc-length resampling : Before computing the summation invariant, range data are resampled uniformly with respect to arc-length. Specifically, for the first row on the range data, we first compute its arc-length and resample it uniformly with respect to the arc-length. Then, we perform the same arc-length resampling on the following rows.
3) Semi-local summation invariants: While the FRGC baseline algorithm uses 3D range data directly, *i.e.* the depth values, we compute the summation invariant $\eta_{1,1}$ semi-locally from a normalized range image and use the results as an invariant feature. In particular, the summation invariant is computed from a local window surrounding each pixel. We compute the semi-local summation invariants from both a horizontal window and a vertical window. The summation invariant $\eta_{0,1}$ [9] computed by using a horizontal window and a vertical window are shown in Figure 2 (b) and (c), respectively. The window size L is experimentally determined, $L = 21$ in our experiments.
4) Specify the region of interest: Instead of using the entire 3D range image as the FRGC baseline algorithm does, we crop invariant features from the 81×81 subregion centered at the nose tip, shown in Figure 2.
5) Alignment refinement: The location of the selected region is refined by minimizing the SSD (sum of squared differences) with the averaged invariant features of the same region. We compute the averaged invariant features of that region during the training stage.
6) Dimension reduction : In order to reduce the dimensionality of the feature vectors, we use principal component analysis

(PCA) to compute their subspace representation. The PCA basis, *i.e.* eigenfaces, is computed by using the training set containing 943 range images.

7) Similarity metric : During the classification, we use Mahalanobis cosine as a similarity metric. Note that in our experiments, the similarity metric is the same as the one in FRGC baseline algorithm. Please refer to [5] for more details.



(a) (b) (c)

Fig. 2. (a) Normalized range image and a selected sub-region; (b) The computed result of $\eta_{0,1}$ using a horizontal window; and (c) The computed result of $\eta_{0,1}$ using a vertical window. The red color represents higher value and blue color represents lower value.

The single-region algorithm works very well on the FRGC v1.0 dataset [9] which contains only neutral expressions. Unfortunately, we observe a significant performance drop when it was applied to the FRGC v2.0 dataset where expression changes exist. The main reason is that it is difficult to find a single region which is rigid across expressions and also large enough to provide good discriminating power. An intuitive method to overcome this limitation is to utilize multiple regions. In the multi-regions algorithm, each region should be relatively small compared to the one used in single-region algorithm so that it is relatively rigid across expressions. Also, the multi-regions algorithm is expected to perform better than combining multiple summation invariants [9]. If we compute different summation invariants from a single region, the extracted features will be highly correlated. Hence, the benefits of integrating multiple summation invariants will be limited. On the other hand, it is reasonable to assume that the feature vectors extracted from different sub-regions on a facial surface are statistically independent. This suggests that different facial sub-regions provide complementary information about the face to be classified. In this work, we will address the fusion of different facial regions rather than the fusion of different summation invariants.

III. Multi-Regions Face Recognition Algorithm

Our goal is to achieve the best possible recognition performance by combing information at hand. Information obtained from multiple resources could be fused by using a number of approaches, at different levels. In general, there are four possible levels of fusion [49]: (1)

sensor level, (2) feature level, (3) opinion/score level, (4) decision level. Figure 3 shows an example of fusion at opinion/score level.

There might be a strong correlation between the information resources (*e.g.*, different feature vectors extracted from the same data) and such a scenario is referred to as tightly coupled integration [50]. On the other hand, the loosely coupled system means little correlation between information resources (*e.g.*, feature vectors extracted from multiple modality) and fusion occurs at the output of independent classifiers. It has been shown that the independence of classifiers plays an important role in performance improvement [51]. Hence, the emphasis of our work will be placed on the design and analysis of a loosely coupled system. The scenario of fusion is as following: Each single-region sub-system matches an image pair independently and generates the corresponding matching score. The matching scores of individual single-region sub-systems are combined by different fusion schemes. Each sub-system is based on one of the 10 facial regions shown in Fig. 4. These regions are located on the normalized range images provided by FRGC baseline algorithm. Note that the location of nose tip and the pixel distance between eye corners are normalized during the preprocessing stage. Interesting readers may refer to [5] for the details on the preprocessing of 3D range images. We will discuss three approaches for information fusion. We demonstrate that information fusion results in significant performance improvements in all of these three cases.

A. Description of Experiments

We perform face recognition experiments on the FRGC v2.0 dataset [52]. In the FRGC dataset, 3D images with a resolution 640×480 consist of both shape and textures channels. The two disjoint data partitions, training and validation partitions, contains 943 and 4007 3D images respectively. Among the challenge problems defined in FRGC v2.0, we focus on experiment 3s which utilizes only shape channel in a 3D image. Performance will be reported on a Receiver Operator Characteristic (ROC) that shows the trade-off between verification and false accept rate. In the FRGC protocol, three masks are defined over the similarity matrix where each entry contains the matching score of an image pair. Each mask collects its own set of entries in the similarity matrix, thus generating three ROC curves which will be referred to as ROC I, II and III. The images within an image pair are called gallery and probe. In ROC I, gallery and probe are from the same semester. In ROC II, gallery and probe are from the same year. In ROC III, gallery and probe are from different semesters. In average, ROC III has the longest time lapse between

gallery and probe and therefore is the most challenge experiment.

B. Sum Rule

The simplest way of fusion is to add the scores from multiple regions. Equal weights are assigned on each region since we don't know the relative importance of each region. Figure 5 shows the results of sum rule on the FRGC v2.0 dataset. Note that the verification performance of single-region algorithm is better than that of FRGC baseline. This, once again, confirms the fact that summation invariants provide a better discriminating capability than depth value does. But, the performance improvement is limited in the presence of facial expressions. This limitation can be overcome by utilizing multiple regions. By combining relatively small regions which are relatively unaffected by expressions, we observe a significant improvement in verification rate.

C. Linear Discriminant Analysis

In the FRGC protocol, there are two non-overlapping data partitions: training and validation. During algorithm development, experiments are conducted on the data in the validation partition. This allows researchers to tune the parameters of their approaches. However, finding a subspace representation and classifier training are required to be conducted on the training partition. For example, PCA-based algorithms should construct eigenfaces from the training set. Here, we use Linear Discriminant Analysis (LDA) [53] to obtain an optimal weighting vector from the training partition. In FRGC v2.0 dataset, there are a total of 943 3D images in the training partition.

