3D FACE RECOGNITION PERFORMANCE UNDER ADVERSARIAL CONDITIONS

Arman Savran [1], Oya Çeliktutan [1], Aydın Akyol [2], Jana Trojanová [3], Hamdi Dibeklioğlu [4], Semih Esenlik [1], Nesli Bozkurt [5], Cem Demirkır [1], Erdem Akagündüz [5], Kerem Çalışkan [6], Neşe Alyüz [4], Bülent Sankur [1], İlkay Ulusoy [5], Lale Akarun [4], Tevfik Metin Sezgin [7]

[1] Boğaziçi University, Department of Electrical and Electronic Engineering, İstanbul, Turkey
[2] İstanbul Technical University, Department of Computer Engineering, Turkey
[3] Department of Cybernetics, University of West Bohemia in Pilsen, Czech Republic
[4] Boğaziçi University, Department of Computer Engineering, Turkey
[5] Middle Eastern Technical University, Dept. of Electrical and Electronics Engineering, Turkey
[6] Middle Eastern Technical University, Informatics Institute, Turkey
[7] University of Cambridge, Computer Laboratory, UK

ABSTRACT

We address the question of 3D face recognition and expression understanding under adverse conditions like illumination, pose, and accessories. We therefore conduct a campaign to build a 3D face database including systematic variation of poses, different types of occlusions, and a rich set of expressions. The expressions consist of a judiciously selected subset of Action Units as well as the six basic emotions. The database is designed to enable various research paths from face recognition to facial landmarking and to expression estimation. Preliminary results are presented on the outcome of three different landmarking methods as well as one registration method. As expected, observed non-neutral and non-frontal faces demand new robust algorithms to achieve an acceptable performance.

KEYWORDS

3D face database – Facial expressions – Facial landmarking – Face recognition

1. INTRODUCTION

The history of automated face recognition dates back to the early 1960s. The overwhelming majority of these techniques are based on 2D face data. The early approaches were focused on the geometry of key points (eyes, mouth and nose) and relations between them (length, angles). In the 1990s, the principal component analysis (PCA) based eigenface algorithm was introduced and it became a standard reference in face recognition. Despite the plethora of 2D algorithms, there is not yet a fully reliable identification and verification method that can operate under adverse conditions of illumination, pose, accessories and expression. Two solutions to this impasse consist in i) Exploring new modalities, such as 3D facial data; ii) The use of multi-biometrics, that is, the employment of more than one biometric modality and their judicious fusion.

Recently face recognizers using 3D facial data have gained popularity due to their lighting and viewpoint independence. This has also been enabled by the wider availability of 3D range scanners. The 3D face processing can be envisioned as a single modality biometric approach in lieu of the 2D version or in a complementary mode in a multi-biometric scheme. Another goal application of 3D facial data is the understanding of facial expressions in an affective human-computer interface.

Based on the ideas above, two main goals are addressed in this project:

1.1. 3D Recognition of Non-cooperative Persons

Most of the existing methods for facial feature detection and person recognition assume frontal and neutral views only. In these studies data are therefore collected from cooperative subjects, who expose their faces in a still position in front of the scanner, frontal poses, and avoid extreme expressions and any occluding material. This may be uncomfortable to subjects. In fact, the second generation techniques vie for improved quality of life via biometry, hence enrollment and updating should proceed in ways that do not encumber the subject. On the other extreme, a subject, aware of person identification cameras, may try to eschew being recognized by posing awkwardly and worse still, by resorting to occlusions via dangling hair, eyeglasses, facial hair etc. Also, the 3D data may have different translation, rotation or scaling due to the controlled environmental parameters such as the acquisition setup and device properties. Both the natural, uncontrolled behaviour of subjects and the mimics and acrobatics of the eschewer seriously damage the performance of both 2D and 3D recognition algorithms. Furthermore, we believe that 3D capture of face data can mitigate significantly most of these effects. Using a database built specifically for this purpose, we test the performance of 3D face identification algorithms as well as those of automatic landmarking and registration under non-frontal poses, in presence of facial expression, gestures and occlusions.

