Three-Dimensional Face Recognition in the Presence of Facial Expressions: An Annotated Deformable Model Approach

Ioannis A. Kakadiaris, *Member*, *IEEE*, Georgios Passalis, George Toderici, Mohammed N. Murtuza, Yunliang Lu, Nikos Karampatziakis, and Theoharis Theoharis

Abstract—In this paper, we present the computational tools and a hardware prototype for 3D face recognition. Full automation is provided through the use of advanced multistage alignment algorithms, resilience to facial expressions by employing a deformable model framework, and invariance to 3D capture devices through suitable preprocessing steps. In addition, scalability in both time and space is achieved by converting 3D facial scans into compact metadata. We present our results on the largest known, and now publicly available, Face Recognition Grand Challenge 3D facial database consisting of several thousand scans. To the best of our knowledge, this is the highest performance reported on the FRGC v2 database for the 3D modality.

Index Terms—Face and gesture recognition, information search and retrieval.

1 Introduction

Among several biometric identification modalities proposed for verification and identification purposes, face recognition is high in the list of subject preference, mainly because of its nonintrusive nature. However, from the operator point of view, face recognition has some significant challenges that hamper its widespread adoption. Accuracy is the most important of these challenges. Current 2D face recognition systems can be fooled by differences in pose, lighting, expressions, and other characteristics that can vary between captures of a human face. This issue becomes more significant when the subject has incentives *not* to be recognized (noncooperative subjects).

It is now widely accepted that, in order to address the challenge of accuracy, different capture modalities (such as 3D or infrared) and/or multiple instances of subjects (in the form of multiple still captures or video) must be employed [1]. However, the introduction of new capture modalities brings new challenges for a field-deployable system. The challenges of 3D face recognition, which concern the current paper, are:

- 3D Challenge 1—*Accuracy Gain:* A significant accuracy gain of the 3D system with respect to 2D face recognition systems must result in order to justify the introduction of a 3D system, either for sole use or in combination with other modalities.
- 3D Challenge 2—*Efficiency:* 3D capture creates larger data files per subject which implies significant storage requirements and slow processing. The conversion of raw 3D data to efficient metadata must thus be addressed.
- 3D Challenge 3—*Automation:* A field-deployable system must be able to function fully automatically. It is therefore not acceptable to assume user intervention such as for the location of key landmarks in a 3D facial scan.
- 3D Challenge 4—*Capture Devices:* 3D capture devices were mostly developed for medical and other low-volume applications and suffer from a number of drawbacks when applied to face recognition, including artifacts, small depth of field, long acquisition time, multiple types of output, and high price.
- 3D Challenge 5—*Testing Databases:* The lack of large and widely accepted databases for objectively testing the performance of 3D face recognition systems.

1.1 Related Work

Despite the introduction of commercial grade 2D face recognition systems, 2D face recognition remains unreliable. Extensive experiments conducted using the FERET data set [2] and during the FRVT 2002 study indicate that the success rate is not sufficient for critical applications. It appears that 2D face recognition techniques have exhausted their potential as they stumble on inherent problems of their modality (mainly pose and illumination differences).

With the shortcomings of the 2D approaches, a number of 3D and 3D + 2D multimodal approaches have recently been proposed. Excellent recent surveys on this field are given by

- *I.A. Kakadiaris is with the Computational Biomedicine Lab (CBL), Department of Computer Science, University of Houston, MS CSC 3010, 219 Philip Guthrie Hoffman Hall (PGH), 4800 Calhoun, Houston, TX 77204-3010. E-mail: ikakadia@central.uh.edu.*
- *G. Passalis and T. Theoharis are with CBL and the Department of Informatics, University of Athens, TYPA Buildings, Panepistimiopolis, Ilisia, 15784, Athens, Greece. E-mail: passalis@yahoo.com, theotheo@di.uoa.gr.*
- *G. Toderici can be reached at E-mail: toderici@cs.uh.edu.*
- *M.N. Murtuza can be reached at E-mail: najamm@gmail.com.*
- *Y. Lu can be reached at E-mail: yunliang.lu@gmail.com.*
- *N. Karampatziakis is currently with the Department of Computer Science, Cornell University, 5132 Upson Hall, Ithaca NY 14853. E-mail: just.nikos@gmail.com.*

Manuscript received 1 Feb. 2006; revised 30 June 2006; accepted 7 July 2006; published online 18 Jan. 2007.
Recommended for acceptance by S. Prabhakar, J. Kittler, D. Maltoni, L. O'Gorman, and T. Tan.
For information on obtaining reprints of this article, please send e-mail to: tpami@computer.org and reference IEEECS Log Number TPAMISI-0100-0206.
Digital Object Identifier no. 10.1109/TPAMI.2007.1017.

Bowyer et al. [1] and Chang et al. [3]. Due to the lack of available 3D databases, the majority of these approaches has not been extensively tested. To address this issue, NIST introduced the Face Recognition Grand Challenge (FRGC) and Face Recognition Vendor Test 2006 [4], [5], and made two multimodal databases publicly available. The first, FRGC v1, includes more than 900 scans, while the second, FRGC v2, includes more than 4,000 scans with facial expressions. Below, we present a small sample of related work that is not meant to be exhaustive, and it is focused on the approaches that utilize these databases, as we expect that the FRGC databases will become the standard score reporting databases in Face Recognition; this will aid both scientific research and potential users of such systems. The performance metrics used to evaluate these approaches are described in Section 3.2.

On the facial-expression-free FRGC v1 database, Pan et al. [6] reported a 95 percent rank-one recognition rate using a PCA approach, while Russ et al. [7] reported a 98 percent verification rate. Our deformable model approach achieved a 99 percent rank-one recognition rate for the same database [8].

On the extensive FRGC v2 database, Chang et al. [9], [10] examined the effects of facial expressions using two different 3D recognition algorithms. They reported a 92 percent rank-one recognition rate. Lu and Jain [11] use a generic 3D model to create user-specific deformable models in the neutral position. In the identification phase, the distance returned by the Iterated Closest Point (ICP) algorithm was used for matching all the user-specific models to the new data set, with 92.1 percent rank-one identification on a subset of FRGC v2. Russ et al. [12] use Principal Components Analysis (PCA) on range images generated after realigning the data to a generic 3D model. The results were presented on subsets of FRGC v2, and show that this improvement outperforms the pure PCA approach, but still suffers from facial expressions. Lin et al. [13] compute summation invariant images from the raw 3D data. Using the baseline PCA approach included in the FRGC v2, but on manually cropped images, using the provided annotation, they reported verification rates between 80.82 percent and 83.13 percent. Husken et al. [14] presented a multimodal approach that uses hierarchical graph matching (HGM). They extended their HGM approach from 2D to 3D but the reported 3D performance is lower than the 2D equivalent. Their fusion, however, offers competitive results, a 96.8 percent verification rate at 0.001 False Acceptance Rate (FAR), compared to 86.9 percent for the 3D only. Maurer et al. [15] presented a multimodal approach tested on the FRGC v2 database and reported a 87 percent verification rate at 0.01 FAR. In our previous work on this database [16], we analyzed the behavior of our approach in the presence of facial expressions. The improvements presented in this paper allowed us to overcome previous shortcomings, as detailed in Section 3, and we now claim the top reported performance in 3D face recognition when tested using the FRGC databases.

