Passive Multimodal 2-D+3-D Face Recognition Using Gabor Features and Landmark Distances

Sina Jahanbin, *Member, IEEE*, Hyohoon Choi, *Member, IEEE*, and Alan C. Bovik, *Fellow, IEEE*

***Abstract*—We introduce a novel multimodal framework for face recognition based on local attributes calculated from range and portrait image pairs. Gabor coefficients are computed at automatically detected landmark locations and combined with powerful anthropometric features defined in the form of geodesic and Euclidean distances between pairs of fiducial points. We make the pragmatic assumption that the 2-D and 3-D data is acquired passively (e.g., via stereo ranging) with perfect registration between the portrait data and the range data. Statistical learning approaches are evaluated independently to reduce the dimensionality of the 2-D and 3-D Gabor coefficients and the anthropometric distances. Three parallel face recognizers that result from applying the best performing statistical learning schemes are fused at the match score-level to construct a unified multimodal (2-D+3-D) face recognition system with boosted performance. Performance of the proposed algorithm is evaluated on a large public database of range and portrait image pairs and found to perform quite well.**

***Index Terms*—Classifier fusion, face recognition, fiducial detection, Gabor wavelets, geodesic distances, range images.**

I. Introduction

SINCE Bledsoe's pioneering "man–machine" system in mid 1960s [1], many face recognition algorithms have been proposed. The majority of those implemented in the early days were based on 2-D (intensity/portrait) images of the face [2]. Although several sophisticated 2-D solutions have been implemented, unbiased evaluations, such as the Face Recognition Vendor Test (FRVT) 2002 [3] show that their collective performances are unsatisfactory, degrading significantly with variations in illumination, head pose, cosmetics, or non-neutral facial expressions [3]. Recent developments in 3-D sensor technology has made the acquisition of 3-D models cheaper, quicker, and more reliable. Since the 3-D shape of the face surface remains invariant to changes in head orientation and illumination, researchers are turning to 3-D images to overcome some of the limitations associated with 2-D images [4].

In fact, the majority of early 3-D approaches were extended versions of holistic 2-D approaches in which the portrait images are replaced by range images [5]. Typically, the input range images were aligned and then reformatted into a feature vector. The dimensionality of these high-dimensional feature vectors would be reduced using standard statistical dimensionality reduction techniques [6], [7]. The main drawback of traditional holistic approaches is their sensitivity to deformations (e.g. open mouth), facial expressions, and improper alignment of faces [5]. In recent years, many of the problems associated with traditional holistic approaches have been mitigated by introducing region ensembles approaches [8]–[12]. In region ensembles approaches, faces are divided into multiple smaller subregions. Comparisons between two faces begin by independently comparing the corresponding subregions. Finally, the committee of similarity scores collected from various subregions are merged into a final decision. Region ensembles approaches are considered as a compromise between holistic and local face recognition approaches.

Since many 3-D scanners also capture portrait images, attempts have been made to further increase the accuracy of face recognition systems by combining portrait and range modalities. The 2-D+3-D face recognition literature is limited and the dominant trend is to create parallel recognition systems using one of the above-mentioned algorithms, eventually combining two systems by some classifier fusion technique [13], [14]. These holistic approaches are generally sensitive to changes in facial expression, deformation in the mouth area (open/closed mouth), and improper scaling or alignment of faces [5]. Other authors have combined PCA/LDA-based 2-D face recognition with a surface matching technique, such as the iterative closest point (ICP) algorithm for 3-D face matching [15], [16]. Surface matching techniques have been shown to be very successful, but their traditional treatment may require heavy computation, especially when the face recognition system is performing identification on a database with a large gallery. In traditional surface matching approaches, the matching routine is called to compute a match score between the probe image and every image in the gallery. More recent algorithms such as [10]–[12] only invoke the surface matching procedure once to align the probe image with a fixed reference model.

Compared to holistic approaches, face recognition algorithms based on local features are generally more robust against changes in facial pose, illumination, noise, holes (missing range values), occlusion, and facial expressions, especially if the features are extracted from rigid parts of the face [5]. One

Manuscript received October 14, 2010; revised June 17, 2011; accepted June 23, 2011. Date of publication July 22, 2011; date of current version November 18, 2011. This work was supported by Dr. Kenneth Castleman. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Patrick J. Flynn.

S. Jahanbin was with the Laboratory for Image and Video Engineering (LIVE), Department of Electrical and Computer Engineering, The University of Texas at Austin, Austin, TX 78712-1084 USA. He is now with KLA-Tencor, San Jose, CA 95035 USA (e-mail: sina.jahanbin@utexas.edu).

H. Choi is with the Samsung Electro-Mechanics R&D Institute, Suwon 443-743, South Korea (e-mail: hyohoon.choi@samsung.com).

A. C. Bovik is with the Laboratory for Image and Video Engineering (LIVE), Department of Electrical and Computer Engineering, The University of Texas at Austin, Austin, TX 78712-1084 USA (e-mail: bovik@ece.utexas.edu).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TIFS.2011.2162585

successful 2-D face recognition algorithm based on local features is the Elastic Bunch Graph Matching (EBGM) proposed by Wiskott *et al.* [17]. In EBGM, faces are represented by labeled graphs where each node corresponding to a fiducial point is labeled with a set of complex Gabor responses (called a jet). Each edge of the graph connecting two nodes is represented by a 2-D displacement vector encoding the topology of the graph. A graph is extracted from a new portrait image by iteratively fitting a general face graph representation, called a face bunch graph (FBG), to the input image such that a target matching function becomes optimized. This similarity function takes into account both Gabor jet similarities and topological deformations of the graphs. Recognition of a face is performed by comparing its graph with all model graphs available in the gallery and selecting the one with the highest similarity. However, potentially useful geometric information provided by pairwise fiducial point displacements is neglected in EBGM.

Motivated by EBGM's excellent performance on 2-D face recognition, as well as by its robustness against pose and expression variations, we propose a framework for multimodal face recognition algorithm whereby local Gabor features from coregistered range and portrait image pairs are used as essential features. Our algorithm also takes into account anthropometric features in the form of Euclidean and geodesic distances (and their ratios) between automatically detected fiducial points. Geodesic distances between fiducials are expression invariant features since recent studies suggest that changes in facial expression can often be well modeled as isometric deformations of facial surfaces under which intrinsic properties, including the geodesic distances between fiducial points, remain unchanged [18]. By contrast to EBGM, we apply statistical feature analysis to 2-D Gabor, 3-D Gabor, and Euclidean/geodesic anthropometric feature sets to select the most discriminative features while discarding redundancies. The results of face recognition using these three modalities are fused at the match score-level using a weighted sum technique to construct a multimodal (2-D+3-D) face recognition system with boosted performance. An underlying assumption of our approach is that the acquired portrait and range data are perfectly coregistered. This is a reasonable assumption, since there now exist 2-D+3-D face imaging systems that passively acquire this type of data. In our view, such passive systems are greatly preferable to laser-based ranging systems, not only because the registration issue is eliminated, but since less invasive passive optical systems are more realistic for real-world applications than those that image the face using lasers.

Accurate localization of facial landmarks plays a crucial role in automation of face recognition approaches that use local features [19]. In EBGM [17], fiducial point detection is computationally expensive since fiducial points can move relative to each other without any restrictions. Wang *et al.* [20] suggested that EBGM's fiducial detection stage could be accelerated by restricting the search range of each fiducial. In our framework, Gabor wavelet responses are instrumentally used to detect landmark points having meaningful anthropometric definitions by combining range and portrait driven attributes. Following Wang *et al.*'s [20] proposition, fiducial point localization is accelerated by restricting the search range of each fiducial. Details of the proposed fiducial point detection step are explained in Section II.

Deploying Gabor wavelets to decompose range images and using geometric features (e.g. distances, ratios, and angles between fiducial points) are topics of recent interest in 3-D face recognition. For example, Xu *et al.* [21] proposed a face recognition algorithm based on Gabor wavelet responses extracted from range and portrait images at five scales and eight orientations. First, range and portrait images are aligned and normalized with respect to a generic face model. Gabor features are calculated at every range and portrait image coordinates, resulting in a very high-dimensional feature vector for each modality (80 times the number of pixels in the image). Raw 2-D and 3-D Gabor features are hierarchically projected to lower dimensional subspaces, then fused at feature-level using a combination of LDA and AdaBoost learning [22]. As with other global algorithms, this one is sensitive to misalignments, global face deformations, facial expressions, and missing range data (which commonly occur in regions of face covered by hair). The extremely high-dimensional feature vector cannot be managed by conventional statistical learning techniques such as LDA or PCA. Handling very high-dimensional features is also computationally expensive (e.g., in [21] at least 80 LDAs are applied in the first level of the hierarchy). In our proposed framework, the Gabor features are extracted from a small number of fiducials, resulting in feature vectors of reasonable size.

Cook *et al.* [23] proposed an algorithm to combine complementary 2-D and 3-D face data to enhance recognition accuracy. In this algorithm, range and portrait images are convolved with log-Gabor filters. Each face is partitioned into 25 square windows on which PCA is applied to learn subspaces from the log-Gabor responses. Each window is considered as an independent observation where the Mahalanobis-cosine distance is used to measure the similarity between probe and gallery windows. All Mahalanobis-cosine distances resulting from comparison of a probe and a gallery face are collected into a feature vector utilized by a linear support vector machine (SVM). Similar to other holistic algorithms which partition the face into subregions, this algorithm is sensitive to face alignment, scale, and pose normalization. The authors show that 3-D log-Gabor features perform significantly better than 2-D log Gabors. However, their feature fusion scheme is not able to effectively exploit complementary range and portrait features, since 2-D+3-D performance is barely improved compared to the 3-D alone.

Wang *et al.* [24] considered face intensity $I(x, y, z)$ to be a function of 3-D spatial coordinates and defined 3-D spherical Gabor filters (3DSGF) to analyze them. Unlike traditional Gabors, 3DSGFs are spherically symmetric, hence lack the directional selectivity property. For a given point on the face, the 3DSGF responses collected from neighboring points are collectively represented by their histogram. No automatic fiducial detection algorithm is presented; instead 3DSGF histograms are calculated, and compared using the "Least Trimmed Hausdorff" distance [25]. As with our approach, the resulting recognition algorithm assumes perfectly coregistered range and portrait image pairs to operate. Evaluation is performed on a database of 960 coregistered range and portrait pairs from 80 subjects captured using stereo imaging technology. In our view,

a drawback of this algorithm is the lack of directionality in the proposed 3DSGF, where Gabor analysis is performed only by changing the kernels' scales. Also, the proposed 3DSGFs are not DC-free causing the face recognition to be sensitive to illumination variations. Instead of extracting features at fiducial points, they are extracted at nodes of a rectangular grid making the algorithm sensitive to pose variations and misalignment.

Hüsken *et al.* [26] applied EBGM [17] separately to range and portrait images to create parallel 2-D EBGM and 3-D EBGM face recognizers. These two recognizers are combined at the match score-level by summing the similarity scores at the classifiers' outputs. Similar to 2-D EBGM [17], the landmark detection step is computationally expensive since search areas of fiducials are not restricted. Although shape and portrait data provide complimentary information, fiducial detection in each modality is independent and no attempt is made to combine them. As with 2-D EBGM, important geometric features, such as the geodesic distances between landmarks, are neglected in this proposed 2-D+3-D EBGM. Gabor features collected from each modality are not analyzed by statistical learning methods to separate discriminatory features from the redundant ones.

In an earlier work [27], we implemented a 3-D face recognition algorithm based on Gabor features extracted at automatically detected landmarks and compared its performance to the corresponding 2-D Gabor-based counterpart. The framework proposed in this paper extends our earlier work in multiple aspects. For example, in [27], Euclidean/geodesic anthropometric features between detected landmarks were not considered. Also in [27], statistical feature analysis was not conducted to choose the most discriminant subset of 3-D and 2-D Gabor features. Instead, a suboptimal regularization LDA method [28] was used to lower the feature dimensionality. Finally, the face recognition algorithms proposed in [27] are unimodal approaches and no effort was made to combine them into a unified multimodal system with enhanced performance.

Geometric features based on distances, angles, and ratios between facial landmarks have also been recently considered. Gupta *et al.* [29] proposed a face recognition algorithm based on geodesic and 3-D Euclidean distances between ten automatically annotated anthropometric fiducial points ("Anthroface 3-D"). "Anthroface 3-D" is not a pure 3-D approach, since portrait clues are used to assist the fiducial detection step. Most of the fiducials (seven out of ten) in Gupta *et al.*'s algorithm are detected using the primary version [30] of the Gabor-based fiducial detection algorithm proposed in this paper. Similarly, Riccio *et al.* [31] proposed a face recognition algorithm based on ratios of Euclidean distances between landmarks on 2-D or 3-D images. The drawback of their algorithm is that landmark detection is performed manually.

We propose a novel hierarchical framework to combine a diverse set of local attributes into a multimodal face recognizer with enhanced performance. Since fiducial detection directly impacts the performance of a local feature-based face recognizer, we start by proposing a novel algorithm combining range and portrait Gabor clues to pinpoint landmarks with significantly better accuracies than those achievable using a single modality alone. Several statistical learning techniques are evaluated to independently reduce the dimensionality of each feature source by keeping the most discriminative features. The three parallel face recognizers resulting from applying the best performing statistical learning schemes are fused at the match score-level to enhance the overall performance. The ultimate goal of this work is to deploy and evaluate a multimodal face recognition algorithm and its constituting subsystems, such as fiducial point detection and statistical feature analysis without focusing on preprocessing steps involved in database preparation. Performance of the proposed algorithms are evaluated and reported on a large public database of range and portrait image pairs described in Section II-B1.