1.2. Facial expression understanding

Understanding of facial expressions has wide implications ranging from psychological analysis to affective man-machine interfaces. Once the expression is recognized, it can also be neutralized for improved person recognition. Automatic recognition of facial expressions is challenging since the geometry of the face can change rapidly as a result of facial muscle contractions. Expressions are often accompanied by purposeful or involuntary pose variations of the subject's face. 3D representation of the human face contains more information than its 2D appearance. We conjecture that not all shape changes on the facial surfaces due to expressions are reflected in the 2D images. Certain surfaces deformations, bulges and creases, especially in the smooth areas of the face, are hard to track in 2D while they are apparent on 3D data.

Among the facial signal processing problems, we want to address to the above stated goals by providing solutions to the following problems:

1.3. 3D face registration

3D shapes need to be aligned to each other and should be brought into a common coordinate frame before any comparative analysis can be made. The 3D face registration is the process of defining a transformation that will closely align two faces. First of all, the permissible transformations should be set. Rigid transformations allow only for translation, rotation or scaling. Non-rigid transformations go one step further and allow patchwise deformation of facial surfaces within the constraints of estimated landmark points. Secondly, a similarity measure should be determined that favors successful recognition. Different similarity measures are possible for registration techniques such as the point-to-point or point-to-surface distances.

1.4. 3D face landmarking

Facial landmarks are essential for such tasks as face registration, expression understanding and any related processing. In general, registration process is guided by a set of fiducial points called landmarks [1] that are used to define the transform between two surfaces. The registration techniques that are examined in the scope of this project make use of a set of landmark points that are labeled for each face surface to be aligned. Preliminary studies of landmarking of uncontrolled facial expressions [2] indicate that it remains still as an open problem.

1.5. 3D facial expression and face recognition database construction

There are very few publicly available databases of annotated 3D facial expressions. Wang et. al [3] studied recognition of expressions by extracting primitive surface features that describe the shape. They use their database dedicated to expressions BU-3DFE [4]. It includes only emotional expressions (happiness, surprise, fear, sadness, anger, disgust) with four intensity levels from 100 people.

In this project, we aim to study a more comprehensive set of expressions by covering many expressions based on Facial Action Coding System (FACS) [5] in addition to the six basic emotional expressions. There are 44 Action Units (AUs) defined in FACS. AUs are assumed to be building blocks of expressions, and thus they can give broad basis for facial expressions. Also, since each of them is related with activation of distinct set of muscles, they can be assessed quite objectively. We also include various poses and occlusion conditions. We have collected data from 81 subjects, preprocessed and manually landmarked the data. We have defined the metadata standard. We have also performed basic registration and classification tests. This multi-expression and multi-pose database, which will eventually be made public, will be instrumental in advancing the expression analysis research and develop algorithms for 3D face recognition under adverse conditions.

The rest of the report is as follows: Section 2 describes the database content, and the database structure is detailed in Section 3. Section 4 gives the necessary details of the data acquisition setup. Landmarking issues are dealt in two separate sections: Section 5 and 6, respectively, presents manual landmarking and automatic landmarking techniques. The 3D face registration method applied to our database is given in Section 7. Finally, Section 8 shows the performance results of registration, automatic landmarking and recognition. Conclusions are drawn in Section 9.

2. DATABASE CONTENT

In this project, on the one hand we model attempts to invalidate 3D face recognition and any other effort to mislead the system or to induce a fake character. To this effect, we capture 3D face data imitating difficult surveillance conditions and non-cooperating subjects, trying various realistic but effective occlusions and poses. On the other hand, we collect 3D facial data corresponding to various action units and to various emotional expressions.

The database consists of 81 subjects in various poses, expressions and occlusion conditions. The images are captured using Inspeck Mega Capturor II [6], and where each scan is manually labeled for facial landmark points such as nose tip, inner eye corner, etc. Our database has two versions:

- Database1 is designed for both expression understanding and face recognition. Here there are 47 people with 53 different face scans per subject. Each scan is intended to cover one pose and/or one expression type, and most of the subjects have only one neutral face, though some of them have two. Totally there are 34 expressions, 13 poses, four occlusions and one or two neutral faces. In addition, Database1 also incorporates 30 professional actors/actresses out of 47, which hopefully provide more realistic or at least more pronounced expressions.
- Database2 includes 34 subjects with only 10 expressions, 13 poses, four occlusions and four neutral faces and four neutral faces, thus resulting in a total of 31 scans per subject.