Let $\mathbf{S} = (s_1, s_2, \ldots, s_n)^T$ denote a score vector, where n is the number of facial regions, $n \leq 10$ in our experiments. Each s_i is the matching score of i-th region, generated from a pair of images. The LDA finds the optimal projection which transforms the original score vector $\mathbf{S} \in \mathbb{R}^n$ to the final matching score $\widetilde{S} \in \mathbb{R}$.

$$\widetilde{S} = \mathbf{W}_{LDA}^T \mathbf{S} \tag{4}$$

where $\mathbf{W}_{LDA}$ is a $n \times 1$ weighting vector obtained by solving

$$\mathbf{W}_{LDA} = \arg\max_{\mathbf{W}} \frac{\mathbf{W}^T \mathbf{S}_B \mathbf{W}}{\mathbf{W}^T \mathbf{S}_W \mathbf{W}} \tag{5}$$

where $\mathbf{S}_B$ is the between-class scatter matrix and $\mathbf{S}_W$ is the within-class scatter matrix.

$$\mathbf{S}_B = (\mathbf{m}_1 - \mathbf{m}_2)(\mathbf{m}_1 - \mathbf{m}_2)^T \tag{6}$$

$$\mathbf{S}_W = \sum_{i=1}^{2} \sum_{\mathbf{S}_k \in \mathbb{S}_i} (\mathbf{S}_k - \mathbf{m}_i)(\mathbf{S}_k - \mathbf{m}_i)^T \tag{7}$$

In Eq. (6) and (7), $\mathbf{m}_i$ is the mean score vector of samples belonging to class i, $\mathbb{S}_i$ denotes the set of samples belonging to class i.

The solution for the criterion function Eq. (5) is given by

$$\mathbf{W}_{LDA} = \mathbf{S}_W^{-1}(\mathbf{m}_1 - \mathbf{m}_2) \tag{8}$$

In LDA-based face recognition such as [54], the number of classes is the number of subjects. Here, we apply LDA on a two classes problem, namely match class and non-match class. Match class contains the score vectors from image pairs of the same person while non-match class is generated by image pairs of the different persons. The LDA provides the optimal projection direction that maximizes the separation between two classes. The result of weighted combination by using LDA is shown in Figure 6(a). We observe a significant improvement over sum rule. The results suggest that the distribution of training data is an important knowledge available for us to construct a better classifier. We should also mention the fact that the training partition contains only neutral expression while validation partition contains different expressions. Hence, the distribution of training samples is expected to be different from the distribution of validation data in FRGC v2.0 protocol. This also points out a future direction that one could expect performance improvement by including expression changes in the training samples. The weights obtained by LDA are shown in Table I. From the table, we observe that LDA puts the highest weight on nose region and second highest one on forehead region. It indicates that nose and forehead are the two most important features in 3D face recognition.

D. Linear Support Vector Machine

Linear support vector machine (LSVM) is the simplest kind of SVM. It is a linear classifier with the maximized margin. The *margin* means the width that a decision boundary can be increased by without contacting a data point. By maximizing margin, it gives us the least chance of making a misclassification. Consider a set of d-dimensional training data belonging to two classes, the decision hyperplane and margins are given by

$$\begin{aligned} \text{Plus plane} &= \{\mathbf{x} : \mathbf{W}_{LSVM}^T\mathbf{x} + b = 1\} \\ \text{Decision plane} &= \{\mathbf{x} : \mathbf{W}_{LSVM}^T\mathbf{x} + b = 0\} \\ \text{Minus plane} &= \{\mathbf{x} : \mathbf{W}_{LSVM}^T\mathbf{x} + b = -1\} \end{aligned}$$

where $\mathbf{x}, \mathbf{W}_{LSVM} \in \mathbb{R}^d$ and $b \in \mathbb{R}$. The margin width M is defined as the shortest Euclidean distance between plus and minus plane.

$$M = \frac{2}{\|\mathbf{W}_{LSVM}\|} \tag{9}$$

So, we just need to search for the $\mathbf{W}_{LSVM}$ and b with the widest margin. This can be formulated as an optimization problem as follows:

given the training data with labels $\{\mathbf{x}_k, y_k\}, k = \{1, \ldots, N\}, y_k \in \{-1, 1\}, \mathbf{x}_k \in \mathbb{R}^d$,

$$\min \|\mathbf{W}_{LSVM}\| \tag{10}$$

subject to

$$y_k(\mathbf{W}_{LSVM}^T \mathbf{x}_k + b) \geq 1 \quad \forall k \tag{11}$$

Note that this formulation considers the cases where data is linearly separable. It can be extended to linear nonseparable cases which is assumed in our research by taking classification errors into account.

$$\min \|\mathbf{W}_{LSVM}\| + C \sum_{k=1}^{N} \epsilon_k$$

subject to

$$\begin{aligned} \mathbf{W}_{LSVM}^T \mathbf{x}_k + b &\geq 1 - \epsilon_k \quad \text{if } y_k = 1 \\ \mathbf{W}_{LSVM}^T \mathbf{x}_k + b &\leq -1 + \epsilon_k \quad \text{if } y_k = -1 \\ \epsilon_k &\geq 0, \quad \forall k \end{aligned}$$

where C is the penalty factor assigned by user and ϵ_k is the classification error. For the data points which have been classified correctly, their ϵ_k are zero. The details about how to solve this optimization problem is beyond the scope of this paper. Readers may refer to [55] for more detailed discussion on solving the optimization problem.

We use the software package LIBSVM [56] to perform support vector classification. Similar to LDA, the score vector $\mathbf{S}$ is projected along the direction $\mathbf{W}_{LSVM}$ to yield the final score.

$$\widetilde{S} = \mathbf{W}_{LSVM}^T \mathbf{S} \tag{12}$$

Figure 6(b) shows the ROC curves of fusion by LSVM. We observe that the LSVM classification yields similar performance as LDA does. However, training a SVM on a large data collection is considered as a challenging problem. In the FRGC v2.0 dataset, the training partition contains 943^2 samples. It takes a long time for linear SVM training. On the other hand, the time complexity of LDA grows linearly with the number of training samples. Note that computing the inverse of $\mathbf{S}_w$ is not a dominant factor since the largest size of $\mathbf{S}_w$ is 10×10. Hence, the LDA is a more efficient method for combining multiple regions than the LSVM. The weights obtained by LSVM are shown in Table II. Similar to LDA results, we observe that LSVM puts the highest weight on nose region and the second highest one on forehead region. It confirms the importance of these two regions in the context of 3D face recognition.