II. Automatic Facial Landmark Detection From Portrait and Range Images

The automatic detection of facial feature points (fiducials) plays an important role in applications such as human–machine interaction and face recognition. Motivated by such practical applications, extensive work has focused on developing methods for automatic feature localization on 2-D images of the face which contain facial texture and color information. The active appearance model (AAM) by Cootes *et al.* [32] is one of the most effective facial landmark detection algorithms on 2-D images. An iterative search algorithm seeks the best location for each feature point using a texture model describing that feature point's surrounding. These feature locations are then fine-tuned using the spatial configuration of feature points stored in a statistical shape model. In a later work, Cristinacce *et al.* [33] improved the AAM algorithm and showed that their new shape optimized search (SOS) algorithm outperforms AAM.

Compared to fiducial detection on 2-D facial images, automatic 3-D facial landmarking is a newer research topic. In some of the earlier 3-D face recognition algorithms, the landmark detection procedure has been founded on questionable heuristics. For example, in [34] and [35], it is assumed that the nose tip is the closest point to the 3-D sensor, hence it has the smallest depth value in the resulting range image. Such assumptions may fail when the landmarking step erroneously detects a streak of hair or the tip of a protruding chin as the nose tip. Many other face recognition algorithms utilize geometric features such as the distances and angles defined between prominent facial landmarks [31], [36] that are pinpointed manually.

In recent years, a variety of 3-D facial landmarking algorithms founded on better heuristics have been introduced. Koudelka *et al.* [37] used a radial symmetry map, gradient zero-crossing maps, and a set of physical constraints to locate the inner eye corners, nose tip, mouth, and sellion point on range images. Evaluation of this algorithm on a dataset containing 943 range images indicates that the average positional error over all detected landmarks is 3.55 mm, and 97.22% of the detected landmarks fall within 10 mm from the ground truth.

Curvature information has been considered by several researchers to detect landmarks on 3-D facial data [38]–[41]. In Chang *et al.*'s 3-D face recognition algorithm [41], the eye pits, nose tip, and bridge of the nose are detected based on surface curvatures computed at each point. Prior to curvature computation, a local coordinate system is established at each point by an-

alyzing a small region around that point using PCA. A quadratic surface is fit to each local region and the mean and Gaussian curvature are estimated from the fit coefficients. The facial surface is then segmented into smaller regions of interest based on curvature type, and landmarks are eventually found in their corresponding regions.

Faltemier *et al.* [42] proposed the "Rotated Profile Signature" (RPS) to detect the nose tip on range images across a large degree of pose angle variations. In RPS, 3-D faces are rotated 180° around the vertical axis at 5° intervals. The right-most face "profile" is extracted at each interval and compared to a variety of nose profile models. As the nose rotates into view, the similarity score between the extracted profile and the nose model peaks, ostensibly leading to the detection of the correct nose tip location.

The body of work focusing on using facial fiducial point localization using a combination of range and portrait information is even more limited. Boehnen *et al.* [43] proposed an algorithm to detect the eyes, nose, and the mouth using complementary range and portrait images. This algorithm uses color portrait images to segment the skin regions. Detected skin regions are refined using the corresponding range information. Subsequently, eye and mouth maps are defined by operations in color space. Eventually, candidate eye and mouth locations are detected by processing these maps. The putative location of the nose is found as the point having the largest distance in 3-D space from the plane containing the eyes and mouth points. Detection accuracies in the 83.33% to 99.61% range are achieved depending on the quality of the portrait images in the testing set.

Inspired by the success of (EBGM) [17] in 2-D face recognition, we extend the concept to model the local appearance of portrait and range images around high-information fiducial points using Gabor coefficients. To detect fiducial points for each modality (range or portrait), a vector of Gabor coefficients, called a "jet," is computed at each pixel in the corresponding search window and compared with a template of that fiducial, called a "bunch." A bunch consists of Gabor coefficients collected from several training images by manually marking that specific fiducial. The desired feature point is located at the pixel whose jet is most similar to the training bunch. We are able to show that more accurate landmarking is possible by summing the similarity scores of range and portrait.

To the best of our knowledge this is the first time that complementary range and portrait Gabor-based appearance clues have been used simultaneously to detect fiducial points on pairs of coregistered range and portrait images. We have tested our 2-D, 3-D, and 2-D+3-D landmarking algorithms on 1146 pairs of range and portrait images from the Texas 3-D Face Recognition Database (T3FRD) [44]. High detection accuracy is achieved using a small number of training images and it is shown that colocalization using Gabor jets on range and portrait images results in better accuracy than using any single image modality. The obtained accuracy is competitive with other techniques in the literature.

A. Background

Gabor filters model the receptive field profiles of cortical simple cells implicated in the decomposition and analysis of visual signals [45]. They have been used successfully for many image analysis tasks since then [46]–[48].

1) Gabor Jets: The local appearance around a point $\vec{x}$ in a gray-scale range or portrait image $I(\vec{x})$ can be encoded using a set of Gabor coefficients $J_j(\vec{x})$ [46]. Each coefficient $J_j(\vec{x})$ is derived by convolving the input image $I(\vec{x})$ with a family of Gabor kernels

$$\psi_j(\vec{x}) = \frac{k_j^2}{\sigma^2}\exp\left(\frac{-k_j^2 x^2}{2\sigma^2}\right)\left[\exp(i\vec{k}_j\cdot\vec{x}) - \exp\left(\frac{-\sigma^2}{2}\right)\right]. \tag{1}$$

Gabor kernels are plane waves modulated by a 2-D Gaussian function. In our implementation, $\sigma = 2\pi$ and each kernel is characterized by a wave vector $\vec{k}_j = [k_v\cos\phi_u \;\; k_v\sin\phi_u]^T$, where $k_v = 2^{-(v+1)}$, $v = 0, 1\ldots, 4$ denote spatial frequencies, and $\phi_u = (\phi/8)u$, $u = 0, 1\ldots, 7$ are the orientations of the Gabor kernels.

A "jet" $\vec{J}$ is a set $\{J_j, j = u + 8v\}$ of 40 complex Gabor coefficients obtained from a single image point. Complex Gabor coefficients are represented as $J_j = a_j\exp(i\phi_j)$, where $a_j(\vec{x})$ is the slowly varying magnitude and $\phi_j(\vec{x})$ is the phase of the jth Gabor coefficient at $\vec{x}$.

The similarity between two jets is effectively measured by the phase sensitive similarity measure

$$S(\vec{J}, \vec{J'}) = \frac{\sum_{i=1}^{40} a_i \acute{a}_i \cos(\phi_i - \acute{\phi}_i)}{\sqrt{\sum_{i=1}^{40} a_i^2 \sum_{i=1}^{40} \acute{a}_i^2}}. \tag{2}$$

This similarity measure returns real values in the range $[-1, +1]$, where a value close to $+1$ means a high similarity between the input jets.

2) Gabor Bunch: To search for a given feature on a new face image, a general representation of that fiducial point is required. As proposed in [17], the general appearance of each fiducial point can be modeled by bundling the Gabor jets extracted from several manually marked examples of that feature point (e.g., eye corners) collected from multiple subjects in a stack-like structure called a Gabor "bunch."

In order to support a wide range of variations in the appearance of faces caused by subjects' different gender, race, age, and facial expression, a comprehensive training set should be selected. For example, the Gabor bunch representing an eye corner should contain jets from open, closed, male, female, and other possible eye corners. In this work, a training set containing 50 pairs of registered portrait and range images was selected to cover possible variations present in the data set.

The similarity measure between a jet and a bunch is naturally defined to be the maximum of the similarity values between the input jet and each constituent jet of that bunch

$$S_B(\vec{J}, \vec{\vec{B}}) = \max_{i=1}^{50} S\left(\vec{J}, \vec{B}_{(i)}\right) \tag{3}$$

where in (3), $\vec{\vec{B}}$ represents a bunch, while $\vec{B}_{(i)}$, $i = 1, \ldots, 50$ are its constituent jets.

B. Materials and Methods

1) Texas 3-D Face Recognition Database: We approach the 2-D+3-D face recognition problem under the premise that

data is acquired using a passive modality, such as optical stereo ranging. We view this as a desirable assumption, owing to the high accuracy that is attainable, the fact that perfectly coregistered portrait and range images can be acquired simultaneously, and the less invasive nature of the acquisition process.

In this paper, we address only the problem of recognition (including fiducial detection, feature extraction, statistical learning, and classification), and not preprocessing of the face data. We tested our algorithms on the Texas 3-D Face Recognition Database (T3FRD) [44]. In our view, since preprocessing and recognition are separate problems, algorithms to accomplish these tasks should be tested separately. T3FRD is a large recently released public database of coregistered 2-D and 3-D face images that is finding significant use. T3FRD is available to researchers free of charge. It contains 1196 pairs of roughly aligned, high resolution, colored portrait and range images from 116 adult subjects. The images were captured at the former Advanced Digital Imaging Research (ADIR) LLC (Friendswood, TX) using a MU-2 stereo imaging system made by 3Q Technologies Ltd. (Atlanta, GA), under contract to NIST.

All the portrait and range pairs in the database are roughly aligned with respect to a fixed generic face in the canonical frontal pose using the iterative closest point (ICP) algorithm [49]. Since all of the 3-D facial images in T3FRD are aligned by an identical procedure, head pose variations are limited and fair comparisons can be made between competing algorithms based on recognition or fiducial detection capabilities only, without having to deal with biases introduced by preprocessing. The database includes images of male and female subjects (68% male and 32% female) with ages ranging from 20 to 75 years old. About 80% of the images in T3FRD are from subjects younger than 40 years of age. T3FRD contain images from different ethnicities with the following mix: 40% of images are of Caucasian subjects, 5% of Africans, 32% of Asians, 22% of East-Indians, and the rest belong to other ethnicities. The database contains 69% neutral and 31% expressive faces. Each range and portrait pair is accompanied by a file containing information about the subject's gender, ethnicity, facial expression, and manually annotated locations of 25 anthropometric facial fiducial points. Examples of range and colored portrait pairs are presented in Fig. 1. The left two columns are image pairs from a single subject captured at two different sessions, one neutral and the other with an expression. The image pair in the right-most column is an example of a subject with an expressive face and missing surface information on areas covered with facial hair. Since camera settings will drift out of calibration throughout the data acquisition, portrait images in T3FRD exhibit a wide range of illumination levels although these have not been measured. Such variations exemplified by the two portrait images of a single individual (the first and second column of Fig. 1) captured several days apart.

T3FRD complements the older publicly available Face Recognition Grand Challenge (FRGC) 2005 database [50] and is a good alternative for researchers who want to evaluate their innovative face recognition algorithms without dealing with extensive preprocessing such as head pose normalization and scaling as required by the FRGC dataset. It is also appropriate for application frameworks such as ours, where the data is obtained with perfect portrait-range registration. It allows isolation of face recognition performance without any bias introduced by choice of complicated preprocessing schemes. It is the largest (in terms of number of images and subjects) database that has been acquired using a stereo imaging system at a high resolution of 0.32 mm along the x, y, and z dimensions. By comparison, images in the FRGC database were acquired using a Minolta Vivid 900/910 laser scanner sampled at a lower average resolution of 0.98 mm along z and y dimensions and 0.5 mm along the z dimension [13].

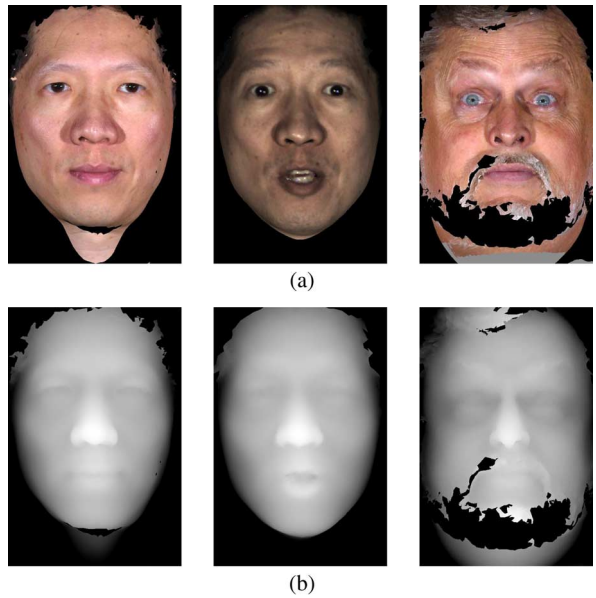


Fig. 1. Example of face images from the Texas 3-D Face Recognition Database. (a) Colored portrait images. (b) Corresponding range images.

Regarding the FRGC data set, during the acquisition of data there was a significant time-lapse between the operation of the laser range finder and the optical camera in the FRGC data acquisition, which caused the acquired 2-D and 3-D images to often be out of correspondence [50]. This time-lapse also caused inconsistencies in facial expressions between the range and portrait images captured in single subject sessions [51]. In addition, since laser scanning is not instantaneous, some of the faces in the FRGC have been reported to be distorted due to head movements during acquisition [51]. By contrast, stereo imaging captures both the shape and the portrait image of the face simultaneously, hence each range and portrait pair are perfectly coregistered in the T3FRD. Furthermore, accuracy assessment of fiducial detection requires access to publicly available ground-truth of manually pinpointed fiducial points, which is not provided by FRGC. Using FRGC for evaluation of fiducial detection algorithm would have required manually pinpointing 11 fiducials on 5000 portrait and 5000 range images. Finally, 3-D face acquisition using laser range finders can cause emotional or physical discomfort in those being scanned, and in our view, is a modality that is highly unlikely to be deployed often in practice. Much of the preprocessing required to utilize the FRGC database is unlikely to be required using a more realistic passive sensing system, such as stereo ranging systems.