1.2 Overview

In this paper, we address the major challenges of a 3D field-deployable face recognition system. We have developed a *fully automatic* system which uses a composite alignment algorithm to register 3D facial scans with a 3D facial model, thus achieving complete *pose-invariance*. Our system employs a deformable model framework to fit the 3D facial model to the aligned 3D facial scans, and in so doing measures the difference between the facial scan and the model in a way that achieves a high degree of *expression invariance* and thus *high accuracy*. The 3D differences (the deformed facial model) are converted to a 2D geometry image and then transformed to the wavelet domain; it has been observed that a small portion of the wavelet data is sufficient to accurately describe a 3D facial scan, thus achieving the *efficiency* goal. Certain issues of 3D *capture devices* were addressed; specifically, artifacts and multiple types of output (range scans and polygon meshes) were handled by suitable preprocessing. Median cut and smoothing filters were applied to handle artifacts and the conversion to a common format (3D polygon data), with down-sampling where necessary, was used to address the issues of multiple types of output from different 3D capture devices. Concerning objective *testing databases*, the FRGC database was augmented by our own 3D face capture project, resulting in a total of almost 5,000 3D facial scans.

The proposed integrated system is based on our previous work in face recognition [8], [16], [17], but new additions to our approach (e.g., normal maps and composite alignment algorithm) along with improvements on the existing methods, result in a significant performance gain. Additionally, multiple-sensor databases are used to the best of our knowledge for the first time to evaluate the performance of such a system. Finally, a prototype 3D face recognition system has been built and it is operational at the University of Houston.

The rest of this paper is organized as follows: Section 2 describes the methods utilized by our approach as well as the specifications and challenges of the prototype system. Section 3 presents a performance evaluation using extensive and publicly available databases, while Section 4 summarizes our approach and proposes future directions.

2 An Integrated 3D Face Recognition System

The main idea of our approach is to describe facial data using a deformed facial model. The deformed model captures the details of an individual's face and represents this 3D geometry information in an efficient 2D structure by utilizing the model's UV parameterization. This structure is analyzed in the wavelet domain and the spectral coefficients define the final metadata that are used for comparison among different subjects. The geometric modeling of the human face allows greater flexibility, better understanding of the face recognition issues, and requires no training.

Our face recognition procedure can be divided in two phases, enrollment and authentication:

Enrollment. Raw data are converted to metadata and stored in the database (Fig. 1) as follows:

1. *Acquisition (Section 2.1.1):* Raw data are acquired from the sensor and converted to a 3D polygonal representation using sensor-dependent preprocessing.
2. *Alignment (Section 2.1.3):* The data are aligned into a unified coordinate system using a scheme that combines three different alignment algorithms.
3. *Deformable Model Fitting (Section 2.1.4):* An Annotated Face Model is fitted to the data.
4. *Geometry Image Analysis (Section 2.1.5):* Geometry and normal map images are derived from the fitted model and wavelet analysis is applied to extract a reduced set of coefficients as metadata.

Authentication. Metadata retrieved from the database are directly compared using a distance metric (Section 2.1.6).

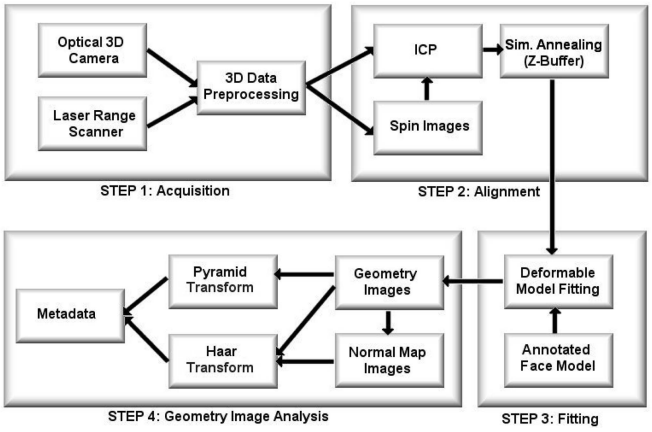


Fig. 1. Enrollment phase of the proposed integrated 3D face recognition system.

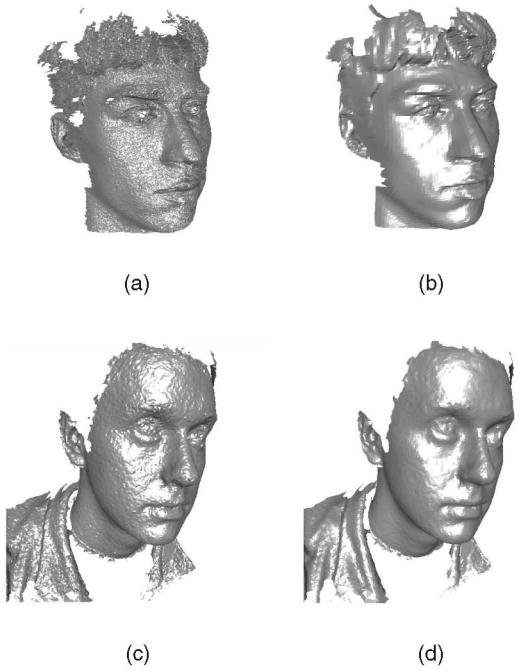


Fig. 2. Sensor-dependent preprocessing. Laser range scanner: (a) raw data (212 K triangles) and (b) processed data (13 K). Stereo camera: (c) raw data (66 K) and (d) processed data (33 K).

We first describe in detail the methods used and subsequently present our field deployable prototype system.

2.1 Methods

2.1.1 Data Preprocessing

The preprocessing's purpose is twofold: to eliminate sensor-specific problems and to unify data from different sensors. The preprocessing consists of the following filters, executed in the given order:

- *Median Cut:* This filter is applied to remove spikes from the data. Spikes are more common in laser range scanners, therefore stronger filtering is needed in this case.
- *Hole Filling:* Since laser scanners usually produce holes in certain areas (e.g., eyes and eyebrows) a hole filling procedure is applied.