E. Discussion

The experiments described above suggest that the LDA perform better than the sum rule. For ROC III, the Verification rate of LDA and sum rule are 90.04% and 87.96% respectively (at FAR = 0.1%). The LDA takes the distribution of training data into account so that we

can construct a better classifier. By fusion at score level, the FAR can be fixed and then compute verification rate in both cases. For FRGC experiment 3s, the proposed algorithm achieves similar performance as the best results in the recent FRGC report [57].

Fig. 7 shows the distributions of matching scores for the selected regions before score-level fusion. The score distributions of combining 10 regions by sum rule and LDA are shown in fig (8). It can be seen from this figure that the separability of two classes has been improved as compared to that of the classifiers based on single-region. The result indicates that the multi-regions system performs better than any of the single-region systems.

Intuitively, facial features are different in terms of their importance in face recognition. To explore the relative importance of each region, we began with 10 regions and then leave one region out at each experimental run. After removing one region, we compute the optimal weights on remaining regions by LDA/LSVM. Table I and II contains the weights obtained by LDA and LSVM respectively. We first observe that using overlapped regions does not improve performance, *i.e.* region 9 and 10 do not have significant contribution. These tables show a clear difference in the values assigned to each region. The results suggest that nose has the highest contribution followed by forehead. Note that the combination of nose and forehead can achieve a verification rate of 84% at false accept rate of 0.1%. This is striking new evidence since psychological experiments typically indicate that eyes are the most important followed by mouth and the nose [58].

How could we reasonably explain this finding? One possibility is that the human visual system relies on the projection of a face , *i.e.* 2D texture image, rather than the face itself, which is a 3D object, in identifying a face. This new evidence might also explain why and how 3D face recognition can help improve the performance of 2D face recognition system. That is, the outputs of 2D face recognition and 3D face recognition may not agree with each other when classifying a face because they are dominated by different facial regions. Hence, the error made by one of them might possibly be covered by the other.

F. Sensitivity to Facial Expressions

In general, facial expressions could be classified by types (smile, angry, etc.) or strength (neutral, weak, strong). For the purpose of recognition, we are concerned about the performance at different levels of expression strengths, not the types. In this experiment, facial expressions are classified into 3 categories: neutral, small and large. The *neutral* category includes the face postures without

obvious expressions, while *small* means natural expressions such as moderate smiles. The *large* category contains the faces with extreme expressions. In many verification/identification scenarios, it is reasonable to assume cooperation from subjects. If a subject is making a weird expression, we can ask him/her to be more cooperative.

The validation partition of FRGC v2.0 dataset, containing 4007 images, is manually classified into these three categories. The manual labeling of images is performed by Geometrix Inc. [59] and labeling results are available for research community. The percentage of neutral, small and large expressions are 60%, 20% and 20% respectively. Fig. (9) shows the ROC curves of following experimental settings (from top to bottom): First, neutral images are matched with neutral images (60% vs 60%). Second, neutral images are matched with the union of neutral and small categories (60% vs 80%). Third, neutral images are matched with all 4007 images (60% vs 100%). Fourth, all 4007 images are matched with all 4007 images (100% vs 100%).

From these experiments, we observe that performance degradation is related to the strength of expressions. With neutral and small expressions, our algorithm can achieve verification rate of 96.90% at 0.1% false accept rate (the second curve from top). This is the performance that can be expected with cooperative subjects, *i.e.* natural expressions are allowable but not exaggerated ones. The lower two curves show significant performance degradation due to extreme expressions.

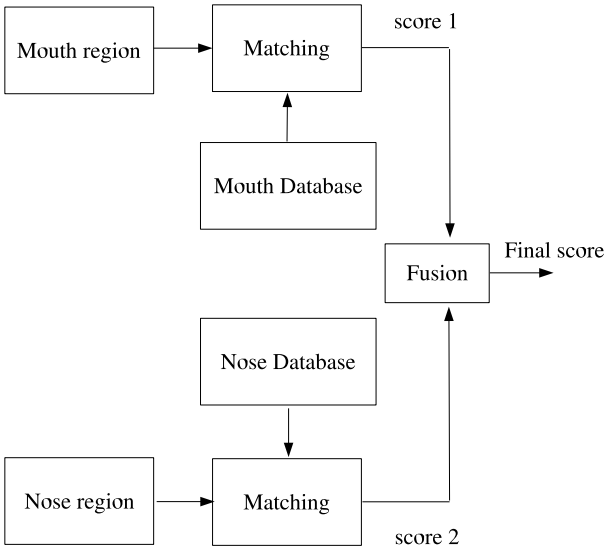


Fig. 3. Example of opinion fusion. The scores from mouth region and nose are combined to yield a final score.

IV. CONCLUSION

We have developed a 3D face recognition system which integrates multiple regions of a facial surface. The proposed system overcomes the limitations of the single-region algorithm. The performance improvement is due to the combination schemes which generate the final matching score with a higher quality than those based on single region. Experiments indicate that the LDA performs better than