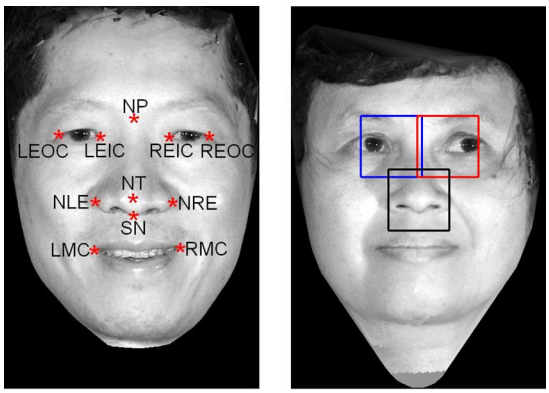


Fig. 2. Example of face images from the Texas 3-D Face Recognition Database. (a) Eleven facial fiducial points manually annotated on a portrait image. (b) Search areas of the nose tip and inner eye corners.

A few preprocessing steps are performed on all of the T3FRD range and portrait images: First, extraneous background regions that are not connected to the face are removed by detecting the face as the biggest connected blob having nonzero range value, and eliminating all remaining smaller blobs. Impulse noise present in the range images are removed by median filtering with a 3×3 square filter. Holes and areas with missing data are filled by bicubic interpolation. All images in the database are of size 751×501. The z value in the range images is represented using 8-bit format with closest point to the camera having the highest value of 255. The portrait images are represented in uncompressed 8-bit RGB format. In our experiments, we reduced the size of all images by a factor of 3 in each direction to reduce computational cost, so the resulting images used were of size 251×167 pixels. Finally, the 2-D colored portrait images were transformed to gray-scale portrait images.

2) Training of Automatic Feature Point Detection Algorithm: To train and evaluate our fiducial point detection algorithm, we partitioned the 1196 pairs of face images in T3FRD into disjoint testing and training sets. A set of 50 pairs of registered range and gray-scale portrait images covering a variety of facial appearances from subjects with different ages, races, and genders was selected for the training and Gabor bunch extraction. This training set contains neutral and expressive faces among which many examples have open/closed mouth or eyes.

We manually marked 11 prominent fiducial points only on the portrait images of these 50 training pairs. Since these portrait and range images are perfectly aligned, the location of fiducials on the range image of the pair is exactly the same as the portrait one. We use the following terminology to refer to these 11 fiducial points. LEIC: left eye inner corner; REIC: right eye inner corner; LEOC: left eye outer corner; REOC: right eye outer corner; LMC: left mouth corner; RMC: right mouth corner; NT: nose tip; NLE: nose left end; NRE: nose right end; SN: subnasal; NP: nasion point. Fig. 2 shows a portrait image from the training set with 11 fiducial points marked with red "*".

Finally, Gabor jets were calculated from images of each modality at the manually marked landmarks. All Gabor jets from a specific feature point (e.g. nose tip) and modality are stacked together to create a bunch representation of that fiducial in that modality. For example, the nose tip's range-bunch describes the nose tip in the range images.

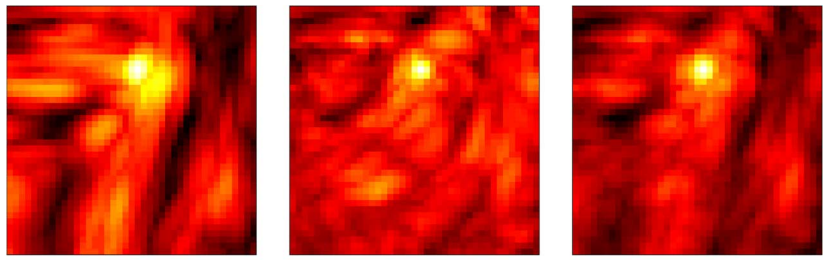
3) Localization Method: In elastic bunch graph matching [17], the "search area" of each feature point is not constrained, causing the iterative optimization algorithm involved to be computationally expensive. Since all of the face images in the T3FRD are coarsely aligned to a frontal view, prior knowledge about the human face can be used to limit the search area of each feature point. For example, the nose tip is expected to be located at the center of the image and the left eye's inner corner is always located above and to the left of the nose tip.

In this work, each fiducial point is searched for over an area centered at the average location of that fiducial in the training data. Each search area is a rectangle box of size 40×40 pixels in the down-sampled 251×167 pixel images (120×120 on the original images). The sides of these rectangular areas are at least 5 times the standard deviation of each fiducial's coordinates in the training set. Our results show that the search window is reasonably large and all fiducials are located in their expected search area. In Fig. 2, the search area of the nose tip and the inner corners of the eyes were marked with rectangular boxes. Similar constraints are used by Lu and Jain [52] and Gupta *et al.* [29] to reduce the search area for fiducial points.

In order to automatically locate a fiducial point on a pair of range and portrait images which have never been seen before, the range and portrait data enclosed by the search area of that feature point are first convolved with the set of 40 Gabor wavelets in (1). As a result, each pixel of the search area is represented by Gabor jets, a "range jet" and a "portrait jet." Next, The jets for each modality are compared to their corresponding bunch using the similarity measure between a jet and a bunch in (3). Consequently, a similarity map is created for each modality demonstrating the similarity between each pixel in the search area and the appropriate bunch describing the appearance of the target feature point. Fig. 3(a) and (b) depicts portrait and range similarity maps resulting from comparing jets in the LEIC search box of an arbitrary subject. Fiducial point detection based on an individual modality (portrait only or range only) can be done by picking the pixel with the highest similarity value in the corresponding similarity map. In an earlier version of the proposed landmark detection algorithm [30], portrait and range similarity maps were combined into a hybrid similarity map by taking the pixel-wise maximum of the similarity scores. In order to improve the performance, we experimented with other strategies such as taking the sum and product of the pixel-wise similarity scores trying to combine range and portrait cues in landmark detection. Ultimately, we decided that the range and portrait information should be combined by summing the pixel-wise similarity scores. The target landmark is located where this sum reaches its maximum. Fig. 3(c) shows the hybrid similarity map from LEIC from the same subject.

C. Accuracy of Feature Point Detection

The landmark detection algorithm was tested on the remaining 1146 pairs of facial images available in T3FRD. The correct location of the feature points is available in the



(a) (b) (c)

Fig. 3. Similarity maps of LEIC: (a) Range similarity map. (b) Portrait similarity map. (c) Hybrid similarity map calculated by pixel-wise summation of portrait and range similarity maps.

TABLE I
STATISTICS OF THE EUCLIDEAN DISTANCE BETWEEN THE AUTOMATICALLY DETECTED FEATURE POINTS AND THE MANUALLY LABELED GROUND-TRUTH

	Mean (mm)			Std (mm)		
fiducial	portrait	range	hybrid	portrait	range	hybrid
LEIC	1.45	1.65	1.17	1.9	1.7	1.0
REIC	1.35	1.63	1.09	1.5	1.5	1.0
LEOC	1.83	3.93	1.47	2.8	4.5	1.8
REOC	1.49	3.77	1.37	1.9	4.5	1.3
LMC	1.45	1.89	1.31	1.7	1.5	1.1
RMC	1.64	1.81	1.35	2.0	1.9	1.2
NLE	1.26	1.04	1.01	1.3	0.7	0.7
NRE	1.18	0.96	0.92	1.2	0.9	0.7
NT	1.26	1.40	1.18	0.9	0.9	0.8
SN	1.07	1.01	0.93	1.1	0.7	0.7
NP	3.56	2.67	2.40	3.6	1.9	1.7
Nasal	1.19	1.10	1.01	1.1	0.8	0.7
Overall	1.59	1.98	1.29	2.1	2.5	1.2

supplemental file accompanying each range and portrait pair in T3FRD. In order to evaluate the accuracy of the automatic fiducial point detection algorithm, the Euclidean distance between the automatically detected landmarks and manually annotated ground-truth was measured in millimeters (mm). Table I provides statistics on the positional error that occurred in the detection of each individual or group of facial fiducial points.

It is evident from the statistics in Table I that combining portrait and range improves detection accuracy, since the mean and standard deviation of the average positional error is reduced significantly as compared to an individual detection strategy for any fiducial point. When considering all 11 landmarks ("overall" row of Table I), detection using the portrait modality outperforms the range with an average positional error of 1.59 mm for portrait-based versus 1.98 mm for range-based detection. Combining the two modalities significantly reduce the overall mean positional error to 1.29 mm. It is interesting to note that portrait texture information provides more powerful cues for finding the eye corners (LEIC, REIC, LEOC, REOC) and mouth corners (LMC, RMC) which are more deformable in the range modality and might be closed or open. By contrast, range data appears to be more relevant for detecting four fiducial points defined on and around the nose tip (NLE, NRE, NP, and SN) which are rigid and do not change much with facial expression variations. From Table I, we can see that the "Nasal"

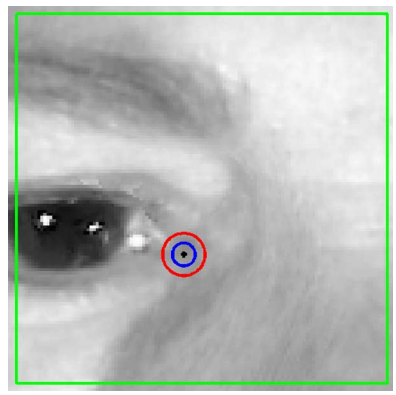


Fig. 4. Correct location of LEIC is shown with a black dot in the search area marked with green square. The blue circle (inner circle) shows the mean hybrid positional error and the red circle (outer circle) marks the mean plus one standard deviation of the hybrid positional error.

group (consisting of NLE, NRE, NT, and SN) and the inner eye corners (LEIC and REIC) are detected more accurately than average. These results reflect the high accuracy of our facial fiducial point detection method. In fact, our algorithm is able to detect the NLE, NRE, and SN with mean positional errors close to 1 mm by combining range and portrait information.

In order to better visualize the accuracy of our fiducial point detection algorithm, we have projected the boundary of the search box for the left inner corner of the eye (LEIC) on an arbitrary portrait image from the T3FRD with a green square in Fig. 4. In this figure, the correct location of LEIC is marked with a black dot. The radius of the blue circle (inner circle) in Fig. 4 is equal to 1.17 mm which is the mean positional error of detection on LEIC using a hybrid of range and portrait information (Table I). The red circle (outer circle) has a radius equal to mean 1.17 mm plus one standard deviation 1.0 mm.

In order to assess the landmark detection accuracy, we normalized the positional error of each detected fiducial by dividing the error by the interocular distance of that face. The normalized positional errors averaged over a group of fiducial points is denoted m_e, adopting the same notation as in [33]. Fig. 5 shows the cumulative distribution of normalized positional error m_e averaged over various groups of fiducial landmarks. As is evident from all five plots, the red curve corresponding to the combination of range and portrait always lies above the individual detections, indicating that combining portrait and range information boosts the performance for all landmark groups. It appears that for "nasal" points only, the blue curve corresponding

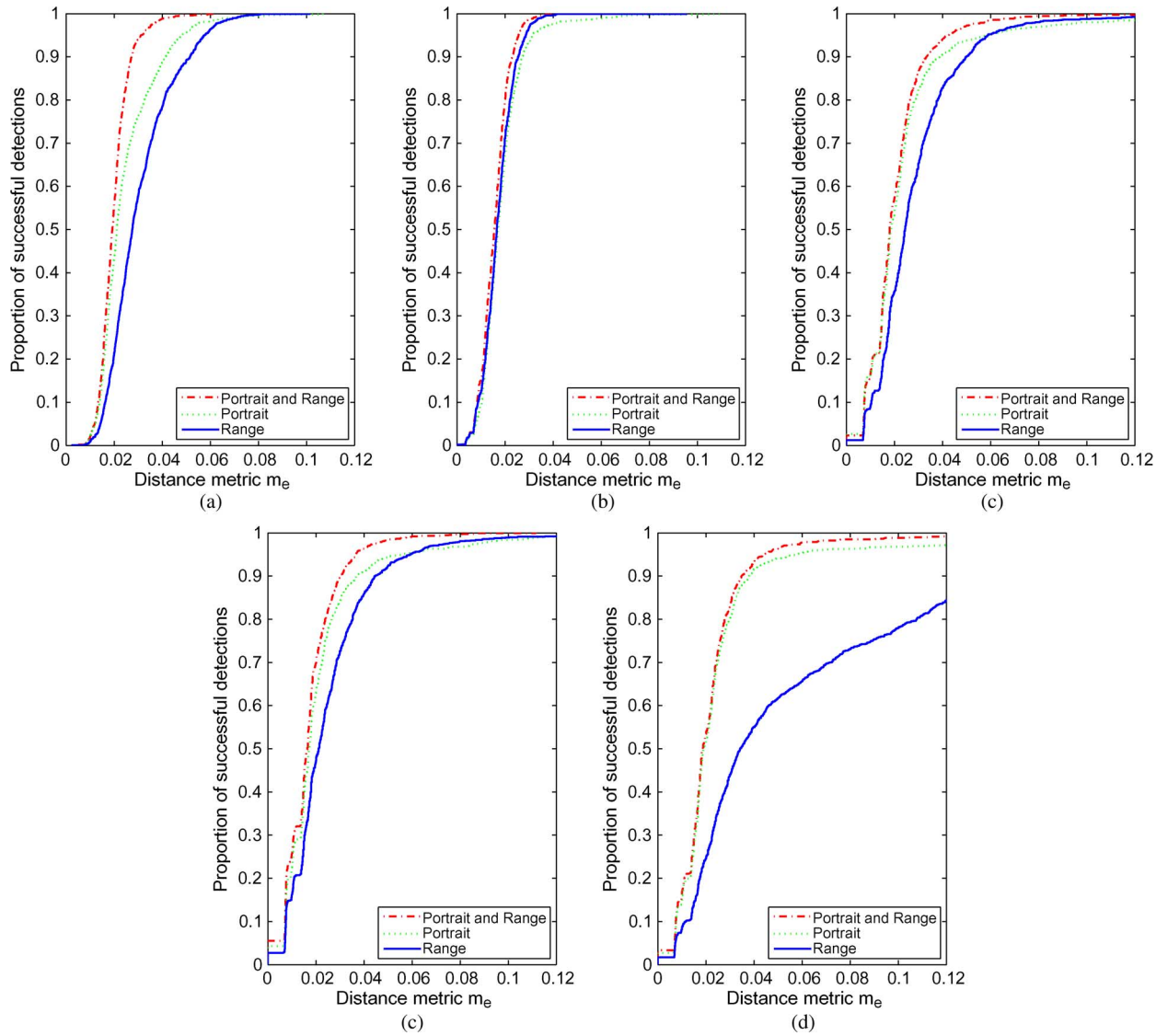


Fig. 5. Average positional error (m_e) for (a) all fiducial points, (b) only nasal fiducial points, (c) mouth fiducial points, (d) inner eye fiducial points, and (e) outer eye fiducial points.