The majority of the subjects are aged between 25 and 35. There are 51 men and 30 women in total, and most of the subjects are Caucasian.

This database can be used for automatic facial landmark detection, face detection, face pose estimation, registration, recognition and facial expression recognition purposes. Hence we also labeled facial fiducial points manually, and provide this information with the database.

In the following subsections, the collected facial expressions, head poses and occlusions are explained.

2.1. Facial Expressions

We have considered two types of expressions. In the first set, the expressions are based on AU of the FACS [5]. However, within this project a subset of them, which are more common and easier to enact, is considered. The selected action units are grouped into 19 lower face AUs, five upper face AUs and three AU combinations. It is important to note that, for some subjects properly producing some AUs may not be possible, because they are not able to activate related muscles or they do not know how to control them. Therefore, in the database some expressions are not available for some subjects. Also, unless all the acquired AUs are validated by trained AU coders, the captured AUs can not be verified.

In the second set, facial expressions corresponding to certain emotional expressions are collected. We have considered the following emotions: happiness, surprise, fear, sadness, anger and disgust. It is stated that these expressions are universal among human races [7].

During acquisition of each action unit, subjects were given explications about these expressions and they were given feedback if they did not enact correctly. Also to facilitate the instructions, a video clip showing the correct facial motion for the corresponding action unit is displayed on the monitor [8, 9]. However, in the case of emotional expressions, there were no video

1. Lower Face Action Units
   - Lower Lip Depressor - AU16
   - Lips Part - AU25
   - Jaw Drop - AU26
   - Mouth Stretch - AU27 (*)
   - Lip Corner Puller - AU12 (*)
   - Left Lip Corner Puller - AU12L
   - Right Lip Corner Puller - AU12R
   - Low Intensity Lip Corner Puller - AU12LW
   - Dimpler - AU14
   - Lip Stretcher - AU20
   - Lip Corner Depressor - AU15
   - Chin Raiser - AU17
   - Lip Funneler - AU22
   - Lip Puckerer - AU18
   - Lip Tightener - AU23
   - Lip Presser - AU24
   - Lip Suck - AU28 (*)
   - Upper Lip Raiser - AU10
   - Nose Wrinkler - AU9 (*)
   - Cheek Puff - AU34 (*)
2. Upper Face Action Units
   - Outer Brow Raiser - AU2 (*)
   - Brow Lowerer - AU4 (*)
   - Inner Brow Raiser - AU1
   - Squint - AU44
   - Eyes Closed - AU43 (*)
3. Some Action Unit Combinations
   - Jaw Drop (26) + Low Intensity Lip Corner Puller
   - Lip Funneler (22) + Lips Part (25) (*)
   - Lip Corner Puller (12) + Lip Corner Depressor (15)
4. Emotions
   - Happiness (*)
   - Surprise
   - Fear
   - Sadness
   - Anger
   - Disgust

The expressions marked with (*) are available in both Database1 and Database2, but the others are only found in Database1.

Table 1: *Lower face action units.*

or photo guidelines so that subjects tried to improvise. Only if they were able to enact, they were told to mimic the expression in a recorded video. Moreover, a mirror is placed in front of the subjects in order to let them check themselves.