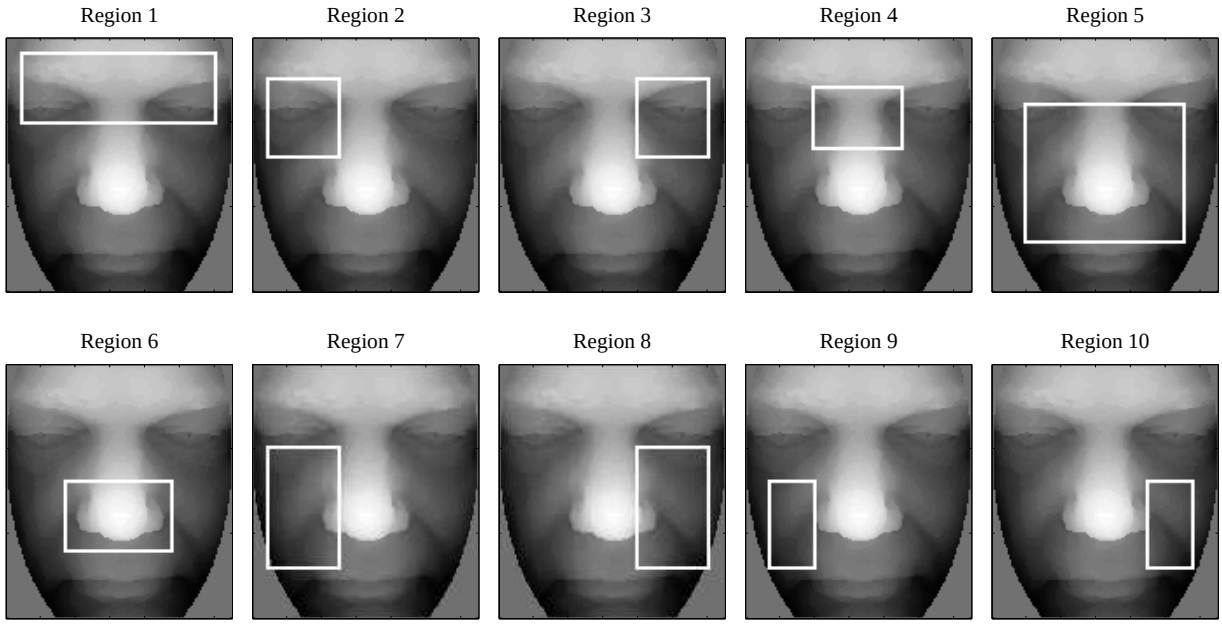
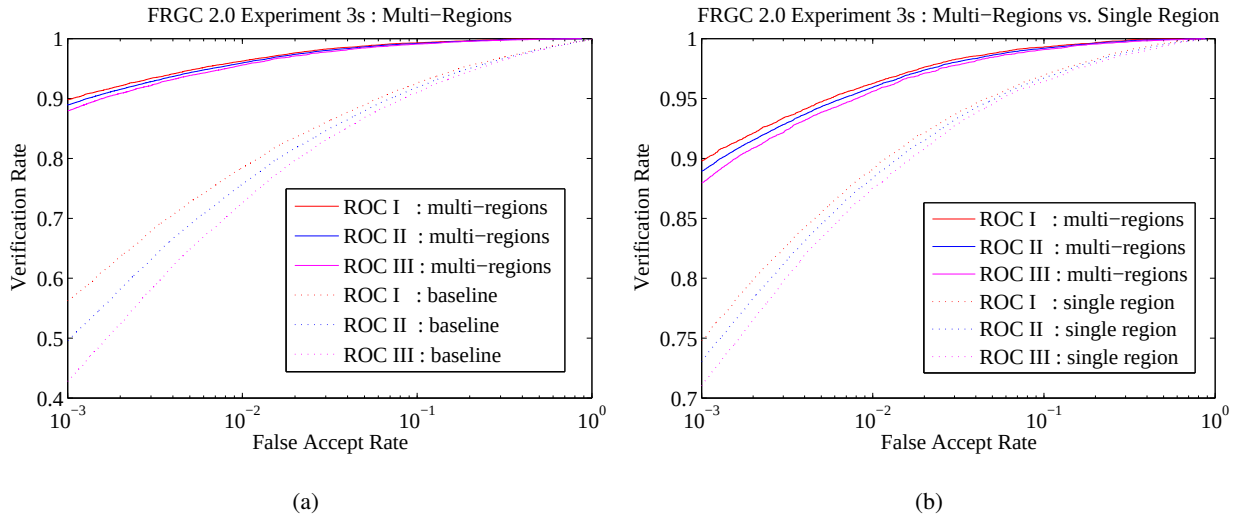


Fig. 4. We specify 10 regions on facial surface. Matching scores obtained from each region are combined to yield the final matching score.



(a) (b)

Fig. 5. ROC performance obtained by using sum rule for information fusion. (a) Comparison with BEE baseline algorithm and (b) Comparison with single-region algorithm

the sum rule. The reason is that the single-region sub-systems do not have the similar performance. If the individual classifiers do not have equal performance, we should assign weights to them. The LDA provides optimal weights for combining multiple classifiers so

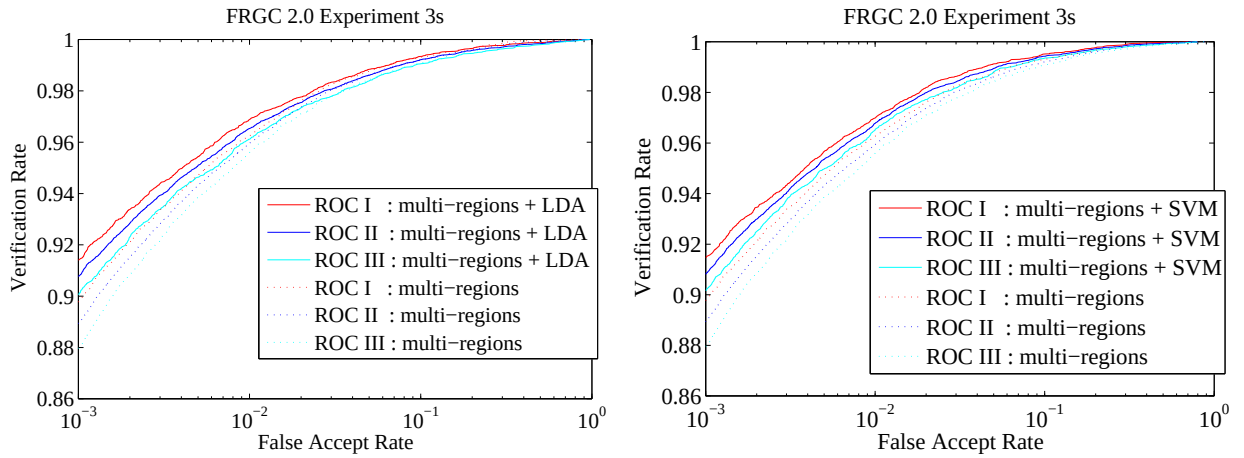


Fig. 6. ROC performance obtained by using (a) LDA and (b) LSVM for information fusion.

that it outperforms the sum rule. LSVM provides another optimal solution for information fusion. The fusion by LSVM achieves similar recognition performance as by LDA while the LSVM is computationally more intensive. Results on the FRGC v2.0 dataset demonstrate that our multiple-regions algorithm is robust in the presence of facial expressions.