to range-based detection is higher than the green curve of portrait-based detection. This means that for nasal points with rigid structure, range information is more important, while for the rest of the fiducial points, portrait is more meaningful.

Fig. 5(a) shows the cumulative probability distribution of m_e averaged over all feature points. This figure indicates that when the acceptable detection error is $m_e \leq 0.03$, the search algorithm based only on range information is successful for 58% of faces, whereas the success rate of a localization algorithm based only on portrait information is 76%. The combination of range and portrait information results in a success rate as high as 94% of faces.

Fig. 5(b) demonstrates the cumulative distribution for four nasal fiducials (i.e., nose tip, subnasal, and nose left/right corners). This figure indicates that by using only portrait information, nasal fiducials are successfully located on more than 93% of faces given that $m_e \leq 0.03$. From Fig. 5(b), it is apparent that using range information outperforms portrait information when finding nasal fiducials for any given value m_e. Range-based detection is successful on 97% of faces when $m_e \leq 0.03$. Table II summarize these results.

TABLE II
PERCENTAGE OF SUCCESSFUL DETECTION WHEN ACCEPTED NORMALIZE ERROR IS EQUAL TO OR LESS THAN 0.03

$m_e \leq 0.03$	portrait	range	hybrid
Overall	76%	58%	94%
Nasal Points	93%	97%	99%
Inner Eye Corners	84%	73%	90%
Mouth corners	83%	65%	87%
Outer eye Corners	80%	44%	83%

The body of papers addressing facial fiducial point localization using a combination of range and portrait information is very limited. Lu and Jain [52] proposed an algorithm to detect seven fiducials (corners of the eyes, corners of the mouth, and nose tip) on perfectly coregistered range and portrait pairs in

TABLE III
COMPARISON OF EUCLIDEAN DISTANCE ERROR STATISTICS

		LEIC	REIC	LEOC	REOC	NT	LMC	RMC
Lu *et al.* [52]	Mean	5.7	6.0	7.9	7.1	5.0	3.6	3.6
	Std	3.0	3.3	5.1	5.9	2.4	2.9	3.3
Gabor based	Mean	1.2	1.1	1.5	1.4	1.2	1.3	1.4
	Std	1.0	1.0	1.8	1.3	0.8	1.1	1.2

frontal upright pose. The first step in their algorithm is to use the range information to detect the nose tip by finding the profile. Once the nose tip is detected, prior knowledge about the placement of landmarks with respect to the nose tip is used to constrain the search region of each fiducial to an ellipsoid around the average location of that fiducial with an axis equal to 1.5 times the standard deviation learned from a training set. Finally, a cornerness feature from the portrait image and a shape index from the range image are combined to determine fiducials in their corresponding search area. This algorithm was evaluated on 946 portrait and range pairs of 267 subjects. Table III shows the statistics of Lu's detection [52] versus ours. It is apparent that our Gabor-based algorithm outperform in detection of any of the seven fiducial points they used.

The performance of our novel feature point localization algorithm is quite competitive with such well-known portrait-based algorithms as AAM [32] and SOS [33]. Cristinacce *et al.* [33] has compared the performance of AAM versus SOS for detecting 17 features on 1521 portrait images. They used a very large training set containing 1055 face images, as compared to our algorithm which needs only 50 image pairs. They reported that when the acceptable normalized positional error is $m_e \leq 0.1$, AAM is successful on 70% and SOS works for 85% of faces. Whereas, with $m_e \leq 0.1$, the success rate of our proposed method exceeds 99% for any fiducial using any combination of range or portrait modalities (see Fig. 5).

Our detection algorithm employs a simple search method with remarkable performance in finding fiducial points on expressive and neutral facial images. Our proposed localization algorithms require a small set of representative faces for training. Since the computational cost is low, it is suitable for real-time applications.

III. MULTIMODAL FACE RECOGNITION USING GABOR FEATURES, DISTANCES, AND CURVATURES BETWEEN LANDMARKS

In Section II, we proposed an automatic fiducial point detection algorithm to locate 11 landmarks using a combination of both portrait and range Gabor jets. Once these landmarks are identified, the following local features can be extracted from the image pairs.

A. Features

1) Geodesic and Euclidean Distances and Global Curvatures: One important pose invariant characteristic defined between each pair of fiducials is their geodesic distance (length of the shortest path on the surface). Fig. 6 shows the shortest path between the nasion point (NP) and the right mouth corner (RMC) on an arbitrary face in the database. It is claimed that facial expressions can be modeled as isometric transformations of the face under which the geodesic distances between points on the face remain constant [18], [36]. We used the "fast marching" algorithm [53] to measure the 55 geodesic distances between all possible pairs of the 11 fiducial points detected by the automatic "hybrid detection" strategy. We also calculated 55 Euclidean distances and the ratio of the geodesic to the Euclidean distance as a global measure of curvature of the geodesic paths. Eventually, the pairwise geodesic distances, Euclidean distances, and the global curvatures calculated between landmarks detected by the proposed "hybrid fiducial detection" strategy are concatenated to create a 165-dimensional feature vector.

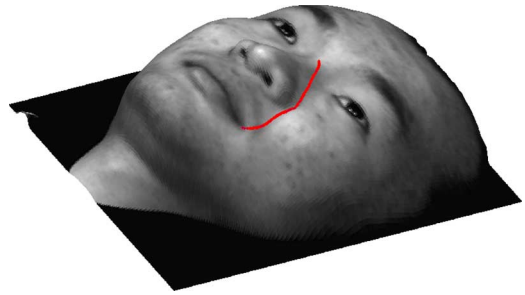


Fig. 6. Shortest path between the NP and RMC shown on an arbitrary face.

2) Gabor Coefficients: The portrait-based and the range-based Gabor wavelet coefficients at the fiducial points detected by the "hybrid fiducial detection" strategy can be used to identify the individual appearing in the image. The absolute value of the Gabor wavelet coefficients at all 11 fiducials detected using a combination of range and portrait Gabor clues are concatenated in each modality (portrait or range), creating two independent 440-dimensional feature vectors per face.

B. Face Recognition Data

In order to evaluate the verification and identification performance of the proposed features, we partitioned the 1196 pairs of images from 116 subjects present in the T3FRD into disjoint training and testing sets. The training set contains 270 pairs of portrait and range images from 18 subjects (15 image pairs per subject). The training set was used to learn the best projection direction using several LDA variants proposed in the literature, or to choose the most discriminative subset of features using stepwise LDA [54]. The test set was further divided into a gallery set and a probe set. The gallery set contains 103 portrait and range pairs from 103 individuals (1 pair per subject) who have at least 2 pairs of images in the test set. The probe set has 810 image pairs from the 103 enrolled subjects in the gallery set, where the number of images per subject varies, with a minimum of 1 pair for some subjects to maximum of 77. There are 13 subjects in the T3FRD with only 1 pair of images per subject, which are included as imposters in the probe set.

C. Statistical Learning Algorithms

As explained in Section III-A, the magnitudes of the complex Gabor jets at detected landmarks are concatenated to create 440-dimensional real-valued feature vectors from each

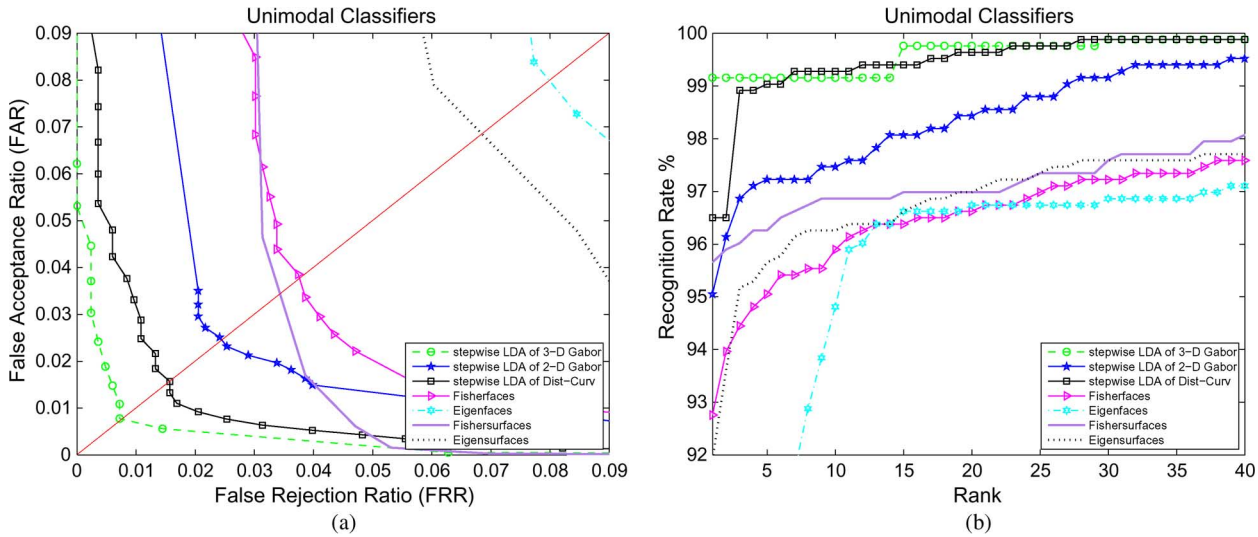


Fig. 7. Performance evaluation of unimodal face classifiers with landmarks detected by "hybrid fiducial detection." (a) ROC curves showing the verification accuracies. (b) CMC curves showing the identification performance.

modality (range and portrait). Dimensionality reduction techniques like PCA and LDA [55] can be used to project the high-dimensional feature vectors to lower dimensional spaces. Unlike PCA, which pools all the training samples regardless of their class affiliations to find the orthogonal basis vectors having least reconstruction error, LDA searches for basis vectors that best discriminate between classes and often has better performance in classification applications [55]. Unfortunately when using LDA with Gabor features (similar to many other realistic face recognition approaches), the "small sample size" problem (SSS) may occur. This problem arises when the total number of training samples is less than the dimensionality of the features. Then, the within-class scatter matrix can become singular, and the computation of projection directions may fail.

We implemented and evaluated several statistical learning techniques such as the regularization method [28], PCA followed by LDA [56], LDA in the Null Space of the within-class scatter matrix [57], D-LDA [58], and stepwise-LDA [54] which are commonly used in the literature to overcome the SSS problem. The evaluation results indicate that the stepwise-LDA outperforms its alternatives when applied to range and portrait Gabor features. Stepwise LDA iteratively adds and removes features so as to maximize the ratio of the between-class to within-class separation. The procedure terminates when no more features can be added or removed based on the specific significance level for the entry or removal of features. In our implementation, both significant levels are set to 0.02.

D. Performance Evaluation of Unimodal Face Classifiers

To measure the accuracy of our face recognition algorithms, we have arranged both verification and identification experiments according to established face recognition evaluation protocols [59]. In a verification scenario, the face recognition system compares the retrieved features of the claimed identity (from the gallery set) with the currently captured features of the user and decides whether a match can be declared based on the Euclidean distance between features in the lower dimensional space. The performances of face recognition algorithms in the verification experiment are evident from their receiver operation characteristics (ROC). The ROC curve displays the trade-off between the false acceptance rate (FAR) and the false rejection rate (FRR) for various operating points (distance thresholds). Two quantitative measures of verification performance, the equal error rate (EER) and area under the curve (AUC), were measured for each overall face recognition algorithm used in this work. In the Identification scenario, the features of the subject are compared to all the enrolled models in the galley set and the results are ranked based on the Euclidean distance of the features in the projection subspace. The identification performance of proposed algorithms are presented by the cumulative match characteristic (CMC) curves and the rank 1 recognition rates (RR1).