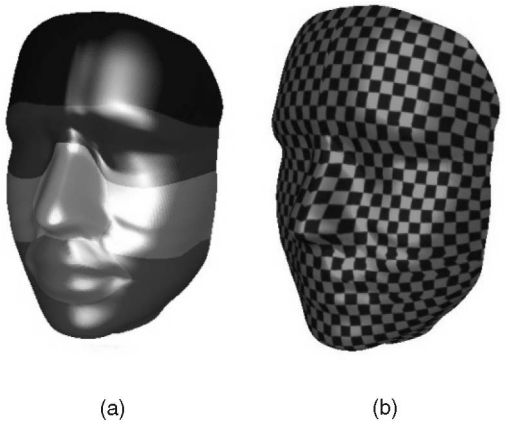


Fig. 3. AFM: (a) Annotated facial areas and (b) texture used to demonstrate parameterization.

- *Smoothing:* A smoothing filter is applied to remove white noise as most high resolution scanners produce noisy data in real-life conditions.
- *Subsampling:* The deformable model fitting (Section 2.1.4) effectively resamples the data, making the method insensitive to data resolution. Therefore, the resolution is decreased to gain efficiency to a level that does not sacrifice accuracy.

In general, the current generation of scanners output either a range image or 3D polygonal data. We implemented the above filters for both representations (Fig. 2). The filters operate on a 1-neighborhood area for both representations. Note that with range images, there is the possibility to first apply the filters and then convert to 3D polygonal data or the opposite. Experiments show that the filters perform better in the data's native representation.

2.1.2 Annotated Face Model

Our approach utilizes an annotated model of the human face (AFM), which needs to be constructed only once and is described in detail in our previous work [8], [16]. This model is subsequently used in alignment and it is deformed in the fitting stage and is the source of the metadata. Based on Farkas' work [18], we ensured that the model is anthropometrically correct and it was annotated into different facial areas, as depicted in Fig. 3a. The key feature of this model is its continuous global UV parameterization, depicted in Fig. 3b. The injective property of the specific parameterization allows us to map all vertices of the model's surface from R^3 to R^2 and vice versa. This allows the transition from the original polygonal representation to a regularly sampled 2D grid representation, called *geometry image* [19], [20], [21].

2.1.3 Alignment

Our previous work on face recognition points out that alignment (pose correction) is a key part of any geometric approach. In fact, an alignment error cannot be rectified in later steps of this or other similar approaches. To this end, we present a novel multistage alignment method that offers robust and accurate alignment even in the presence of facial expressions.

The general idea is that we align each new raw data set with the AFM before the fitting process starts. The alignment computes a rigid transformation that includes rotation and

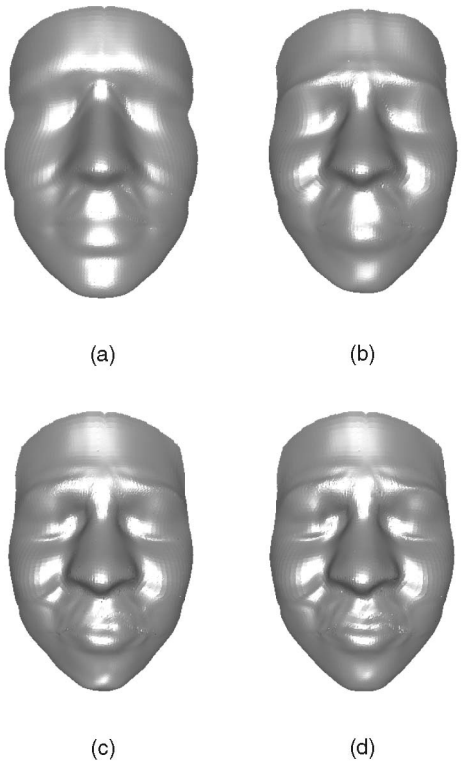


Fig. 4. Fitting progress: AFM after (a) 0, (b) 8, (c) 32, and (d) 64 iterations.

translation. The multistage alignment method consists of three algorithmic steps. Each step uses as input the output of the previous one; early steps offer greater resilience to local minimums while later steps offer greater alignment accuracy:

- *Spin Images:* The purpose of the first step is to establish a plausible initial correspondence between the model and the data. If we do not expect arbitrary rotations and translations in the database, this step can be omitted. We utilize the spin image algorithm presented by Johnson [22]. A spin image is a representation of the geometric neighborhood around a specific point. To register two shapes, the correspondences between the individual spin images must be found. These correspondences are grouped into geometrically consistent groups and the transformations they yield are verified by checking if they rotate the data by an acute angle (based on the assumption that a given face does not have an upside down pose nor does it have an opposite orientation from the camera). This check is essential due to the bilateral symmetry property of the human face.
- *Iterative Closest Point (ICP):* The main step of our alignment pipeline uses the ICP algorithm [23], extended in a number of ways. The ICP algorithm solves the registration problem by minimizing the distance between the two sets of points. Pairs containing points on surface boundaries are rejected [24]. This ensures that no residual error is introduced into ICP's metric from the nonoverlapping parts of two surfaces. Finally, if the resulting transformation is not satisfactory, we have an option of running the trimmed ICP algorithm [25].
- *Simulated Annealing on Z-Buffers:* This is a refinement step that ensures that the model and the data are correctly aligned. The idea is to refine alignment by minimizing the differences between the z-buffers of the model and data. We employ a global optimization technique known as Enhanced Simulated Annealing (ESA) [26] to minimize the z-buffer difference [27]. The higher accuracy of this step can be attributed to the fact that the z-buffers effectively resample the data which results in independence from the data's triangulation.

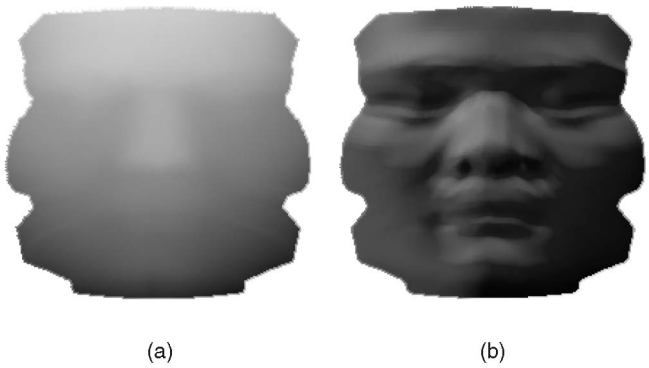


Fig. 5. (a) Geometry and (b) normal map images of a subject's face area.

Note that the proposed multistage alignment process was the result of extensive testing on facial databases. However, we believe that it is a very efficient rigid object alignment method in the general case.

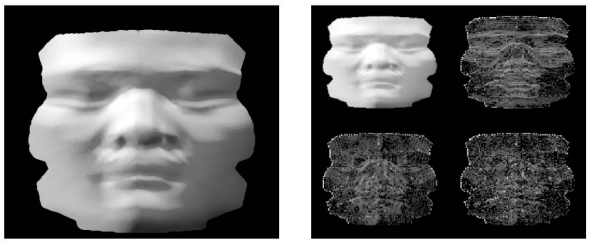
2.1.4 Deformable Model Fitting

The AFM is fitted to each individual data set in order to capture the geometric characteristics of the subject's face using a deformable model-based approach [8], [16]. This is achieved using the elastically adapted deformable model framework of Metaxas and Kakadiaris [28]. Based on the work of Mandal et al. [29], [30], the framework is combined with Loop subdivision surfaces [31]. The solution is approximated iteratively and depends on simulated physical properties. An example of the fitting progress is presented in Fig. 4.