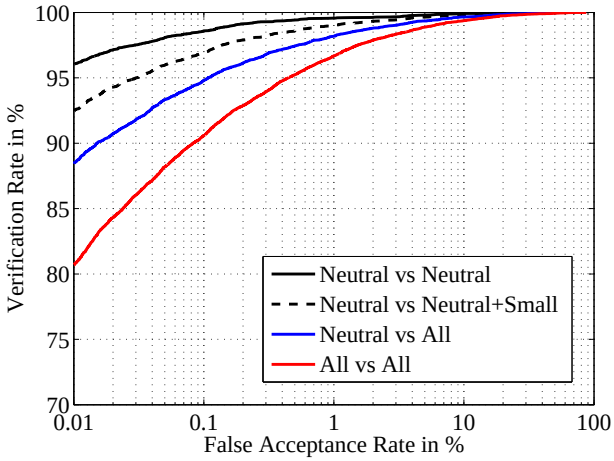


Fig. 9. ROC performance for different levels of expression strength.

REFERENCES

[1] K. W. Bowyer, K. Chang, and P. Flynn, "A survey of approaches to three-dimensional face recognition," in *Proc. of Intl. Conf. on Pattern Recognition*, vol. 1, 2004, pp. 358–61.

[2] W. Zhao, R. Chellappa, P. J. Phillips, and A. Rosenfeld, "Face recognition: A literature survey," *ACM Computing Surveys*, vol. 35, no. 4, pp. 399–458, 2003.

[3] P. J. Phillips, H. Moon, S. A. Rizvi, and P. J. Rauss, "The feret evaluation methodology for face-recognition algorithms," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 22, no. 10, pp. 1090–104, 2000.

[4] P. J. Phillips, P. Grother, R. Micheals, D. M. Blackburn, E. Tabassi, and M. Bone, "Face recognition vendor test 2002," in *Intl. Workshop on Analysis and Modeling of Faces and Gestures*, 2003, p. 44.

[5] K. I. Chang, K. W. Bowyer, and P. J. Flynn, "An evaluation of multimodal 2D+3D face biometrics," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 27, no. 4, pp. 619–24, 04 2005.

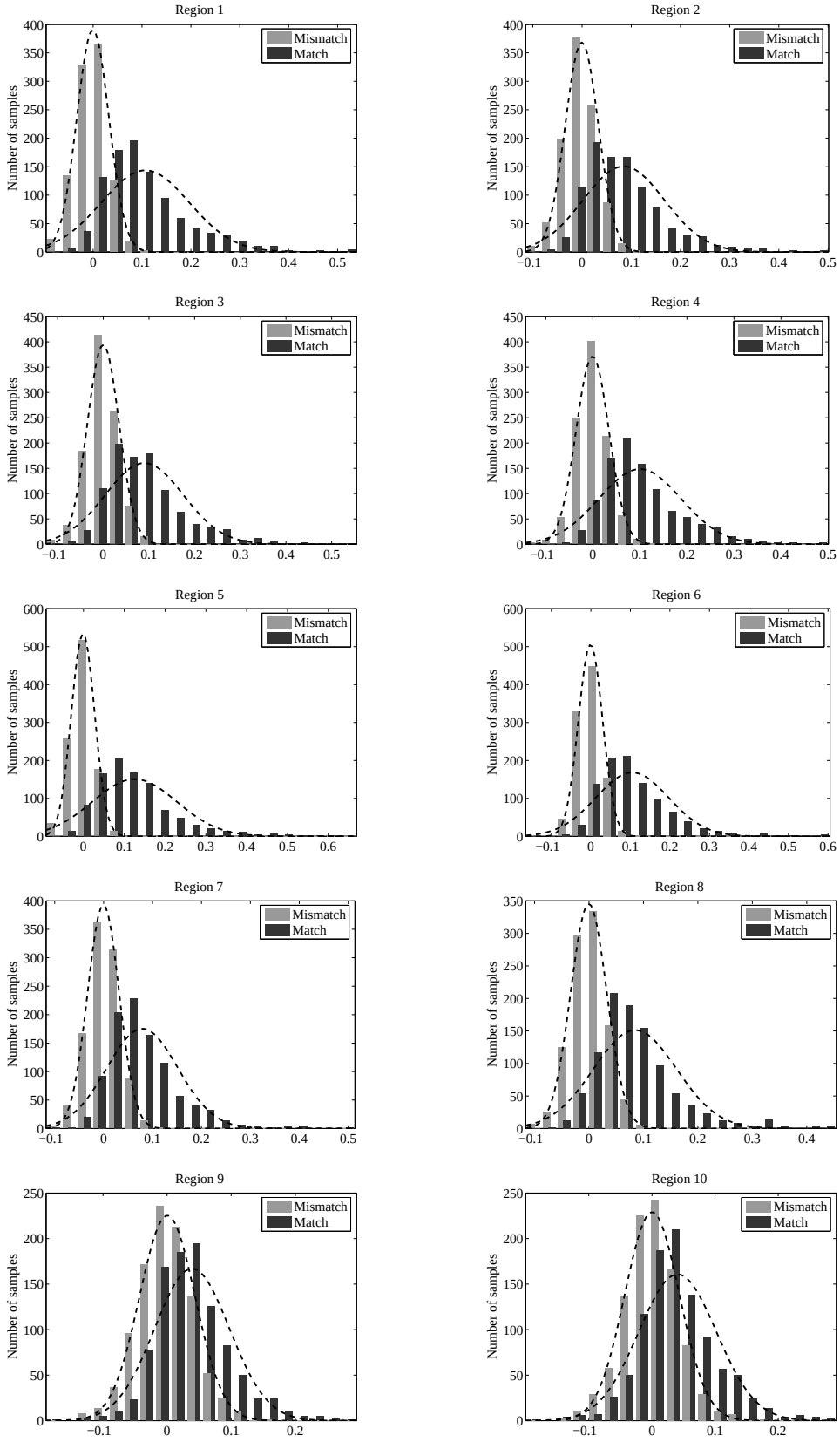


Fig. 7. Score distribution of the selected regions. Higher score means more similar. (dashed lines denote the Gaussian approximation of score distribution).