The performance of the unimodal face recognition classifiers implemented using range Gabor, portrait Gabor, and pairwise geometric features (Dis_Curv) calculated at fiducial points detected by the "hybrid fiducial detection" strategy are displayed by CMC and ROC curves in Fig. 7. Table IV summarizes the EER, AUC, and RR1 values for these unimodal algorithms. For comparison purposes, four baseline face classifiers are evaluated: The baseline Eigensurfaces and Fishersurfaces are implemented by applying PCA and LDA to the z values of the range images, and the baseline Eigenfaces and Fisherfaces are implemented by applying PCA and LDA to the intensity values. As mentioned before, stepwise-LDA has the best performance among the statistical learning techniques applied to the range Gabor features with AUC $= 0.0005$ and EER $= 0.76\%$. This verification performance shows significant improvement compared to the corresponding range-based baselines (Eigensurfaces and Fishersurfaces) with corresponding AUC $= 0.0294$, EER $= 6.85\%$ and AUC $= 0.0143$, EER $= 3.44\%$. Similarly, the identification performance of stepwise-LDA applied to the range Gabor features is significantly better than the baseline Eigensurfaces and Fishersurfaces algorithms. Range Gabor features selected by stepwise-LDA produce rank 1 recognition

TABLE IV
OBSERVED EER, AUC, AND RR1 VALUES FOR UNIMODAL ALGORITHMS WITH LANDMARKS DETECTED BY "HYBRID FIDUCIAL DETECTION"

Algorithm	EER (%)	AUC	RR1 (%)
Eigensurfaces	6.85	0.0294	91.91
Eigenfaces	7.99	0.0342	80.92
Fishersurfaces	3.44	0.0143	95.65
Fisherfaces	3.77	0.0104	92.75
Stepwise LDA of 3-D Gabor	0.76	0.0005	99.15
Stepwise LDA of 2-D Gabor	2.45	0.0044	95.05
Stepwise LDA of Dist_Curv	1.57	0.0012	96.50

results as high as RR1 $= 99.15\%$ while the rank 1 recognition rates of baseline Eigensurfaces and Fishersurfaces do not exceed 95.65%.

Similarly, the portrait Gabor features collected at landmarks detected by the "hybrid fiducial detection" and selected by stepwise-LDA outperform the baseline 2-D face recognition algorithms, Eigenfaces, and Fisherfaces, in both identification and verification tasks. In verification mode, the portrait Gabor features selected by stepwise-LDA have AUC $= 0.0044$ and EER $= 2.45\%$, while Eigenfaces and Fisherfaces have AUC $= 0.0342$, EER $= 7.99\%$ and AUC $= 0.0104$, EER $= 3.77\%$. In the identification test, stepwise LDA has rank 1 recognition of RR1 $= 95.05\%$ while the performances of the baseline algorithms do not exceed RR1 $= 92.75\%$.

We compared the identification and verification performance of stepwise-LDA applied to range Gabor features extracted from fiducials (detected by "hybrid detection") with the performance of stepwise-LDA applied to portrait Gabor features at landmarks also detected by "hybrid detection." As reflected in Table IV and Fig. 7, it is evident that range Gabor features significantly outperform portrait Gabor features. Fig. 7 and Table IV also summarize the identification and verification performance of the Dist_Curv classifier which uses a discriminative subset of geodesic/Euclidean distances and curvatures, calculated from fiducials detected by "hybrid detection", selected by stepwise LDA. With EER $= 1.54$, AUC $= 0.0012$, and RR1 $= 96.50\%$, the Dist_Curv classifier performs better than the portrait Gabor-based classifier but cannot reach the performance of stepwise-LDA applied to the range Gabor features.

E. Fusion of Unimodal Face Classifiers

Finally, we integrated the information available from different feature sources into a more accurate multimodal face recognition system. Fusion in biometric systems can be accomplished at the sensor level, the feature level, the match score-level, or the decision level [60]. In this study, we use fusion at the match score-level via a weighted sum approach (the most common approach in multibiometrics systems) to combine the following classifiers given that features are extracted at fiducial points automatically detected using the proposed "hybrid detection" algorithm:

1) stepwise LDA of 3-D Gabor;
2) stepwise LDA of 2-D Gabor;
3) stepwise LDA of Dist_Curv.

The scores that are fused in this study are the Euclidean distances, d_i, $i = 1, 2, 3$, measured by each classifier. However, a combination of distances (match scores) is only meaningful when the distances are in the same range. We have used *min-max* normalization to transform distances obtained from each classifier to a common range. The weighted sum fusion of distances D_f is calculated

$$D_f = \sum_{j=1}^{3} w_j * d_j^n \tag{4}$$

where d_j^n and w_j are the normalized distance and weight of the jth classifier, respectively, with the condition $\sum_{j=1}^{3} w_j = 1$. Each classifier's weight is a function of its performance estimated using the training data

$$w_j = \frac{1 - (\mathrm{FAR}_j + \mathrm{FRR}_j)}{3 - \sum_{i=1}^{3}(\mathrm{FAR}_i + \mathrm{FRR}_i)} \tag{5}$$

where FAR_i and FRR_i are the false acceptance and false rejection rates of the ith classifier. In this study, we have used the constant EER_i instead of FAR_i and FRR_i that are threshold dependent. The weights calculated for "stepwise LDA of 3-D Gabor," "stepwise LDA of 2-D Gabor," and "stepwise LDA of Dist_Curv" are correspondingly equal to $w_1 = 0.3390$, $w_2 = 0.3274$, and $w_3 = 0.3336$ indicating the importance of each classifier present in this fusion.

For comparison purposes, two baseline multimodal face classifiers were implemented. The first baseline multimodal classifier is constructed by combining the Eigenfaces and Eigensurfaces (PCA applied to intensity and range values) via the weighted sum approach explained above. The second baseline multimodal classifier combines Fisherfaces and Fishersurfaces using the weighted sum approach.

Fig. 8 and Table V reflect the significant boost achieved in the verification and identification performance by fusion of different modalities. The verification improvement is evident as the fused classifier, "2-D Gabor+3-D Gabor+Dist_Curv," has EER $= 0.25\%$ and AUC $= 0.0001$ which is significantly reduced as compared to each individual classifier present in the fusion. The rank 1 recognition rate of "2-D Gabor+3-D Gabor+Dist_Curv" also increased to RR1 $= 99.76\%$, which outperforms each contributing classifier. The baseline multimodal classifiers are not close to our proposed multimodal classifier. The baseline "Eigenfaces + Eigensurfaces" has enhanced identification and verification performance compared to baseline Eigenfaces and Eigensurfaces, but its performance is not comparable to our proposed unimodal classifiers. The multimodal "Fisherfaces + Fishersurfaces" has EER $= 3.11\%$, AUC $= 0.0119$, and RR1 $= 95.90\%$ which also suffers in comparison to "2-D Gabor+3-D Gabor+Dist_Curv."

The multimodal classifier, "2-D Gabor + 3-D Gabor + Dist_Curv," using features calculated from fiducials automatically detected by "hybrid detection" algorithm achieves excellent performance with EER $= 0.25\%$ and AUC $= 0.0001$ and rank 1 recognition rate of RR1 $= 99.76\%$. Table VI compares the results of the proposed multimodal face recognition algorithm with several state-of-the-art algorithms and baseline algorithms that have been evaluated on T3FRD. By comparison, Gupta *et al.* [29] proposed a face recognition algorithm based on geodesic and 3-D Euclidean distances between 10

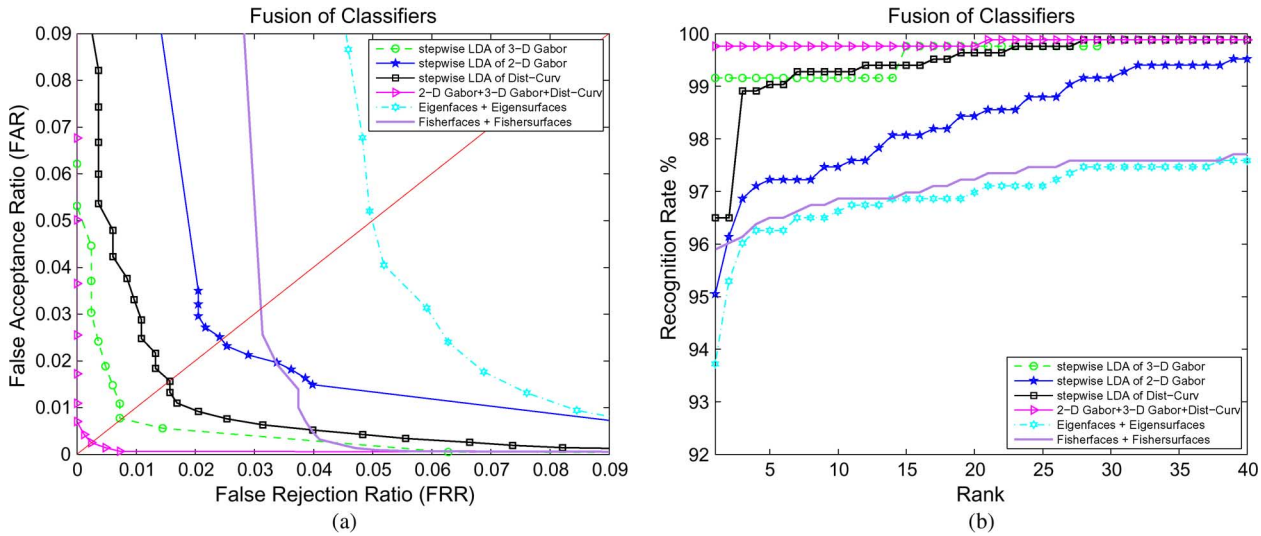


Fig. 8. Performance evaluation of the baseline multimodal classifiers and the proposed unimodal and multimodal face classifiers with landmarks detected by "hybrid fiducial detection." (a) ROC curves showing the verification accuracies. (b) CMC curves showing the identification performance.

TABLE V
EER, AUC, AND RR1 VALUES FOR THE BASELINE MULTIMODAL CLASSIFIERS AND THE PROPOSED UNIMODAL AND MULTIMODAL FACE CLASSIFIERS WITH LANDMARKS DETECTED BY "HYBRID FIDUCIAL DETECTION"

Algorithm	EER (%)	AUC	RR1 (%)
Eigenfaces + Eigensurfaces	5.00	0.0229	83.72
Fisherfaces + Fishersurfaces	3.11	0.0119	95.90
Stepwise LDA of 3-D Gabor	0.76	0.0005	99.15
Stepwise LDA of 2-D Gabor	2.45	0.0044	95.05
Stepwise LDA of Dist_Curv	1.57	0.0012	96.50
2-D Gabor+3-D Gabor+Dist_Curv	0.25	0.0001	99.76

TABLE VI
OBSERVED EER, AUC, AND RR1 VALUES FOR ALGORITHMS EVALUATED ON THE TEXAS 3-D FACE RECOGNITION DATABASE

Algorithm	EER (%)	AUC	RR1 (%)
2-D Gabor+3-D Gabor+Dist_Curv	0.25	0.0001	99.76
"Anthroface 3D" [29]	1.65	0.0014	97.3
Warped Examples [61]	2.5	NA	NA
Eigenfaces + Eigensurfaces [62]	5.00	0.0229	83.72
Fisherfaces + Fishersurfaces	3.11	0.0119	95.90

automatically annotated anthropometric facial fiducial points ("Anthroface 3-D"). Similaly, stepwise LDA [54] is used in "Anthroface 3-D" to select the most discriminative features from raw Euclidean and geodesic distances. Unlike "2-D Gabor + 3-D Gabor + Dist_Curv," Gabor features are not considered as face recognition features in "Anthroface 3-D." The performance of Anthroface 3-D was also evaluated on T3FRD with comparable size training, probe, and gallery sets. The highest performance reported by Gupta *et al.* [29] is achieved on the recognition of neutral faces yielding EER $= 1.65\%$ and AUC $= 0.0014$ in a verification experiment and RR1 $= 97.3\%$ in an identification experiment. Our proposed multimodal face recognition approach has significantly better performance than the algorithms in [29].

Le Zou *et al.* [61] introduced "warping coefficients," a 3-D face recognition system based on warped range images. In this algorithm, a number of selected range images constitute a set of example faces, while another range image is selected as a "generic face." The generic face is then warped to match each of the example faces. Each such warp is specified by a vector of displacement values. In the feature extraction phase, when the algorithm is provided with a new range image, the generic face is warped to match it. The geometric transformation used in this warping can be approximated as a linear combination of example face warping vectors. The coefficients in the linear combination are used as features and passed to a Mahalanobis-distance based classifier. The "warping coefficients" achieved EER $= 2.5\%$ when evaluated using a subset of range images available in T3FRD. AUC and RR1 were not reported in [61].

IV. ROBUSTNESS EVALUATIONS

In Sections III-D and III-E, the performances of the proposed unimodal and multimodal face recognition algorithms were studied under the assumption that the fiducial points were automatically detected by fusing range and portrait Gabor information. The performance evaluations were conducted using a probe set containing both expressive and neutral faces. In the following subsections, we discuss the effects of automatic landmark detection on overall face recognition performance. We also evaluate the robustness of the proposed face recognition algorithms against facial expression changes. Finally, we assess the generalization capabilities of the proposed face recognition model by using test and training sets that are subject independent.