2.1.5 Geometry Image Analysis

The deformed model that is the output of the fitting process is converted to a geometry image, as depicted in Fig. 5a. The geometry image regularly samples the deformed model's surface and encodes this information on a 2D grid. The grid resolution is correlated with the resolution of the AFM's subdivision surface. From the geometry image, a normal map image (Fig. 5b) is constructed. The normal map distributes the information evenly among its three components, in contrast with the geometry image, where most information is concentrated in the Z component.

We treat the three channels (X, Y, and Z) of the normal map and geometry image as separate images. Each component is analyzed using a wavelet transform and the coefficients are stored. We use the Haar and Pyramid transforms, thus obtaining two sets of coefficients. The Pyramid transform is significantly more computationally expensive. We apply the Haar Wavelets on the combined normal/geometry images and the Pyramid transform only on the geometry images.



(a) (b)

Fig. 6. Haar wavelet analysis for the normal map image from Fig. 5: (a) Zero Level. (b) First level.

- *Haar Wavelets*: The first transform is a decimated wavelet decomposition using tensor products of the full Walsh wavelet packet system [32] (Fig. 6). We used this transform extensively in our previous work [8], [16]. The 1D Walsh wavelet packet system is constructed by repeated application of the Haar filters (low-pass: $g = \frac{1}{\sqrt{2}}[1\ \ 1]$ and high-pass: $h = \frac{1}{\sqrt{2}}\ [[1\ -1]]$). For images, we use tensor products of these 1D filters. This means that the filter bank operations are applied separately to the rows and columns of the image, resulting in a four-channel filter bank with channels LL, LH, HL, and HH (corresponding to the filters $g^t * g$, $g^t * h$, $h^t * g$, and $h^t * h$, respectively). We recursively apply this decomposition to each of the four output channels to construct the full Walsh wavelet packet tree decomposition. We store the same subset of coefficients from each subject, allowing an efficient direct comparison of coefficients without the need of reconstruction.
- *Pyramid Transform*: The second transform decomposes the images using the complex version [33] of the steerable pyramid transform [34], a linear multiscale, multiorientation image decomposition algorithm. The image is first divided into high-pass and low-pass subbands by using two initialization filters H_0 and L_0. The low-pass subband is then fed into a set of steerable bandpass filters, which produce a set of oriented subbands and a lower-pass subband. This lower-pass subband is subsampled by 2 and recursively applied the same set of steerable bandpass filters. Such pyramid wavelet representation is translation-invariant and rotation-invariant. This advantage is desirable to address possible positional and rotational displacements caused by facial expressions. To maintain reasonable image resolution and computational complexity our algorithm applies a 3-scale, 10-orientation complex steerable pyramid transform to decompose each component of the geometry image. Only the oriented subbands at the farthest scale are stored. This enables us to compare the subband coefficients of the two images directly without the overhead of reconstruction.

2.1.6 Distance Metrics

In the authentication phase, the comparison between two subjects (gallery and probe), is performed using the metadata. In this paper, we introduce a novel approach that utilizes and combines two different distance metrics for the two transform types (Haar and Pyramid):

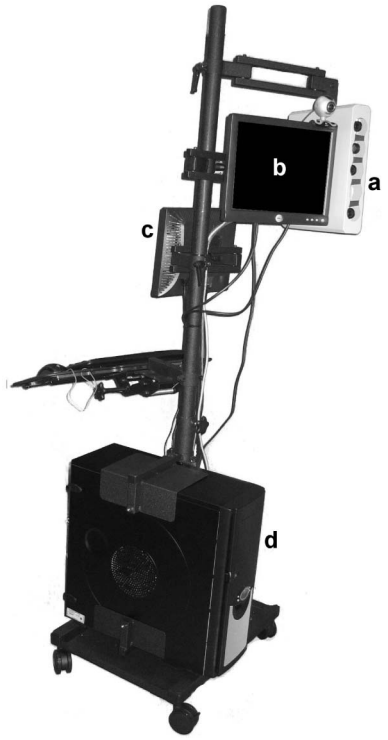


Fig. 7. Prototype system using a 3dMD™ optical scanner with a 1-pod configuration. Individual components: (a) 1-pod scanner, (b) subject's monitor, (c) operator's monitor, and (d) host computer.

Haar Metric. For the Haar wavelet coefficients, we employ a simple L^1 metric on each component independently. For example, the X component is computed as follows:

$$d_x^h(P,G) = \sum_{i,j} |P_x[i,j] - G_x[i,j]|,$$

where P and G are the probe and gallery images, respectively. The total distance is the sum of the distances computed on all components:

$$d^h(P,G) = d_x^h(P,G) + d_y^h(P,G) + d_z^h(P,G).$$

Pyramid Metric. In order to quantify the distance between the two geometry images of the probe and gallery, we need to compare their oriented subband coefficients and assign a numerical score to each area F_k of the face. Each F_k is defined according to the annotation of the face model. Note that because of the presence of facial expression, F_k may be distorted in different ways. These distortions are mostly scaling, translational, and rotational displacements. To that end, based on this, we employ the CW-SSIM index algorithm. CW-SSIM is a translational insensitive image similarity measure inspired by the structural similarity (SSIM) index algorithm [35]. CW-SSIM iteratively measures the similarity indices between the two sliding windows placed in the same positions of the two images and uses the weighted sum as a final similarity score. In our context, a window of size 3 is placed in the oriented subbands and moved across pixels in each subband one step at a time. In each step, we extract all the coefficients associated with F_k within the window, resulting in two sets of coefficients $P_w = \{p_{w,i} | i = 1, \ldots, N\}$ and $G_w = \{g_{w,i} | i = 1, \ldots, N\}$, drawn from the probe and the gallery, respectively. The distance between these two sets is