TABLE I

THE WEIGHTS ON SELECTED REGIONS OBTAINED BY LDA

Number of selected regions	Weights on the selected regions									VR @ FAR = 0.1%		
	#1	#2	#3	#4	#5	#6	#7	#8	#9	ROC-I	ROC-II	ROC-III
9 regions	.3586	.3222	.2665	.3661	.4849	.4008	.2396	.2749	.1831	.9185	.9120	.9045
8 regions	.3630	.3162	.2675	.3784	.4958	.4067	.2564	.2763	0	.9065	.8992	.8910
7 regions	.3708	.3125	.2709	.3759	.6225	0	.2881	.2819	0	.8867	.8751	.8614
6 regions	.3922	.3400	0	.3906	.6289	0	.2994	.3044	0	.9046	.8984	.8918
5 regions	.3980	.3736	0	.3923	.6679	0	0	.3192	0	.9086	.9034	.8983
4 regions	.4052	.3754	0	.3950	.7339	0	0	0	0	.8863	.8771	.8663
3 regions	.4692	0	0	.4374	.7671	0	0	0	0	.8450	.8302	.8133
2 regions	.5367	0	0	0	.8437	0	0	0	0	.8630	.8536	.8414
1 region	1.000	0	0	0	0	0	0	0	0	.7591	.7429	.7242

TABLE II

THE WEIGHTS ON SELECTED REGIONS OBTAINED BY LSVM

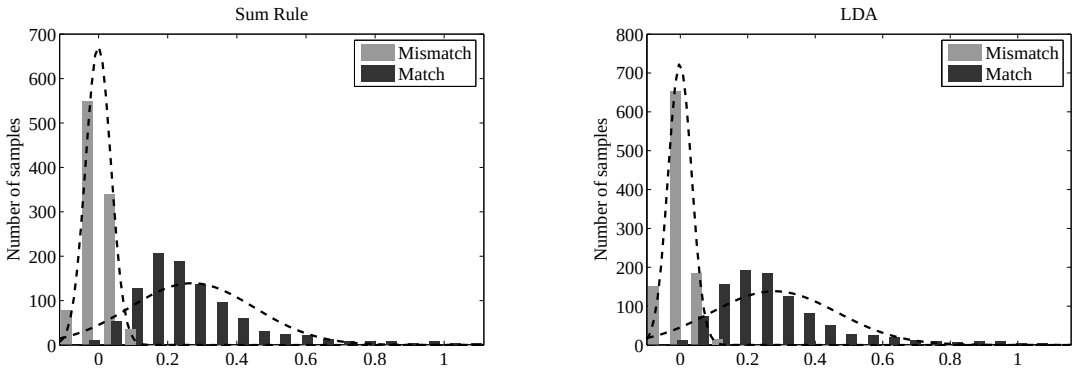
Number of selected regions	Weights on the selected regions										VR @ FAR = 0.1%		
	#1	#2	#3	#4	#5	#6	#7	#8	#9	#10	ROC-I	ROC-II	ROC-III
10 regions	2.534	1.293	1.423	1.991	2.792	1.889	1.444	1.501	.8265	1.032	.9145	.9080	.9014
9 regions	2.518	1.298	1.498	1.815	2.809	2.103	1.567	1.666	0	1.206	.9138	.9081	.9014
8 regions	2.288	1.259	1.654	1.896	3.100	2.179	1.879	1.789	0	0	.9152	.9096	.9028
7 regions	2.457	0	1.780	2.012	3.077	2.298	1.996	1.747	0	0	.9132	.9077	.9012
6 regions	2.609	0	1.904	1.973	3.700	2.347	2.244	0	0	0	.9136	.9084	.9030
5 regions	3.048	0	0	2.321	3.963	2.513	2.258	0	0	0	.9123	.9054	.8978
4 regions	3.683	0	0	0	4.376	2.519	2.313	0	0	0	.8861	.8797	.8715
3 regions	3.961	0	0	0	5.246	2.822	0	0	0	0	.8820	.8738	.8656
2 regions	4.317	0	0	0	6.917	0	0	0	0	0	.8594	.8525	.8439
1 region	0	0	0	0	1	0	0	0	0	0	.7591	.7429	.7242

[6] M. Jones and P. Viola, "Face Recognition Using Boosted Local Features," *Mitsubishi Electric Research Laboratories Technical Report Number: TR2003-25. Date: April*, 2003.

[7] A. Pentland, B. Moghaddam, and T. Starner, "View-based and modular eigenspaces for face recognition," *Proc. of IEEE Conf. on Computer Vision and Pattern Recognition (CVPR'94)*, pp. 84–91, 1994.

[8] W. Y. Lin, N. Boston, and Y. H. Hu, "Summation invariant and its application to shape recognition," in *Proc. IEEE Intl. Conf. on Acoustics, Speech, and Signal Processing*, vol. V, 2005, pp. 205–208.

[9] W. Y. Lin, K. C. Wong, N. Boston, and Y. H. Hu, "Fusion of summation invariants in 3D human face recognition,"



(a) (b)

Fig. 8. Score distribution of combining 10 regions. (a) Sum rule. (b) LDA.

in *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, vol. 2, 2006, pp. 1369 – 1376.

[10] J. Kittler, M. Hatef, R. P. W. Duin, and J. Matas, "On combining classifiers," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 20, no. 3, pp. 226–239, 1998.

[11] L. Hong and A. K. Jain, "Integrating faces and fingerprints for personal identification," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 20, no. 12, pp. 1295–1307, 1998.

[12] A. Ross and A. Jain, "Information fusion in biometrics," *Pattern Recogn. Lett.*, vol. 24, no. 13, pp. 2115–2125, 2003.

[13] D. Wolpert, "Stacked generalization," *Neural Networks*, vol. 5, no. 2, pp. 241–259, 1992.

[14] S. Fine, R. Gilad-Bachrach, and E. Shamir, "Query by committee, linear separation and random walks," *Theor. Comput. Sci.*, vol. 284, no. 1, pp. 25–51, 2002.