A. Sensitivity to Fiducial Detection Errors

In order to assess the effects of the "hybrid fiducial detection" errors on the performance of the proposed unimodal and multimodal face recognition, the algorithms are re-evaluated using the same probe and gallery sets (containing neutral and expressive faces) when the landmarks are pinpointed manually. Fig. 9 and Table VII summarize the face recognition results achieved in the absence of landmarking errors. Comparing the results in Tables VII and V, it is evident that the performance of each

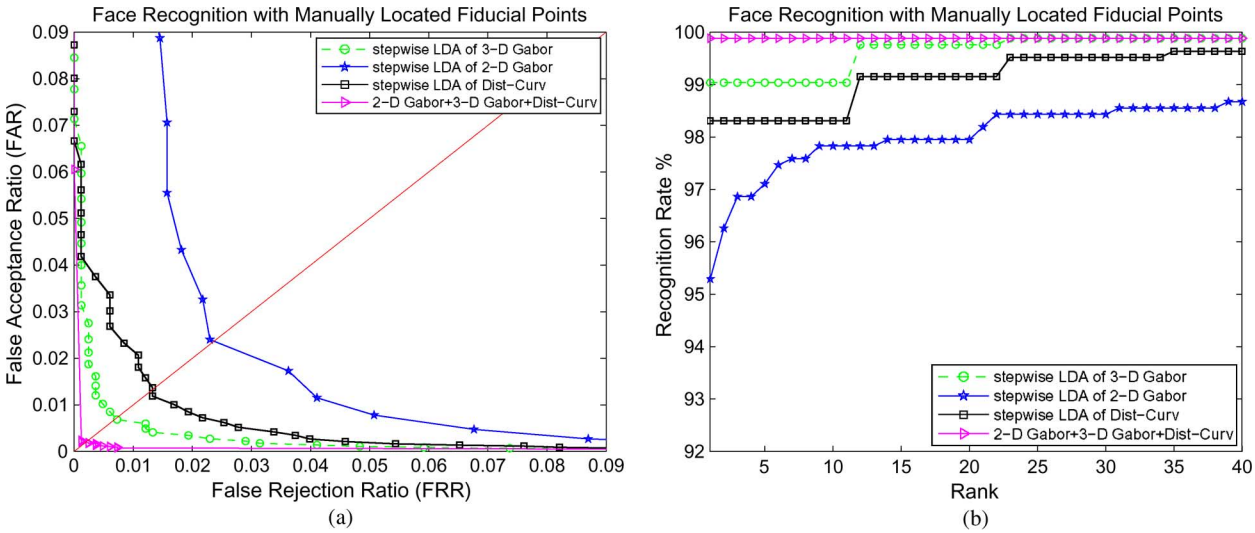


Fig. 9. Performance evaluation of the proposed unimodal and multimodal face classifiers with "manually" detected landmarks. (a) ROC curves showing the verification accuracies. (b) CMC curves showing the identification performance.

TABLE VII
EER, AUC, AND RR1 VALUES FOR THE PROPOSED UNIMODAL AND MULTIMODAL FACE CLASSIFIERS WITH "MANUALLY" DETECTED LANDMARKS

Algorithm	EER (%)	AUC	RR1 (%)
Stepwise LDA of 3-D Gabor	0.71	0.0003	99.03
Stepwise LDA of 2-D Gabor	2.36	0.0034	95.29
Stepwise LDA of Dist_Curv	1.33	0.0007	98.31
2-D Gabor+3-D Gabor+Dist_Curv	0.18	0.0001	99.88

unimodal face recognition algorithm is degraded slightly due to the "hybrid detection" landmarking errors. The performance declines observed in the unimodal recognizers eventually take a toll on the performance of the eventual multimodal face recognizer that combines the match scores created by the contributing unimodal components. In the absence of landmarking errors, the multimodal, "2-D Gabor + 3-D Gabor + Dist_Curv," face recognizer achieves excellent EER $= 0.18\%$ and AUC $= 0.0001$ in the verification test and rank one recognition rate as high as 99.88% in the identification test. These results can be considered as an upper bound for the performance of the proposed algorithms.

In order to evaluate the benefits achieved in face recognition performance by the improved accuracy of the "hybrid fiducial detection" as compared with unimodal landmarking schemes, we conducted an experiment in which 3-D Gabor features are extracted from fiducial points detected using only range information, 2-D Gabor features are extracted from fiducial points detected using only portrait information, and Dist_Curv feature are calculated from points detected using only portrait information. The results achieved by this fictional setup are presented in Fig. 10 and Table VIII. Comparing the results in Tables VIII and V, it is evident that the more accurate fiducial detection provided by combining range and portrait features has improved the performance of each unimodal face recognition component. This gain in face recognition performance is more evident when the final multimodal face recognizer benefiting from hybrid landmark detection is compared with its counterpart using a less accurate unimodal landmarking strategy. For example, hybrid landmark detection reduced the observed equal error rate from EER $= 0.57\%$ to EER $= 0.25\%$. Although the "hybrid fiducial detection" algorithm has improved the results, "2-D Gabor + 3-D Gabor + Dist_Curv" still achieved competitive results with a less accurate unimodal landmark detection algorithm. This strongly suggests that the proposed statistical feature analysis and classifier fusion strategy play a large role in delivering excellent results.

B. *Sensitivity to Facial Expression*

In order to evaluate the robustness of the proposed multimodal face recognition algorithm and its unimodal components against facial expression variations, an experiment was conducted in which subjects in T3FRD are enrolled in the gallery set as a pair of neutral portrait and range images, and the probe set contains only expressive portrait and range pairs. The results of expressive face recognition experiments are summarized in Fig. 11 and Table IX. Comparing the results in Tables IX and V, it is evident that the unimodal face recognition components perform less reliably when subjected to expressive faces. This performance slip is also evident in the performance of the final multimodal face recognizer as the observed verification equal error rate increase from EER $= 0.25\%$ to EER $= 0.59\%$ and the rank one recognition rate falls from RR1 $= 99.76\%$ to RR1 $= 98.70\%$.

Although the performance of the proposed multimodal, "2-D Gabor + 3-D Gabor + Dist_Curv," reduces when facing expressive faces, this reduction is marginal. The proposed multimodal face recognizer still achieves better performance in recognizing expressive faces than do other algorithms in recognizing combinations of expressive and neutral faces from the same database (Table VI). The robustness of the proposed algorithm against facial expression changes can be explained by the fact that this algorithm is founded on local features that are robust against such adverse factors. Naturally, the success of any local feature-based face recognition depends on how accurately fiducial points are

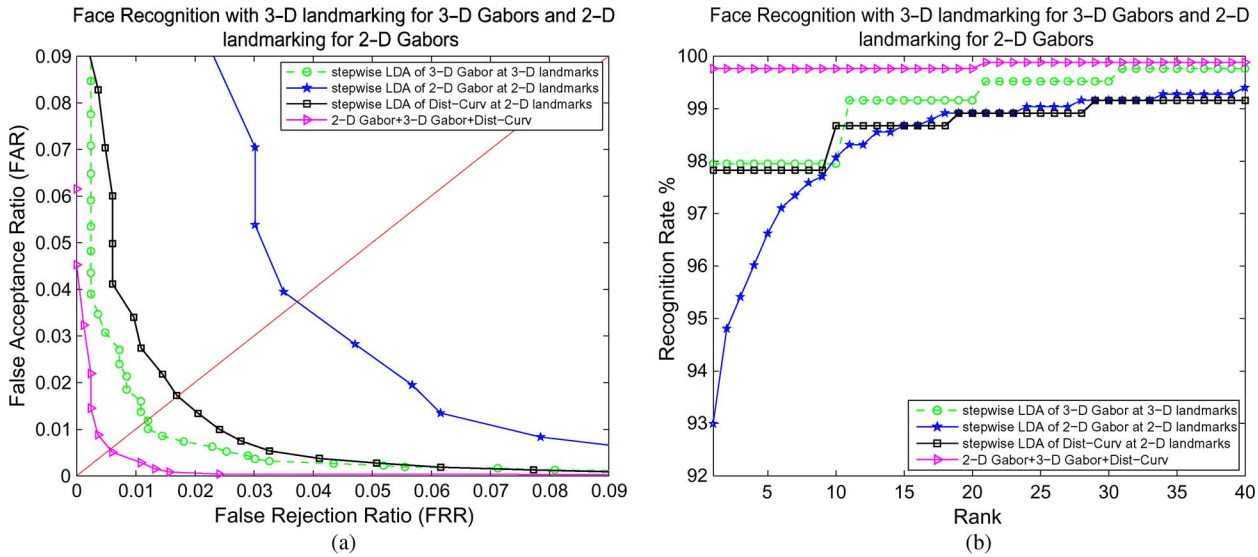


Fig. 10. Performance evaluation of the proposed unimodal and multimodal face classifiers with less accurate unimodal landmark detections. (a) ROC curves showing the verification accuracies. (b) CMC curves showing the identification performance.

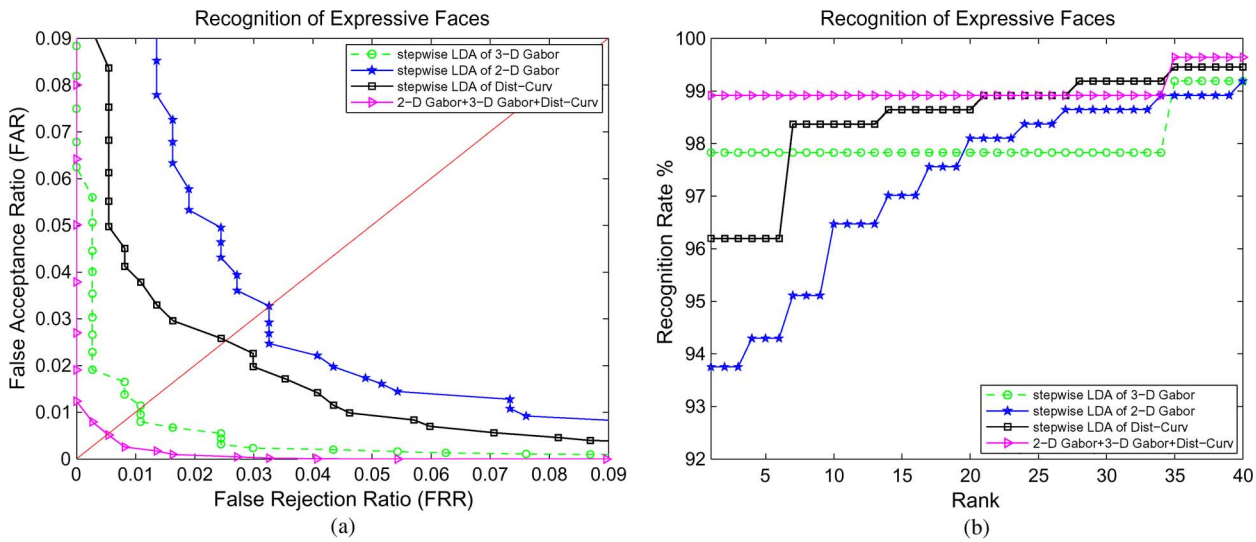


Fig. 11. Performance evaluation of the proposed unimodal and multimodal face classifiers in recognizing "expressive" faces. (a) ROC curves showing the verification accuracies. (b) CMC curves showing the identification performance.

TABLE VIII
EER, AUC, AND RR1 VALUES FOR THE PROPOSED UNIMODAL AND MULTIMODAL FACE WITH LESS ACCURATE UNIMODAL LANDMARK DETECTIONS

Algorithm	EER (%)	AUC	RR1 (%)
3-D Gabor at 3-D landmarks	1.20	0.0008	97.95
2-D Gabor at 2-D landmarks	3.73	0.0091	92.99
Dist_Curv at 2-D landmarks	1.71	0.0012	97.82
2-D Gabor+3-D Gabor+Dist_Curv	0.57	0.0002	99.70

detected. In order to ensure good performance of the face recognition algorithm, a landmarking algorithm was proposed which combines portrait and range information to accurately detect fiducial on both neutral and expressive faces. Finally, the proposed algorithm hierarchically selects and combines a rich set of diverse local features collected from different regions of the face into a robust face recognition algorithm.