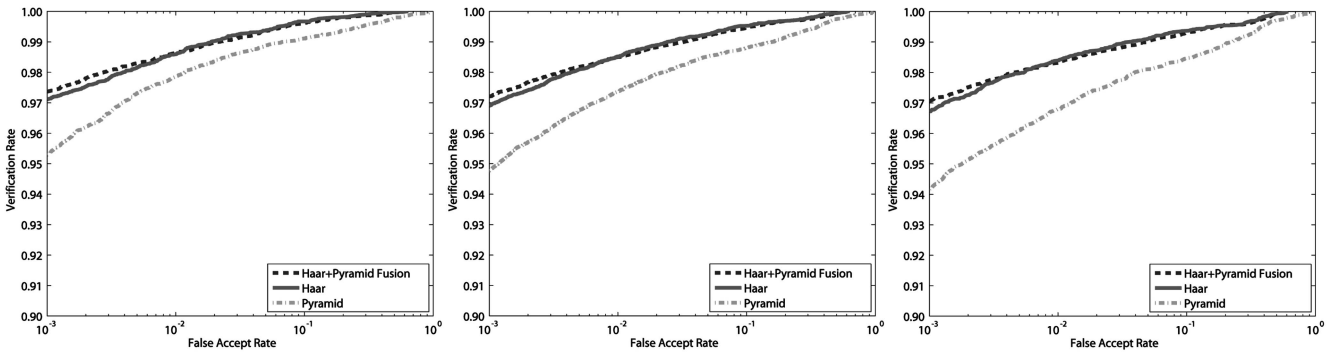


Fig. 8. Performance of our system using the Haar and Pyramid transforms as well as their fusion on the FRGC v2 database. Results reported using: (a) ROC I, (b) ROC II, and (c) ROC III.

measured by a variation of the CW-SSIM index equation originally proposed by Wang and Simoncelli [36]:

$$\begin{aligned}\tilde{S}(P_w, G_w) &= 1 - A_{p_w,g_w} B_{p_w,g_w} C_{p_w,g_w} D_{p_w,g_w},\\ A_{P_w,G_w} &= 2\sum_{i=1}^{N} |p_{w,i}||g_{w,i}| + K,\\ B_{P_w,G_w} &= \left(\sum_{i=1}^{N} |p_{w,i}|^2 + \sum_{i=1}^{N} |g_{w,i}|^2 + K\right)^{-1},\\ C_{P_w,G_w} &= \left(2|\sum_{i=1}^{N} p_{w,i} g^*_{w,i}| + K\right)^{r},\\ D_{P_w,G_w} &= \left(2\sum_{i=1}^{N} |p_{w,i} g^*_{w,i}| + K\right)^{-r}.\end{aligned}$$

The first component measures the equivalence of the two coefficient sets. If $P_w = G_w$, then the subtrahend would be 1 and distance 0 is achieved. The second component reflects the consistency of phase changes, which is insensitive to the translational changes caused by facial expressions. The parameter r is used to tune the amount of distortions we expect. Experimentally, we found that $r = 7$ is optimal for our data sets, but r should be increased if strong facial expression between P and G is known or detected. The parameter K is a small positive number to ensure stable behaviors in the presence of small numbers. In our experiments, we chose K to be 0.01.

As the sliding window moves, the local $\tilde{S}(p_w, g_w)$ at each step w is computed and stored. The weighted sum of the local similarity scores from all computed windows gives the distance score of F_k:

$$e^p(G, P, F_k) = \sum_{w=1}^{N} (b_w \cdot \tilde{S}(P_w, G_w)),$$

where b_w is a predefined weight depending on which subband and component the local window lies on. Experimentally, using the FRGC v1 data set, we found b_w to assume values between 0.4 and 0.8. Finally, the discrete sum of the scores for all F_ks is the overall distance (d^p) between the probe image P and the gallery image G:

$$\begin{aligned}d^p_x(P, G) &= \sum_{k=1}^{N} e^p_x(P, G, F_k) \text{ and}\\ d^p(P, G) &= d^p_x(P, G) + d^p_y(P, G) + d^p_z(P, G).\end{aligned}$$

TABLE 1
Verification Rates of Our System at 0.001 FAR Using Different Transforms on the FRGC v2 Database

	ROC I	ROC II	ROC III
Fusion	97.3%	97.2%	97.0%
Haar	97.1%	96.8%	96.7%
Pyramid	95.2%	94.7%	94.1%

2.2 Prototype

A field-deployable prototype system has been built and is operational at the University of Houston (Fig. 7). A 3dMD optical camera using a 1-pod configuration is currently mounted on the system. This camera system supports multiple pods, with each pod containing two black-and-white cameras for stereo capture, a color camera for texture capture, a speckle pattern projector, and a flash. Each of the cameras has a resolution of 1.2 megapixels. The entire capture process takes less than 2 ms, and it produces a mesh with less than 0.5 mm RMS error (as quoted by the manufacturer).

The system's field-deployable characteristics are:

- *Automation*: All methods utilized are fully automated, requiring no interaction with a user. The system is capable of detecting when a subject is within range and initiating the enrollment or authentication procedures automatically.
- *Space efficiency*: The raw 3D data produced by most scanners are of several MiB. After the enrollment phase, the system needs to keep only the metadata.
- *Time efficiency*: In the enrollment phase, the time delay to convert the raw scanner data to the final metadata is 15 seconds. In the authentication phase, only the stored metadata are utilized. The system can compare the metadata of enrolled subjects at a rate of 1,000/sec., on a typical modern PC (3.0 Ghz P4, 1 GB RAM).

3 Performance Evaluation

3.1 Databases

We use two databases, the publicly available FRGC v2 to allow comparison with other methods and a novel multiple-sensor database to demonstrate the sensor-invariance of our system.

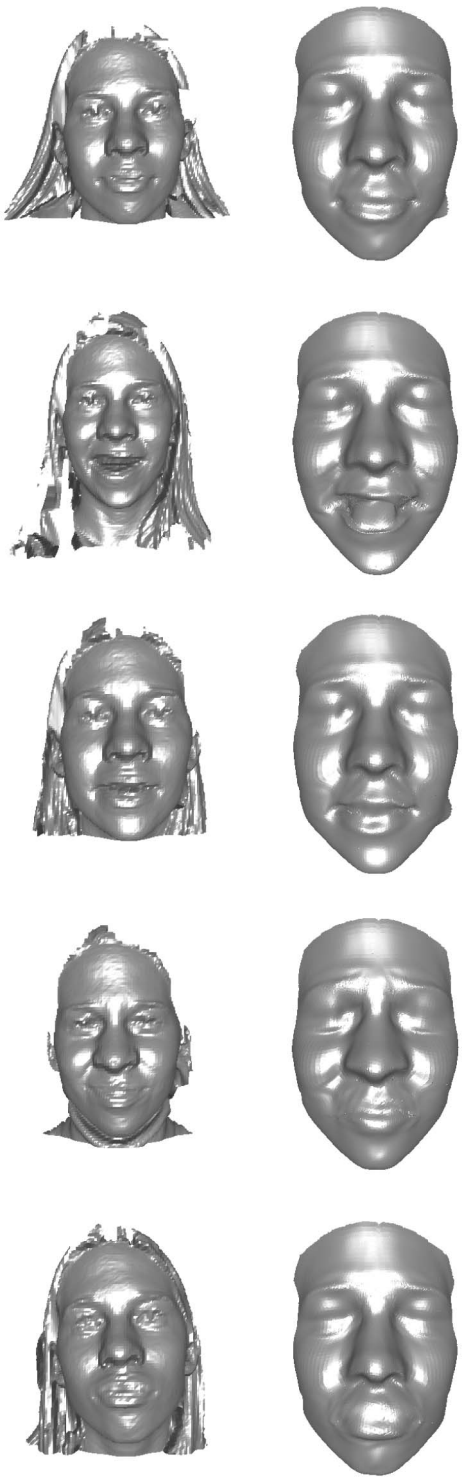


Fig. 9. A single subject with neutral, surprise, happiness, disgust, and sadness expressions along with the corresponding fitted models.