[15] Y. Freund, "Boosting a weak learning algorithm by majority," *Inf. Comput.*, vol. 121, no. 2, pp. 256–285, 1995.

[16] H. S. Seung, M. Opper, and H. Sompolinsky, "Query by committee," in *Computational Learning Theory*, 1992, pp. 287–294.

[17] J. Twomey and A. Smith, "Committee networks by resampling," *Intelligent Engineering Systems through Artificial Neural Networks*, vol. 5, pp. 153–158, 1995.

[18] H. Drucker, C. Cortes, L. D. Jackel, Y. LeCun, and V. Vapnik, "Boosting and other ensemble methods," *Neural Computat.*, vol. 6, pp. 1289–1301, 1994.

[19] L. K. Hansen and P. Salamon, "Neural network ensembles," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 12, no. 10, pp. 993–1001, 1990.

[20] A. Krogh and J. Vedelsby, "Neural network ensembles, cross validation, and active learning," *G. Tesauro, D. Touretzky, and T. Leen, editors, Advances in Neural Information Processing Systems*, vol. 7, pp. 231–238, 1995.

[21] M. Perrone and L. N. Cooper, "When networks disagree: ensemble method for neural networks," *Artificial Neural Networks for Speech and Vision, R. J. Mammone, Ed.*, pp. 126–142, 1993.

[22] P. Sollich and A. Krogh, "Learning with ensembles : How over-fitting can be useful," *Advances in Neural Information Processing Systems*, vol. 8, pp. 190–196, 1996.

[23] K. Tumer and J. Ghosh, "Error correlation and error reduction in ensemble classifiers," *Connection Science*, vol. 8, no. 3-4, pp. 385–403, 1996.

[24] M. I. Jordan and R. A. Jacobs, "Hierarchical mixtures of experts and the EM algorithm, Tech. Rep. AIM-1440, 1993.

[25] M. I. J. L. Xu and G. E. Hinton, "An alternative model

for mixture of experts," *Advances in Neural Information Processing Systems*, pp. 633–640, 1995.

[26] M. Jordan and L. Xu, "Convergence properties of the em approach to learning in mixture-of-experts architectures," *Technical Report 9302, Department of Brain and Cognitive Science, MIT*, 1993.

[27] Z. Chair and P. Varshney, "Optimal data fusion in multiple sensor detection systems," *Multidimensional Systems and Signal Processing*, vol. AES-22, no. 1, pp. 9–31, 1986.

[28] B. Chen and P. Varshney, "A bayesian sampling approach to decision fusion using hierarchical models," *IEEE Transactions on Signal Processing*, vol. 50, no. 8, pp. 1809–1818, Aug. 2002.

[29] L. Xu, A. Krzyzak, and C. Y. Suen, "Methods of combining multiple classifiers and their applications to handwriting recognition," *IEEE Transactions on Systems, Man and Cybernetics*, vol. 22, no. 3, pp. 418–435, May 1992.

[30] K. Woods, W. P. Kegelmeyer Jr., and K. Bowyer, "Combination of multiple classifier using local accuracy estimates," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19, no. 4, pp. 405–410, Apr. 1997.

[31] C. Ji and S. Ma, "Combinations of weak classifiers," *IEEE Transaction on Neural Networks*, vol. 8, no. 1, pp. 32–42, Jan. 1997.

[32] M. Petrakos, J. A. Benediktsson, and I. Kanellopoulos, "The effect of classifier agreement on the accuracy of the combined classifier in decision level fusion," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 39, no. 11, pp. 2539–2546, Nov. 2001.

[33] Y. Huang and C. Suen, "A method of combining multiple experts for the recognition of unconstrained handwritten numerals," *IEEE Transaction on Pattern Analysis and Machine Intelligence*, vol. 17, no. 1, pp. 90–94, Jan. 1995.

[34] K. W. Bowyer, K. Chang, and P. Flynn, "A survey of approaches and challenges in 3D and multi-modal 3D + 2D face recognition," *Computer Vision and Image Understanding*, vol. 101, no. 1, pp. 1 – 15, 2006.

[35] J. Y. Cartoux, J. T. Lapreste, and M. Richetin, "Face authentification or recognition by profile extraction from range images," in *Proc. of Workshop on Interpretation of 3D Scenes*, 1989, pp. 194–9.

[36] J. C. Lee and E. Milios, "Matching range images of human faces," in *Proc. of Third Intl. Conf. on Computer Vision*, 1990, pp. 722–6.

[37] G. G. Gordon, "Face recognition based on depth and curvature features," in *Proc. IEEE Conf. on CVPR*, 1992, pp. 808–10.

[38] H. T. Tanaka, M. Ikeda, and H. Chiaki, "Curvature-based face surface recognition using spherical correlation. principal directions for curved object recognition," in *Proc. of Third IEEE Intl. Conf. on Automatic Face and Gesture Recognition*, 1998, pp. 372–7.

[39] G. Medioni and R. Waupotitsch, "Face modeling and recognition in 3-d," in *2003 IEEE Intl. Workshop on Analysis and Modeling of Faces and Gestures*, 2003, pp. 232–3.

[40] C. Hesher, A. Srivastava, and G. Erlebacher, "A novel technique for face recognition using range imaging," in *Proc. of the Seventh Intl. Symposium on Signal Processing and its Applications*, vol. 2, 2003, pp. 201–4.

[41] T. Heseltine, N. Pears, and J. Austin, "Three-dimensional face recognition: An eigensurface approach," in *Proc. of the Intl. Conf. on Image Processing*, vol. 5, 2004, pp. 1421–1424.

[42] I. A. Kakadiaris, G. Passalis, T. Theoharis, G. Toderici, I. Konstantinidis, and N. Murtuza, "Multimodal face recognition: Combination of geometry with physiological information," in *Proc. IEEE Conf. on CVPR*, vol. 2, 2005, pp. 1022–1029.

[43] K. I. Chang, K. Bowyer, and P. J. Flynn, "Adaptive rigid multi-region selection for handling expression variation in 3D face recognition," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05) - Workshops*, 2005, p. 157.

[44] G. Passalis, I. A. Kakadiaris, T. Theoharis, G. Toderici,

and N. Murtuza, "Evaluation of 3D face recognition in the presence of facial expressions: an annotated deformable model approach," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05) - Workshops*, 2005, p. 171.