C. Generalization Capabilities

In order to assess the generalization capabilities of the proposed unimodal and multimodal face recognition models, performance evaluations were conducted on a "training-subject independent" testing set that does not contain any images from the 18 subjects present in the training phase (e.g., stepwise-LDA training). After removing training dependent subjects, the "training subject independent" testing set contains 300 pairs of range and portrait images from 85 subjects. The results of these generalization evaluation experiments are summarized in Fig. 12 and Table X. Comparing the results in Tables X and V, it is evident that the majority of the performance measures corresponding to "stepwise LDA of Dist_Curv" and "Stepwise LDA of 3-D Gabor" face recognizers remain level. On the other hand, the performances of the "Stepwise LDA of 2-D Gabor" and eventual multimodal "2-D Gabor + 3-D Gabor +

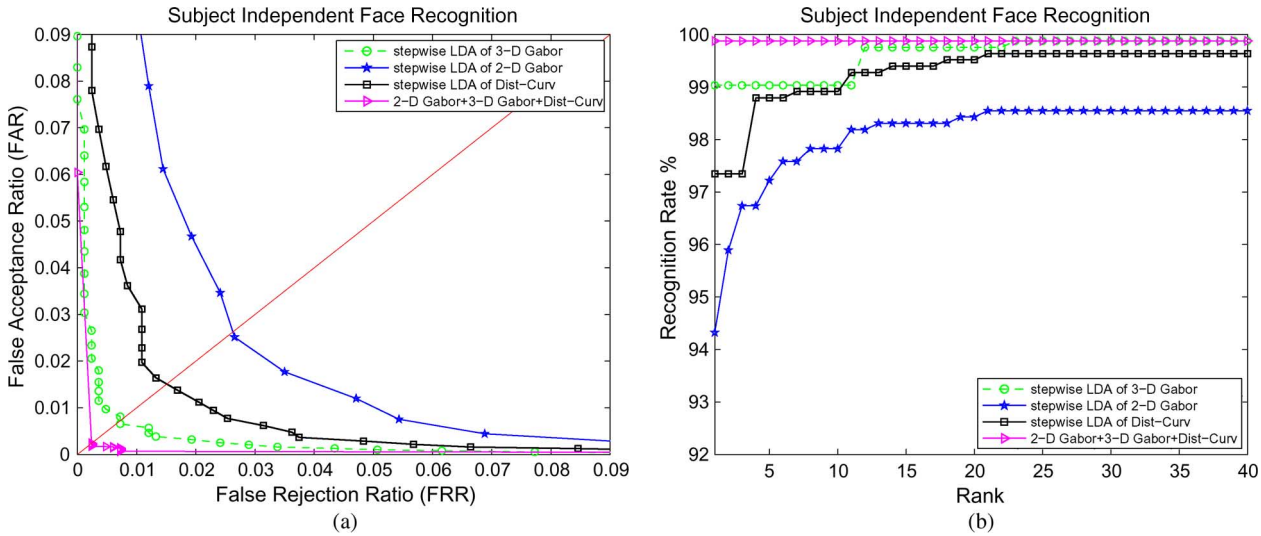


Fig. 12. Performance evaluation of the proposed unimodal and multimodal face classifiers evaluated using training-subject independent test set. (a) ROC curves showing the verification accuracies. (b) CMC curves showing the identification performance.

TABLE IX
EER, AUC, AND RR1 VALUES FOR THE PROPOSED UNIMODAL AND MULTIMODAL FACE CLASSIFIERS IN RECOGNIZING "EXPRESSIVE" FACES

Algorithm	EER (%)	AUC	RR1 (%)
Stepwise LDA of 3-D Gabor	1.09	0.0006	97.82
Stepwise LDA of 2-D Gabor	3.26	0.0055	93.75
Stepwise LDA of Dist_Curv	2.53	0.0022	96.19
2-D Gabor+3-D Gabor+Dist_Curv	0.59	0.0003	98.70

TABLE X
EER, AUC, AND RR1 VALUES FOR THE PROPOSED UNIMODAL AND MULTIMODAL FACE CLASSIFIERS EVALUATED USING TRAINING-SUBJECT INDEPENDENT TEST SET

Algorithm	EER (%)	AUC	RR1 (%)
Stepwise LDA of 3-D Gabor	0.72	0.0003	99.03
Stepwise LDA of 2-D Gabor	2.63	0.0040	94.32
Stepwise LDA of Dist_Curv	1.51	0.0012	97.34
2-D Gabor+3-D Gabor+Dist_Curv	0.26	0.0002	99.67

TABLE XI
EER, AUC, AND RR1 VALUES FOR THE PROPOSED UNIMODAL AND MULTIMODAL FACE CLASSIFIERS EVALUATED USING "EXPRESSIVE" TRAINING SUBJECT INDEPENDENT TEST SET

Algorithm	EER (%)	AUC	RR1 (%)
Stepwise LDA of 3-D Gabor	0.98	0.0004	98.18
Stepwise LDA of 2-D Gabor	3.47	0.0060	92.39
Stepwise LDA of Dist_Curv	2.73	0.0033	97.01
2-D Gabor+3-D Gabor+Dist_Curv	0.61	0.0001	98.75

Dist_Curv" face recognizers show only a slight decline. These results confirm that the proposed training scheme is not biased towards subjects present in the training set.

In order to measure the generalization capabilities of the proposed algorithm for recognizing expressive faces, an expressive testing set was created that does not contain image pairs from those 18 subjects involved in the training set. This testing set contains 48 expressive facial image pairs from 38 subjects. The results of these expressive face recognition experiments are summarized in Fig. 13 and Table XI. Comparing the results in Tables IX and XI reveals that the "Stepwise LDA of 3-D Gabor" performed slightly better on training-subject independent data while the performance of other unimodal and the eventual multimodal face recognizer show only a slight decline in performance. The observed performance is better than other state of the art algorithms tested on T3FRD (on training-subject dependent data). This confirms that our face recognition model can generalize to subjects that it has not been trained on.

V. CONCLUSION

We have described a novel face recognition model that utilizes local features calculated from coregistered portrait and range image pairs. The model embodies several unique contributions to the fields of multimodal face recognition and automatic landmark detection. Since portrait Gabor, range Gabor, and geodesic/Euclidean distances used in the proposed multimodal face recognition model are local features which are calculated around or between pairs of facial landmarks, one important contribution is a fast and accurate landmarking scheme incorporating both 2-D and 3-D Gabor clues. In particular, the 2-D fiducial detection method proposed in EBGM [17] is extended to independently detect fiducial points on portrait and range representations of the face. The fiducial detection throughput is enhanced by restricting the search range corresponding to each target fiducial, thereby removing the computationally expensive iterative scheme present in the original EBGM. We conducted a detailed study to quantitatively measure and compare the performance of the unimodal fiducial detection schemes in pinpointing 11 fiducial points.

A second important contribution is a unique method to combine portrait and range Gabor clues to boost the landmark detection accuracies whenever coregistered range and portrait pairs are available. To the best of our knowledge this is the first time that complementary range and portrait Gabor-based appearance clues have been used simultaneously to detect fiducial points on

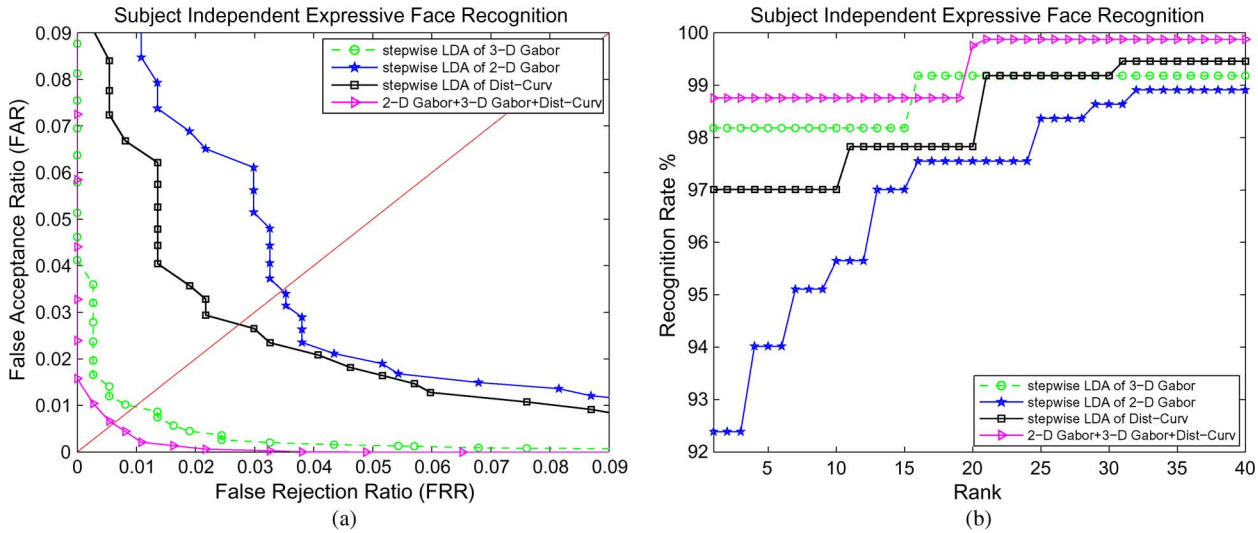


Fig. 13. Performance evaluation of the proposed unimodal and multimodal face classifiers evaluated using "expressive" training-subject independent test set. (a) ROC curves showing the verification accuracies. (b) CMC curves showing the identification performance.

pairs of coregistered range and portrait images. The higher accuracies achieved by the proposed "hybrid fiducial detection" algorithm improve the overall performance of the proposed multimodal face recognition model. Performance evaluations discussed in Section IV-A (Fig. 10 and Table VIII) show that by using "hybrid fiducial detection" instead of its less accurate unimodal counterparts, the observed equal error rate is significantly reduced from EER $= 0.57\%$ to EER $= 0.25\%$. This accurate fiducial detection algorithm can be very useful in many related applications such as human–machine interaction and face expression recognition.

Another unique contribution of this work is the use of statistical feature analysis techniques in a hierarchical scheme designed to combine features from diverse sources into a unified multimodal face recognition algorithm. One challenge faced while creating this multimodal face recognition model was that concatenated Gabor coefficients and Euclidean/geodesic anthropometric features were of high dimensionality, causing "small sample size" problems when applying LDA. A small sample size problem happens when the total number of training samples is less than the dimensionality of the features. Instead of blindly reducing feature vector dimensionality by applying PCA prior to LDA [56], which is a common trend in the literature, we chose stepwise-LDA [54] to statistically analyze the input features, selecting those features important for classification and removing irrelevant ones. The evaluations indicate that the three parallel unimodal face classifiers deliver performances significantly better than corresponding baseline algorithms. Additional robustness evaluations indicate that the statistically selected features incorporated in these unimodal face recognizers are robust against facial expression variations and fiducial detection errors.

Another contribution of this model is the final match score-level fusion stage of the hierarchy, whereby three parallel face recognizers are combined into a unified multimodal (2-D+3-D) face recognition algorithm that achieves significantly better performance than each of the participating classifiers. The implemented match score-level fusion makes decisions based on the weighted sum of the contributing face recognizer where weights are proportional to individual performances. Hence, the "stepwise LDA of 3-D Gabor" has the highest contribution followed by the "stepwise LDA of Dist_Curv" and, finally, the "stepwise LDA of 2-D Gabor." Our novel multimodal classifier, "2-D Gabor + 3-D Gabor + Dist_Curv," using features calculated from fiducials automatically detected by the "hybrid detection" algorithm achieves excellent performance with EER $= 0.25\%$ and AUC $= 0.0001$ and rank 1 recognition rate of RR1 $= 99.76\%$. This is highly competitive with state-of-the-art algorithms and baseline algorithms evaluated on the T3FRD. Additional evaluations confirm the robustness of the proposed multimodal face recognition algorithm against facial expression variations and fiducial detection errors.

ACKNOWLEDGMENT

The authors would like to thank Dr. K. Castleman for providing them with the Texas 3-D Face Recognition Database (T3FRD) of facial range and portrait images. This work draws continued inspiration from Dr. Castleman's ideas.

REFERENCES

[1] W. Bledsoe, Man-Machine Facial Recognition Panoramic Research Inc., Palo Alto, CA, 1966.
[2] W. Zhao, R. Chellappa, P. Phillips, and A. Rosenfeld, "Face recognition: A literature survey," *ACM Comput. Surv.*, vol. 35, no. 4, pp. 399–459, 2003.
[3] P. Phillips, P. Grother, R. Micheals, D. Blackburn, E. Tabassi, and J. Bone, Face Recognition Vendor Test 2002: Overview and Summary National Institute of Standards and Technology, NISTIR 6965, Tech. Rep., 2003 [Online]. Available: www.frvt.org
[4] K. Bowyer, K. Chang, and P. Flynn, "A survey of approaches and challenges in 3D and multi-modal 3D+2D face recognition," *Comput. Vis. Image Understand.*, vol. 101, no. 1, pp. 1–15, 2006.
[5] S. Gupta, M. K. Markey, and A. C. Bovik, "Advances and challenges in 3D and 2D+3D human face recognition," *Pattern Recognit. Biology*, p. 63, 2007.
[6] B. Gokberk, A. A. Salah, and L. Akarun, "Rank-based decision fusion for 3D shape-based face recognition," *Lecture Notes Comput. Sci.*, vol. 3546, p. 1019, 2005.
[7] T. Heseltine, N. Pears, and J. Austin, "Three-dimensional face recognition: A fishersurface approach," *Image Anal. Recognit.*, pp. 684–691, 2004.

[8] C. Queirolo, L. Silva, O. Bellon, and M. Segundo, "3d face recognition using the surface interpenetration measure: A comparative evaluation on the frgc database," in *Proc. 19th Int. Conf. Pattern Recognition, 2008 (ICPR 2008)*, Dec. 2008, pp. 1–5.

[9] T. Faltemier, K. Bowyer, and P. Flynn, "A region ensemble for 3-D face recognition," *IEEE Trans. Inf. Forensics Security*, vol. 3, no. 1, pp. 62–73, Mar. 2008.

[10] I. Kakadiaris, G. Passalis, G. Toderici, M. Murtuza, Y. Lu, N. Karampatziakis, and T. Theoharis, "Three-dimensional face recognition in the presence of facial expressions: An annotated deformable model approach," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 4, pp. 640–649, Apr. 2007.

[11] R. McKeon and T. Russ, "Employing region ensembles in a statistical learning framework for robust 3D facial recognition," in *Proc. 2010 Fourth IEEE Int. Conf. Biometrics: Theory Applications and Systems (BTAS)*, pp. 1–7.

[12] C. Boehnen, T. Peters, and P. Flynn, "3D signatures for fast 3D face recognition," *Adv. Biometrics*, vol. 5558, pp. 12–21, 2009.