3.1.1 FRGC v2

We utilize the FRGC v2 database, containing 4,007 3D scans of 466 persons. The data were acquired using a Minolta 910 laser scanner that produces range images with a resolution of 640×480. The scans contain various facial expressions (e.g., happiness and surprise), and subjects are 57 percent male and 43 percent female, with the following age distribution: 65 percent are 18-22 years old, 18 percent are 23-27, and 17 percent are 28 years or over [5].

3.1.2 Extended Database

We have extended the FRGC v2 database with the UH database, which contains 884 3D facial data sets acquired using our 3dMD-based prototype system (with 1-pod and 2-pod setups) over a period of one year. The data acquisition protocol was the following:

For each subject:

- Remove any accessories (e.g., glasses).
- Acquire a data set with neutral expression.
- Acquire several data sets while the subject reads loudly a predefined text (thus, assuming facial expressions).
- Put on the accessories and acquire a data set with neutral expression.

The UH database is more challenging, compared to the FRGC v2, as the subjects were encouraged to assume various extreme facial expressions and, in some cases, accessories are present. The resulting extended database contains a total of 4,891 data sets, 82 percent acquired using a laser scanner, 18 percent acquired using an optical camera and, to the best of our knowledge, is the largest 3D facial database reported.

3.2 Performance Metrics

We employ two different scenarios for our experiments: identification and verification. In an identification scenario, we divide the database into probe and gallery sets so that each subject in the probe set has exactly one match in the gallery set. To achieve this, we mark the first data set of every individual as gallery and the rest as probes. During the experiment, each probe is compared against all gallery sets, which is one-to-many matching. The performance is measured using a Cumulative Match Characteristic (CMC) curve and the rank-one recognition rate is reported.

In a verification scenario, we measure the verification rate at 0.001 FAR. Each probe is compared to a gallery set and the result is compared against a threshold. The results are summarized using Receiver Operating Characteristic (ROC) curves. For the FRGC v2 database, in order to produce comparable results, we utilize the three masks provided by FRGC along with the database. These masks, referred to as ROC I, ROC II, and ROC III, are defined over the square similarity matrix ($4,007 \times 4,007$), and they are of increasing difficulty (the difficulty reflects the time elapsed between the probe and gallery acquisition sessions). In the experiments below, we have used parameters which maximize the rank-1 recognition rate on the FRGC v1 database. The value K was chosen to be 0.01, while the b_w weights assumed values between 0.4 and 0.8.

3.3 Experiment 1: Transforms

Experiment 1 is performed on the FRGC v2 database and its purpose is to evaluate the two transforms that we employ, as well as to provide a reference score for our system using publicly available data sets and methods. In this experiment, our system using a fusion of the two transforms yielded a verification rate of 97.3 percent (for ROC I at 0.001 FAR), while separately for the Haar transform a rate of 97.1 percent and for

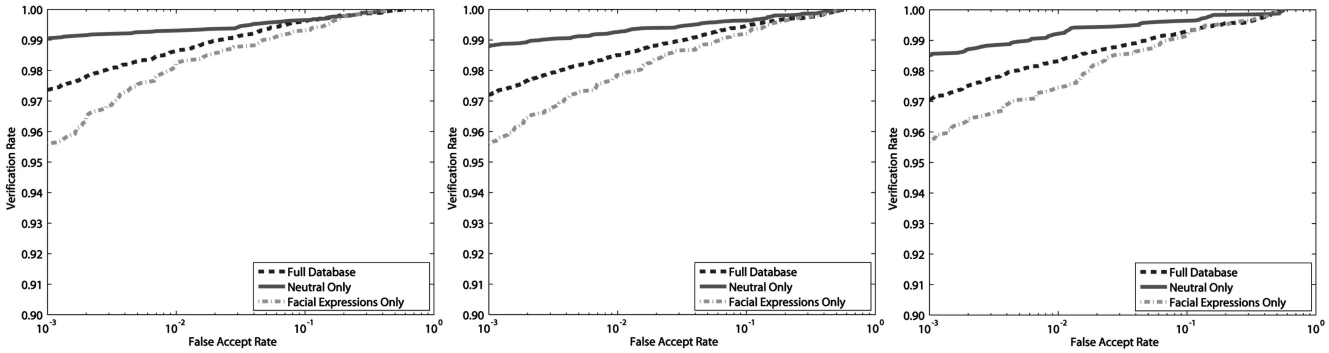


Fig. 10. Division of the FRGC v2 database into two subsets, (the first containing only nonneutral facial expressions and the other one only neutral expressions) and comparison of performance versus the full database. Results reported using: (a) ROC I, (b) ROC II, and (c) ROC III.

the Pyramid transform a rate of 95.2 percent were achieved (Fig. 8). For the fusion, we experimentally found that the weighted sum is the most efficient rule.

Even though the Pyramid transform is computationally more expensive, it is outperformed by the simpler Haar wavelet transform (this can be attributed to the fact that in the current implementation, the Pyramid transform utilizes only the geometry images and not the normal map images). The fusion of the two transforms offers more descriptive power, yielding higher scores especially in the more difficult experiments of ROC II and ROC III, as depicted in Table 1. To the best of our knowledge, this is the highest performance reported on the FRGC v2 database for the 3D modality.

3.4 Experiment 2: Facial Expressions

The second experiment is focused on the effect of facial expressions on performance. An example of these facial expressions for a single individual is depicted in Fig. 9. The FRGC v2 database provides a categorization of the expressions that each individual assumes, allowing an easy division on two subsets: one containing only data sets where facial expressions are present, the other containing only data sets with neutral expressions.

The performance on the two subsets is measured and compared to the performance on the full database utilizing a verification scenario (Fig. 10). The analysis of Table 2 shows that the verification rate is not decreased by a significant amount when expressions are present. The average decrease of 1.56 percent of the verification rate at 0.001 FAR between the full database and the facial expressions-only subset is very modest (compared to other existing algorithms) given the fact that this subset contains the most challenging data sets from the whole database. This can be attributed to the use of the deformable model framework.

TABLE 2
Performance of Our System at 0.001 FAR on the Full FRGC v2 Database, on a Subset Containing Only Nonneutral Facial Expressions and on a Subset Containing Only Neutral Expressions

	ROC I	ROC II	ROC III
Full Database	97.3%	97.2%	97.0%
Non-neutral Expression	95.6%	95.6%	95.6%
Neutral Expression	99.0%	98.7%	98.5%

3.5 Experiment 3: Multiple Sensors

The third experiment evaluates the performance of our system using a multiple-sensor database. Verification experiments depend heavily on the selected facial pairs. In the absence of standard such experiments (e.g., FRGC's ROC experiments), we opted for an identification experiment.