[45] A. M. Bronstein, M. M. Bronstein, and R. Kimmel, "Three-dimensional face recognition," *Int. J. Comput. Vision*, vol. 64, no. 1, pp. 5–30, 2005.

[46] M. P. do Carmo, *Differential Geometry of Curves and Surfaces*. Prentice-Hall, 1976.

[47] M. Fels and P. J. Olver, "Moving coframes: I. a practical algorithm," *Acta Applicandae Mathematicae*, vol. 51, no. 2, pp. 161 – 213, 1998.

[48] ——, "Moving coframes: Ii. regularization and theoretical foundations," *Acta Applicandae Mathematicae*, vol. 55, no. 2, pp. 127 – 208, 1999.

[49] M. Faundez-Zanuy, "Data fusion in biometrics," *IEEE Aerospace and Electronic Systems Magazine*, vol. 20, no. 1, pp. 34–38, 2005.

[50] J. J. Clark and A. L. Yuille, *Data Fusion for Sensory Information Processing Systems*. Norwell, MA, USA: Kluwer Academic Publishers, 1990.

[51] L. I. Kuncheva, C. J. Whitaker, C. A. Shipp, and R. P. W. Duin, "Is independence good for combining classifiers?" in *ICPR*, 2000, pp. 168–171.

[52] P. J. Phillips, P. J. Flynn, T. Scruggs, K. W. Bowyer, J. Chang, K. Hoffman, J. Marques, J. Min, and W. Worek, "Overview of the face recognition grand challenge," in *Proc. IEEE Conf. on CVPR*, vol. 1, 2005, pp. 947–54.

[53] R. Duda, P. Hart, and D. Stork, *Pattern Classification*, 2nd ed.

[54] P. N. Belhumeur, J. Hespanha, and D. J. Kriegman, "Eigenfaces vs. fisherfaces: Recognition using class specific linear projection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19, no. 7, pp. 711–720, 1997.

[55] V. N. Vapnik, *The nature of statistical learning theory*. New York, NY, USA: Springer-Verlag New York, Inc., 1995.

[56] C.-C. Chang and C.-J. Lin, *LIBSVM: a library for support vector machines*, 2001. [Online]. Available: http://www.csie.ntu.edu.tw/ cjlin/libsvm

[57] P. Phillips, P. Flynn, T. Scruggs, K. Bowyer, and W. Worek, "Preliminary face recognition grand challenge results," in *7th International Conference on Automatic Face and Gesture Recognition*, 2006, pp. 15–24.

[58] P. Sinha, B. Balas, Y. Ostrovsky, and R. Russell, "Face recognition by humans: 20 results all computer vision researchers should know about," *(under review)*.

[59] T. Maurer, D. Guigonis, I. Maslov, B. Pesenti, A. Tsaregorodtsev, D. West, and G. Medioni, "Performance of geometrix activeid 3D face recognition engine on the FRGC data," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05) - Workshops*, vol. 3, 2005, pp. 154–154.

PLACE PHOTO HERE

Kin-chung Wong Kin-chung Wong received the MSc degree in electrical engineering from the University of Wisconsin-Madison in 2006. As part of his graduate research, he and other researchers in the University of Wisconsin Face Recognition Group worked together to develop an invariant-based 3D face recognition algorithm. His research interests are pattern recognition, image processing and computer vision.

PLACE PHOTO HERE

Wei-Yang Lin Wei-Yang Lin received BSEE from National Sun Yat-sen University, Taiwan, in 1994. He received MSEE and PhD degrees from University of Wisconsin-Madison in 2004, and 2006, respectively. Since 2006, he has been with the Department of Computer Science and Information Engineering, National Chung Cheng University, Taiwan, where he is currently an assistant professor. His research interests include computer vision, biometric authentication, and multimedia signal processing.

PLACE PHOTO HERE

Yu Hen Hu Yu Hen Hu received BSEE from National Taiwan University in 1976. He received MSEE and PhD degrees from University of Southern California in 1980, and 1982 respectively. From 1983 to 1987, he was an assistant professor at the Electrical Engineering Department of Southern Methodist University, Dallas, Texas. Since 1987, he has been with the Department of Electrical and Computer Engineering, University of Wisconsin, Madison where he is currently a professor.

Dr. Hu has broad research interests ranging from design and implementation of signal processing algorithms, computer aided design and physical design of VLSI, pattern classification and machine learning algorithms, and image and signal processing in general. He has published more than 200 technical papers, edited several books in these areas.

He has served as an associate editor for the IEEE Transaction of Acoustic, Speech, and Signal Processing, IEEE signal processing letters, European Journal of Applied signal Processing, IEEE Multimedia Magazine, and Journal of VLSI Signal Processing. He has served as the secretary and an executive committee member of the IEEE signal processing society, a board of governors of IEEE neural network council representing the signal processing society, the chair of signal processing society neural network for signal processing technical committee, and is the current chair of IEEE signal processing society multimedia signal processing technical committee (2004-2005). He is also a steering committee member of the international conference of Multimedia and Expo, IEEE Transactions on Multimedia on behalf of IEEE Signal processing society.

Dr. Hu is a fellow of IEEE.

PLACE PHOTO HERE

Nigel Boston Nigel Boston grew up in England and attended Cambridge and Harvard. After a year in Paris and two in Berkeley, he went to the University of Illinois for twelve years and then joined the faculty at the University of Wisconsin. His original work was in algebraic number theory. In recent years, he has moved towards engineering, holding joint appointments in the Mathematics and Electrical and Computer Engineering Departments at UW, where he has collaborated on coding theory, cryptography, face recognition, and watermarking and was founding director of the Wireless and Sensor Networks (WiSeNet) Consortium. Since August, he has been Williams-Hedberg-Hedberg Chair of Mathematics at the University of South Carolina.

PLACE PHOTO HERE

Xueqin Zhang Xueqin Zhang, born in 1972,China, got Master degree at 1998 in mechanics. Now I am a PH.D candidate in Detection Techniques and Automatic Devices at East China University of Science and Technology(ECUST),Shanghai,China.

After I graduated from ECUST at 1998, I became a tutor in the same university. Now I am an assistant professor in the department of electron and communication engineering of ECUST. My research interests are machine learning and information security.