[13] K. Chang, K. Bowyer, and P. Flynn, "Face recognition using 2D and 3D facial data," in *Proc. ACM Workshop on Multimodal User Authentication*, 2003, pp. 25–32.

[14] A. Godil, S. Ressler, and P. Grother, "Face recognition using 3D facial shape and color map information: Comparison and combination," *SPIE, Biometric Technology for Human Identification*, vol. 5404, pp. 351–361, 2005.

[15] X. Lu, A. K. Jain, and D. Colbry, "Matching 2.5 D face scans to 3D models," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 1, pp. 31–43, Jan. 2006.

[16] X. Lu and A. K. Jain, "Integrating range and texture information for 3D face recognition," in *Proc. Seventh IEEE Workshops on Application of Computer Vision (WACV/MOTION'05)*, Washington, DC, 2005, vol. 1, pp. 156–163, IEEE Computer Society.

[17] L. Wiskott, J.-M. Fellous, N. Kuiger, and C. von der Malsburg, "Face recognition by elastic bunch graph matching," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 19, no. 7, pp. 775–779, Jul. 1997.

[18] A. M. Bronstein, M. M. Bronstein, and R. Kimmel, "Three-dimensional face recognition," *Int. J. Comput. Vis.*, vol. 64, no. 1, pp. 5–30, 20050801.

[19] A. Jain and S. Li, *Handbook of Face Recognition*. New York: Springer, 2005.

[20] Y. Wang, C. S. Chua, and Y. K. Ho, "Facial feature detection and face recognition from 2D and 3D images," *Pattern Recognit. Lett.*, vol. 23, no. 10, pp. 1191–1202, 2002.

[21] C. Xu, S. Li, T. Tan, and L. Quan, "Automatic 3D face recognition from depth and intensity Gabor features," *Pattern Recognit.*, vol. 42, no. 9, pp. 1895–1905, 2009.

[22] Y. Freund and R. Schapire, "A desicion-theoretic generalization of on-line learning and an application to boosting," in *Computational Learning Theory*. New York: Springer, 1995, pp. 23–37.

[23] J. Cook, C. McCool, V. Chandran, and S. Sridharan, "Combined 2D/3D face recognition using log-gabor templates," in *Proc. IEEE Int. Conf. Video and Signal Based Surveillance*, 2006, p. 83, IEEE Computer Society.

[24] Y. Wang and C. Chua, "Face recognition from 2D and 3D images using 3D Gabor filters," *Image Vis. Comput.*, vol. 23, no. 11, pp. 1018–1028, 2005.

[25] D. Sim, O. Kwon, and R. Park, "Object matching algorithms using robust Hausdorff distance measures," *IEEE Trans. Image Process.*, vol. 8, no. 3, pp. 425–429, Mar. 1999.

[26] M. Husken, M. Brauckmann, S. Gehlen, and C. Von der Malsburg, "Strategies and benefits of fusion of 2D and 3D face recognition," in *Proc. 2005 IEEE Computer Society Conf. Computer Vision and Pattern Recognition*, 2005, pp. 174–174.

[27] S. Jahanbin, H. Choi, R. Jahanbin, and A. Bovik, "Automated facial feature detection and face recognition using Gabor features on range and portrait images," in *Proc. 15th IEEE Int. Conf. Image Processing, 2008 (ICIP 2008)*, 2008, pp. 2768–2771.

[28] W. Zhao, R. Chellappa, and P. Phillips, Subspace Linear Discriminant Analysis for Face Recognition Center for Automation Research, University of Maryland, College Park, Tech. Rep. CAR-TR-914, 1999, vol. 137, pp. 138–144.

[29] S. Gupta, M. Markey, and A. Bovik, "Anthropometric 3d face recognition," *Int. J. Comput. Vis.*, vol. 90, no. 3, pp. 331–349, 2010.

[30] S. Jahanbin, A. C. Bovik, and H. Choi, "Automated facial feature detection from portrait and range images," in *Proc. IEEE Southwest Symp. Image Analysis and Interpretation, 2008 (SSIAI 2008)*, 2008, pp. 25–28.

[31] D. Riccio and J. L. Dugelay, "Geometric invariants for 2D/3D face recognition," *Pattern Recognit. Lett.*, vol. 28, no. 14, pp. 1907–1914, 2007.

[32] T. Cootes, G. Edwards, and C. Taylor, "Active appearance models," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 23, no. 6, pp. 681–685, Jun. 2001.

[33] T. Cristinacce and D. Cootes, "A comparison of shape constrained facial feature detectors," in *Proc. Sixth IEEE Int. Conf. Automatic Face and Gesture Recognition, 2004*, May 17–19, 2004, pp. 375–380.

[34] S. Jahanbin, H. Choi, A. Bovik, and K. Castleman, "Three dimensional face recognition using wavelet decomposition of range images," in *Proc. IEEE Int. Conf. Image Processing, 2007 (ICIP 2007)*, San Antonio, TX, 2007, vol. 1.

[35] Y. Lee and T. Yi, "3D face recognition using multiple features for local depth information," in *Proc. 4th EURASIP Conf. Focused on Video/Image Processing and Multimedia Communications, 2003*, Jul. 2003, vol. 1, no. 2–5, pp. 429–434.

[36] S. Gupta, M. K. Markey, J. K. Aggarwal, and A. C. Bovik, *Three Dimensional Face Recognition Based on Geodesic and Euclidean Distances*, L. J. Latecki, D. M. Mount, and A. Y. Wu, Eds. Bellingham, WA: SPIE, 2007, vol. 6499, p. 64990D, no. 1.

[37] M. Koudelka, M. Koch, and T. Russ, "A prescreener for 3D face recognition using radial symmetry and the hausdorff fraction," in *Proc. IEEE Computer Society Conf. Computer Vision and Pattern Recognition-Workshops, 2005 (CVPR Workshops)*, 2005, p. 168.

[38] G. Gordon, "Face recognition based on depth and curvature features," in *Proc. 1992 IEEE Computer Society Conf. Computer Vision and Pattern Recognition*, Reading, MA, pp. 808–810, TASC, 19920101, Cat. No. 92CH3168-2.

[39] A. Moreno, Á. Sánchez, and J. Vélez, "FD íaz. Face recognition using 3d surface-extracted descriptor," in *Proc. Irish Machine Vision and Image Processing Coleraine*, Ireland, Sep. 2003.

[40] A. Colombo, C. Cusano, and R. Schettini, "3D face detection using curvature analysis," *Pattern Recognit.*, vol. 39, no. 3, pp. 444–455, 2006.

[41] K. Chang, W. Bowyer, and P. Flynn, "Multiple nose region matching for 3D face recognition under varying facial expression," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 28, no. 10, pp. 1695–1700, Oct. 2006.

[42] T. Faltemier, K. Bowyer, and P. Flynn, "Rotated profile signatures for robust 3d feature detection," in *Proc. 8th IEEE Int. Conf. Automatic Face & Gesture Recognition, 2008 (FG'08)*, 2009, pp. 1–7.

[43] C. Boehnen and T. Russ, "A fast multi-modal approach to facial feature detection," in *Proc. Seventh IEEE Workshop Application Computer Vision (WACV/MOTIONS '05)*, Jan. 2005, vol. 1, pp. 135–142.

[44] S. Gupta, K. C. Castleman, M. K. Markey, and A. C. Bovik, "Texas 3D face recognition database," in *Proc. IEEE Southwest Symp. Image Analysis and Interpretation, 2010 (SSIAI 2010)*, 2010, pp. 25–28.

[45] J. G. Daugman, "Uncertainty relation for resolution in space, spatial frequency, and orientation optimized by two-dimensional visual cortical filters," *J. Opt. Soc. Amer. A* vol. 2, no. 7, pp. 1160–1169, 1985 [Online]. Available: http://josaa.osa.org/abstract.cfm?URI=josaa-2-7-1160

[46] M. Clark, A. Bovik, and W. Geisler, "Texture segmentation using Gabor modulation/demodulation," *Pattern Recognit. Lett.*, vol. 6, no. 4, pp. 261–267, 1987.

[47] A. Bovik, M. Clark, and W. Geisler, "Multichannel texture analysis using localized spatial filters," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 12, no. 1, pp. 55–73, Jan. 1990.

[48] J. Havlicek, D. Harding, and A. Bovik, "Multidimensional quasi-eigenfunction approximations and multicomponent am-fm models," *IEEE Trans. Image Process.*, vol. 9, no. 2, pp. 227–242, Feb. 2000.

[49] P. J. Besl and H. D. McKay, "A method for registration of 3-D shapes," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 14, no. 2, pp. 239–256, Feb. 1992.

[50] 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 Computer Society Conf. Computer Vision and Pattern Recognition, 2005 (CVPR 2005)*, 2005, vol. 1, pp. 947–954.

[51] 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 *Proc. Computer Vision and Pattern Recognition Workshop*, 2005, p. 154.

[52] X. Lu and A. K. Jain., Multimodal Facial Feature Extraction for Automatic 3D Face Recognition Department of Computer Science, Michigan State University, East Lansing, MI, Tech. Rep. MSU-CSE-05-22, Aug. 2005.

[53] J. Sethian, "Evolving interfaces in computational geometry, fluid mechanics, computer vision, and materials science," in *Level Set Methods and Fast Marching Methods*. Cambridge, U.K.: Cambridge Univ. Press, 1999.
[54] S. Sharma, *Applied Multivariate Techniques*. Hoboken, NJ: Wiley, 1995.
[55] R. Duda, P. Hart, and D. Stork, *Pattern Classification*. Hoboken, NJ: Wiley, 2001.
[56] P. Belhumeur, J. Hespanha, and D. Kriegman, "Eigenfaces vs. fisherfaces: Recognition using class specific linear projection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 19, no. 7, pp. 711–720, Jul. 1997.
[57] L. Chen, H. Liao, M. Ko, J. Lin, and G. Yu, "A new LDA-based face recognition system which can solve the small sample size problem," *Pattern Recognit.*, vol. 33, no. 10, pp. 1713–1726, 2000.
[58] H. Yu and J. Yang, "A direct LDA algorithm for high-dimensional datawith application to face recognition," *Pattern Recognit.*, vol. 34, no. 10, pp. 2067–2070, 2001.
[59] P. Grother, R. Micheals, and P. Phillips, "Face recognition vendor test 2002 performance metrics," *Audio-and Video-Based Biometric Person Authentication*, pp. 937–945, 2003.
[60] A. A. Ross, K. Nandakumar, and A. K. Jain, *Handbook of Multibiometrics*. New York: Springer-Verlag, 2006.
[61] S. C. Le Zou, Z. Xiong, M. Lu, and K. R. Castleman, "3-D face recognition based on warped example faces," *IEEE Trans. Inf. Forensics Security*, vol. 2, no. 3, pp. 513–528, Sep. 2007.
[62] K. I. Chang, K. W. Bowyer, and P. J. Flynn, "An evaluation of multimodal 2D+3D face biometrics," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 4, pp. 619–624, Apr. 2005.

Sina Jahanbin (S'05–M'08) received the B.S. degree in electrical engineering from Sharif University of Technology, Tehran, Iran, in 2002, the M.S. degree in electrical engineering from the Pennsylvania State University, in 2004, and the Ph.D. degree in electrical engineering from the University of Texas at Austin, in May 2011.

In 2005, he joined the Laboratory for Image and Video Engineering (LIVE) at the University of Texas at Austin developing visual inspection and multimodal face recognition algorithms. Currently, he is an algorithm engineer at KLA-Tencor, San Jose, CA, developing machine vision algorithms for silicon wafer inspection systems.

Hyohoon Choi (S'02–M'06) received the B.S. degree in electrical engineering from Kangwon National University, Korea, in 1999, and the M.S. and Ph.D. degrees in biomedical engineering from the University of Texas at Austin, in 2001 and 2006, respectively.

He was with the Laboratory for Image and Video Engineering (LIVE), Department of Electrical and Computer Engineering, The University of Texas at Austin, from 2002 to 2006. He has worked at ADIR from 2001 to 2006 developing image processing algorithms for multiple NIH and NIST funded research projects. From 2006 to 2008, he worked at Sealed Air Corporation as a computer vision researcher in San Jose, CA. Currently, he is a senior researcher at Samsung Electro-Mechanics, Korea, developing machine vision algorithms for PCB inspection systems.

Alan C. Bovik (S'80–M'80–F'96) is the Curry/Cullen Trust Endowed Chair Professor at The University of Texas at Austin, where he directs research in the Laboratory for Image and Video Engineering (LIVE). He is well known for his fundamental work in perceptual image and video processing and computational modeling of visual perception. He is also noted for innovations in engineering education, including his popular books and the widely used Signal, Image and Video Audiovisual (SIVA) Demonstration Gallery. He has also been a major contributor to engineering service, including innovating and creating the *IEEE International Conference on Image Processing*, first held in Austin, TX, in November, 1994, and playing an instrumental role in proposing and creating the IEEE TRANSACTIONS ON IMAGE PROCESSING, for which he served as Editor-in-Chief for six years.

Dr. Bovik has received a number of major awards from the IEEE Signal Processing Society, including: The Best Paper Award (2009); the Education Award (2007); the Technical Achievement Award (2005), the Distinguished Lecturer Award (2000); and the Meritorious Service Award (1998). In 2011, he was named "Imaging Scientist of the Year" by IS&T and SPIE. He is a Fellow of the Optical Society of America (OSA) and a Fellow of the Society of Photo-Optical and Instrumentation Engineers (SPIE).