We first measured the performance on the two parts of the extended database separately, obtaining a 97.0 percent rank-one recognition rate for the FRGC v2 and a 93.8 percent rate for the UH (Fig. 11). The experiment on the extended database yielded a rank-one recognition rate of 96.5 percent. The drop in performance in the extended database compared to the FRGC v2 part is marginal, indicating our system's robustness when data from multiple sensors are included on the same database.

4 Conclusion

We presented algorithmic solutions to the majority of the challenges faced by field-deployable 3D facial recognition

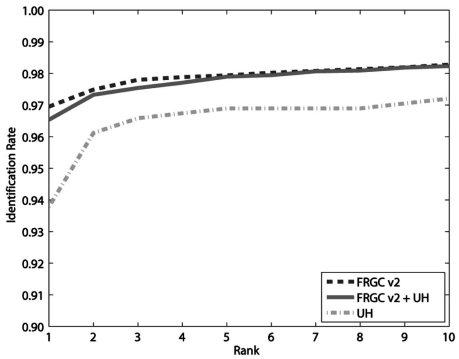


Fig. 11. System performance for identification experiment on different databases: FRGC v2 database with 466 gallery and 3,541 probes (laser scanner), UH database with 240 gallery and 644 probes (optical scanner), and FRGC v2+UH database with 706 gallery and 4,185 probes (both scanners).

systems. By utilizing a deformable model, we map the 3D geometry information onto a 2D regular grid, thus combining the descriptiveness of 3D data with the computational efficiency of 2D data. A multistage fully automatic alignment algorithm and the advanced wavelet analysis resulted in robust state-of-the-art performance on the publicly available FRGC v2 database. Our multiple-sensor database pushed the evaluation envelope one step further, showing that both accuracy and robustness can be achieved when data from different sensors are present, through sensor-oriented preprocessing. Proof of concept is provided by our prototype system which combines competitive accuracy with storage and time efficiency.

ACKNOWLEDGMENTS

The authors are grateful for the support provided by the Department of Computer Science, University of Houston, the Texas Learning and Computation Center, University of Houston, the Central Intelligence Agency, and the Southwest Public Safety Technology Center, University of Houston.

REFERENCES

[1] K. Bowyer, K. Chang, and P. Flynn, "A Survey of Approaches and Challenges in 3D and Multi-Modal 3D + 2D Face Recognition," *Computer Vision and Image Understanding,* vol. 101, no. 1, pp. 1-15, Jan. 2006.

[2] P. Phillips, A. Martin, C. Wilson, and M. Przybocki, "An Introduction to Evaluating Biometric Systems," *Computer,* vol. 33, no. 2, pp. 56-63, Feb. 2000.

[3] K. Chang, K. Bowyer, and P. Flynn, "An Evaluation of Multi-Modal 2D + 3D Face Biometrics," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 27, no. 4, pp. 619-624, Apr. 2005.

[4] "Face Recognition Vendor Test 2006," http://www.frvt.org/FRVT2006/.

[5] P. Phillips, P. Flynn, T. Scruggs, K. Bowyer, J. Chang, K. Hoffman, J. Marques, J. Min, and W. Worek, "Overview of the Face Recognition Grand Challenge," *Proc. IEEE Computer Vision and Pattern Recognition,* pp. 947-954, June 2005.

[6] G. Pan, S. Han, Z. Wu, and Y. Wang, "3D Face Recognition Using Mapped Depth Images," *Proc. IEEE Workshop Face Recognition Grand Challenge Experiments,* pp. 175-181, June 2005.

[7] T. Russ, K. Koch, and C. Little, "3D Facial Recognition: A Quantitative Analysis," *Proc. 45 Ann. Meeting of the Inst. Nuclear Materials Management (INMM),* pp. 338-344, July 2004.

[8] I. Kakadiaris, G. Passalis, T. Theoharis, G. Toderici, I. Konstantinidis, and N. Murtuza, "Multimodal Face Recognition: Combination of Geometry with Physiological Information," *Proc. IEEE Computer Vision and Pattern Recognition,* pp. 1022-1029, June 2005.

[9] K. Chang, K. Bowyer, and P. Flynn, "Effects on Facial Expression in 3D Face Recognition," *Proc. SPIE Biometric Technology for Human Identification II,* vol. 5779, pp. 132-143, 2005.

[10] K. Chang, K. Bowyer, and P. Flynn, "Adaptive Rigid Multi-Region Selection for Handling Expression Variation in 3D Face Recognition," *Proc. IEEE Computer Vision and Pattern Recognition,* pp. 157-164, 2005.

[11] X. Lu and A. Jain, "Deformation Modeling for Robust 3D Face Matching," *Proc. IEEE Computer Vision and Pattern Recognition,* pp. 1377-1383, June 2006.

[12] T. Russ, C. Boehnen, and T. Peters, "3D Face Recognition Using 3D Alignment for PCA," *Proc. IEEE Computer Vision and Pattern Recognition,* pp. 1391-1398, June 2006.

[13] W. Lin, K. Wong, N. Boston, and Y. Hu, "Fusion of Summation Invariants in 3D Human Face Recognition," *Proc. IEEE Computer Vision and Pattern Recognition,* pp. 1369-1376, June 2006.

[14] M. Husken, M. Brauckmann, S. Gehlen, and C. von der Malsburg, "Strategies and Benefits of Fusion of 2D and 3D Face Recognition," *Proc. IEEE Workshop Face Recognition Grand Challenge Experiments,* pp. 174-181, June 2005.

[15] 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," *Proc. IEEE Workshop on Face Recognition Grand Challenge Experiments,* June 2005.

[16] G. Passalis, I. Kakadiaris, T. Theoharis, G. Toderici, and N. Murtuza, "Evaluation of 3D Face Recognition in the Presence of Facial Expressions: An Annotated Deformable Model Approach," *Proc. IEEE Workshop Face Recognition Grand Challenge Experiments,* pp. 171-179, June 2005.

[17] I. Kakadiaris, G. Passalis, G. Toderici, N. Karampatziakis, N. Murtuza, Y. Lu, and T. Theoharis, "Expression-Invariant Multispectral Face Recognition: You Can Smile Now!" *Proc. SPIE Defense and Security Symp.,* Apr. 2006.

[18] L. Farkas, *Anthropometry of the Head and Face.* Raven Press, 1994.

[19] I. Kakadiaris, M. Papadakis, L. Shen, D. Kouri, and D. Hoffman, "m-HDAF Multiresolution Deformable Models," *Proc. 14th Int'l Conf. Digital Signal Processing,* pp. 505-508, July 2002.