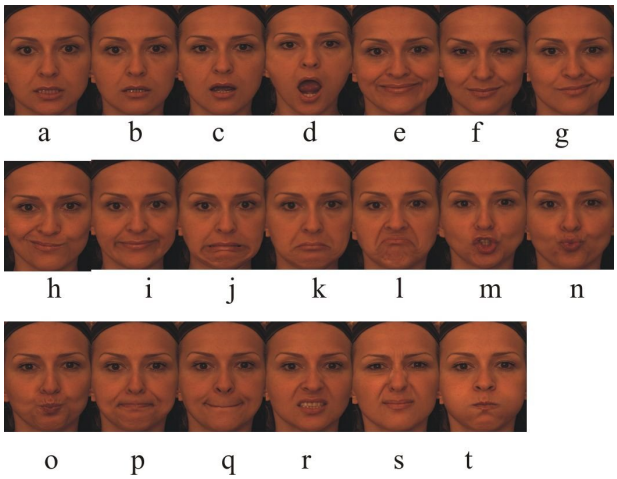

Figure 1: *Lower Face Action Units: lower lip depressor (a), lips part (b), jaw drop (c), mouth stretch (d), lip corner puller (e), low intensity lower lip depressor (f), left lip corner puller (g), right lip corner puller (h), dimpler (i), lip stretcher (j), lip corner depressor (k), chin raiser (l), lip funneler (m), lip puckerer (n), lip tightener (o), lip presser (p), lip suck (q), upper lip raiser (r), nose wrinkler (s), cheek puff (t).*

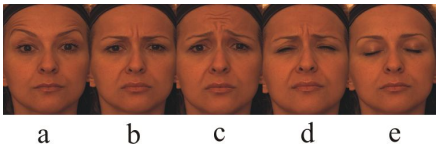

Figure 2: *Upper Face Action Units: outer brow raiser (a), brow lowerer (b), inner brow raiser (c), squint (d), eyes closed (e).*

Figure 3: *Some Action Unit Combinations: lip corner puller + lip corner depressor (a), jaw drop + lip corner puller (b), lip funneler + lips part (c).*

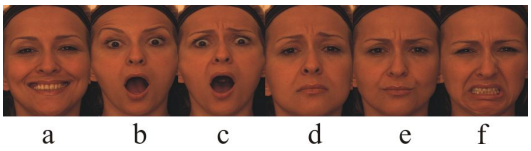

Figure 4: *Emotional expressions: happiness (a), surprise (b), fear (c), sadness (d), angry (e), disgust (f).*

The total of 35 expressions consist of 19 AUs belonging to lower part of the face (Fig. 1), five AUs from the upper face (Fig. 2), three AU combinations (Fig. 3) and six emotional expressions (Fig. 4) as listed in Table 1. In these tables, the expressions marked with (*) are available in both Database1 and Database2, but the others are only to be found in Database1.

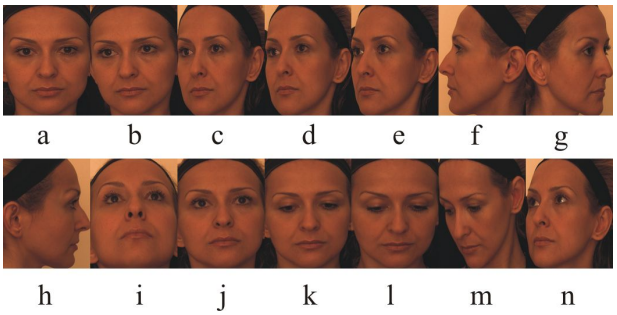

Figure 5: *Poses: neutral pose (a); yaw rotations of +10° (b), +20° (c), +30° (d), +45° (e), +90° (f), -45° (g) and -90° (h) respectively; pitch rotations of strong upwards (i), slight upwards (j), slight downwards (k), strong downwards (l); bottom-right (m) and upper right (n).*

2.2. Head Poses

Various poses of head are acquired for each subject (Fig. 5). There are three types of head which are seven angles of yaw, four angles of pitch, and two cross rotations which incorporate both yaw and pitch. For the yaw rotations, subjects align themselves by rotating the chair on which they sit to align with straps placed on the floor corresponding to various angles. For pitch and cross rotations, we requested the subjects to look at marks placed on the walls by turning their heads only (i.e., no eye rotation). Thus, we can obtain a coarse approximation of rotation angles. The head poses are listed Table 2.