[20] I. Kakadiaris, L. Shen, M. Papadakis, D. Kouri, and D. Hoffman, "g-HDAF Multiresolution Deformable Models for Shape Modeling and Reconstruction," *Proc. British Machine Vision Conf.,* pp. 303-312, Sept. 2002.

[21] X. Gu, S. Gortler, and H. Hoppe, "Geometry Images," *Proc. ACM SIGGRAPH,* pp. 355-361, July 2002.

[22] A. Johnson, "Spin-Images: A Representation for 3-D Surface Matching," PhD dissertation, Robotics Inst., Carnegie Mellon Univ., Pittsburgh, Penn., Aug. 1997.

[23] P. Besl and N. McKay, "A Method for Registration of 3-D Shapes," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 14, no. 2, pp. 239-256, Feb. 1992.

[24] G. Turk and M. Levoy, "Zippered Polygon Meshes from Range Images," *Proc. ACM SIGGRAPH,* pp. 311-318, 1994.

[25] D. Chetverikov, D. Stepanov, and P. Krsek, "Robust Euclidean Alignment of 3D Point Sets: The Trimmed Iterative Closest Point Algorithm," *Image Vision Computing,* vol. 23, no. 3, pp. 299-309, 2005.

[26] P. Siarry, G. Berthiau, F. Durbin, and J. Haussy, "Enhanced Simulated Annealing for Globally Minimizing Functions of Many-Continuous Variables," *ACM Trans. Math. Software,* vol. 23, no. 2, pp. 209-228, 1997.

[27] G. Papaioannou, E. Karabassi, and T. Theoharis, "Reconstruction of Three-Dimensional Objects through Matching of Their Parts," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 24, no. 1, pp. 114-124, Jan. 2002.

[28] D. Metaxas and I. Kakadiaris, "Elastically Adaptive Deformable Models," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 24, no. 10, pp. 1310-1321, Oct. 2002.

[29] C. Mandal, H. Qin, and B. Vemuri, "Dynamic Smooth Subdivision Surfaces for Data Visualization," *Visualization,* pp. 371-377, Oct. 1997.

[30] C. Mandal, H. Qin, and B. Vemuri, "A Novel FEM-Based Dynamic Framework for Subdivision Surfaces," *Computer-Aided Design,* vol. 32, nos. 8-9, pp. 479-497, 2000.

[31] C. Loop, "Smooth Subdivision Surfaces Based on Triangles," master's thesis, Dept. of Math., Univ. of Utah, 1987.

[32] E. Stollnitz, T. DeRose, and D. Salesin, *Wavelets for Computer Graphics: Theory and Applications.* Morgan Kaufmann, 1996.

[33] J. Portilla and E. Simoncelli, "A Parametric Texture Model Based on Joint Statistic of Complex Wavelet Coefficients," *Int'l J. Computer Vision,* vol. 40, pp. 49-71, 2000.

[34] E. Simoncelli, W. Freeman, E. Adelson, and D. Heeger, "Shiftable Multi-Scale Transforms," *IEEE Trans. Information Theory,* vol. 38, pp. 587-607, 1992.

[35] Z. Wang, A. Bovik, H. Sheikh, and E. Simoncelli, "Image Quality Assessment: From Error Visibility to Structural Similarity," *IEEE Trans. Image Processing,* vol. 13, no. 4, pp. 600-612, Apr. 2004.

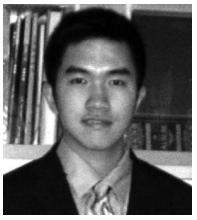
[36] Z. Wang and E. Simoncelli, "Translation Insensitive Image Similarity in Complex Wavelet Domain," *Proc. IEEE Int'l Conf. Acoustics, Speech, and Signal Processing,* vol. II, pp. 573-576, Mar. 2005.

Ioannis A. Kakadiaris received the Ptychion (BSc) degree in physics from the University of Athens, Greece, in 1989, the MSc degree in computer science from Northeastern University, Boston, Massachusetts, in 1991, and the PhD degree in computer science from the University of Pennsylvania, Philadelphia, in 1997. Dr. Kakadiaris joined the University of Houston (UH) in August 1997 after completing a postdoctoral fellowship at the University of Pennsylvania. He is the founder and director of UH's Computational Biomedicine Laboratory (formerly the Visual Computing Lab) and director of the Division of Bio-Imaging and Bio-Computation at the UH Institute for Digital Informatics and Analysis. Professor Kakadiaris' research interests include biomedical image analysis, biometrics, computer vision, and pattern recognition. He is the recipient of the year 2000 NSF Early Career Development Award, UH Computer Science Research Excellence Award, UH Enron Teaching Excellence Award, James Muller VP Young Investigator Prize, and the Schlumberger Technical Foundation Award. He is a member of the IEEE.

Georgios Passalis received the BSc degree from the Department of Informatics and Telecommunications, University of Athens. He subsequently received the MSc degree from the Department of Computer Science, University of Houston. Currently, he is a PhD candidate at the University of Athens and a research associate at the Computational Biomedicine Lab, University of Houston. His thesis is focused on the domains of computer graphics and computer vision. His research interests include object retrieval, face recognition, hardware accelerated voxelization, and object reconstruction.

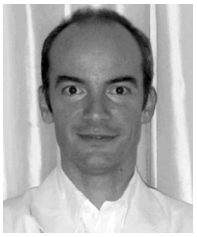
George Toderici received the BSc degree in computer science and mathematics from the University of Houston. Currently, he is a PhD candidate at the University of Houston. He is a member of the Computational Biomedicine Lab focusing on face recognition research. His research interests include machine learning, pattern recognition, object retrieval, and their possible applications on the GPU.

Mohammed N. Murtuza received the BSc degree in computer science from Texas A&M University, College Station, in 2003. Currently, he is a research assistant at the Computational Biomedicine Lab. His research interests include face recognition, ear recognition, 3D biometric systems, genetic algorithms, and image processing.



Yunliang Lu received the BSc degree from the University of Houston (with a major in computer science and a minor in mathematics) in 2006. During his studies, he served as a research assistant at the Computational Biomedicine Lab, University of Houston. Currently, he is a seismic software programmer in Veritas DGC Inc.

Nikos Karampatziakis received the BSc and MSc (with honors) degrees from the Department of Informatics and Telecommunications, University of Athens, Greece, in 2003 and 2005, respectively. He is currently pursuing the PhD degree at the Department of Computer Science, Cornell University, New York. His research interests include computer vision, artificial intelligence, and machine learning.



Theoharis Theoharis received the DPhil degree in computer graphics and parallel processing from the University of Oxford in 1988. He subsequently served as a research fellow (postdoctoral) at the University of Cambridge and as a consultant with Andersen Consulting. He is currently an associate professor with the University of Athens and adjunct faculty member with the Computational Biomedicine Lab, University of Houston. His main research interests lie in the fields of computer graphics, visualization, biometrics, and archaeological reconstruction.