1. Yaw Rotations
   - +10°
   - +20°
   - +30°
   - +45°
   - +90°
   - -45°
   - -90°
2. Pitch Rotations
   - Strong upwards
   - Slight upwards
   - Slight downwards
   - Strong upwards
3. Cross Rotations
   - Yaw and pitch 1 (approximately 20° pitch and 45° yaw)
   - Yaw and pitch 2 (approximately -20° pitch and 45° yaw)

Table 2: *Head poses*

2.3. Occlusions

This database also contains several types of occlusions that are listed in Table 3 and shown in Fig. 6.

For the occlusion of eyes and mouth, subjects choose a natural pose of themselves; for example, as if they were cleaning their eyes or they were surprised by putting their hands over their mouth. Second, for the eyeglasses occlusion, we made so that subjects used different eyeglasses from a pool. Finally, if subjects' hairs were long enough, we have also scanned their faces with hair partly occluding their face.

Figure 6: *Occlusions: eye occlusion (a), mouth occlusion (b), eye glasses (c), hair (d).*

- Occlusion of eye with hand - as natural as possible
- Occlusion of mouth with hand - as natural as possible
- Eye glasses (not sunglasses, normal eyeglasses)
- Hair

Table 3: *Occlusions*

3. THE STRUCTURE OF THE DATABASE

In many image retrieval database applications two approaches are widely used. One is storing any image-like data onto hard disk and the rest of the data to an RDBMS (Relational Database Management System) while the second approach is storing image and its metadata directly on the hard disk. While the first one provides easy querying, it has the disadvantage of using and managing third-party RDBMS software. Also not every RDBMS may work properly in cross-platform systems. In addition it necessitates knowledge of using SQL (Structured Query Language). We suspect that not all researchers are well versed in SQL. For these reasons we decided to use the second alternative, that is, storage of images and their metadata on the harddisk. For this purpose, data storage and mapping structures are designed and a traditional database design used in RDBMSs is considered. In this design we make use of primary key and weak referencing. A primitive referential integrity is used inside the software for preventing redundancy and making relations between tree-like structures. The first version of the design is given in Appendix 12.1.

One can map the entire database and the storage structure based on the above design. We choose to use XML structure in data storage because in addition to having a human readable format XML is also commonly used in non RDBMS applications in the recent years. A record is a row in a table, each column in a record shows us an attribute of the record. Every table defines its attributes and lists its data after declaration of attributes. Content of the subject table exported from Subject.xml is given as an example in Appendix 12.2.

Populating the database manually is always very hard and thus a problematic application. One can easily commit errors during acquisition and processing stages. In order to overcome this problem we decided to write a program using our database structure which helps the acquisition application. Class diagram of the program shows that it is a general time-saving stratagem. Initial class designs involve inheritance of the form structures for easy usage and also involve usage of basic anti-patterns like Singleton for assuring that only single instances of Business Classes exist inside the program for managing data sessions. A schema of the class diagram is found in Fig. 7.

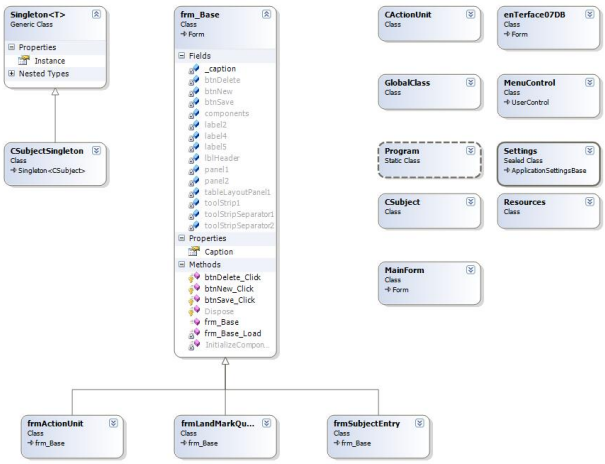

Figure 7: *A schema of the class diagram. Business classes are used for constructing the objects like Subject, Session and Records. Since these objects are used as object construction architecture, each class has its persistency structure, letting it to be written to database in any change. This is a standard mechanism called "Reflection" in "Object Relational"*

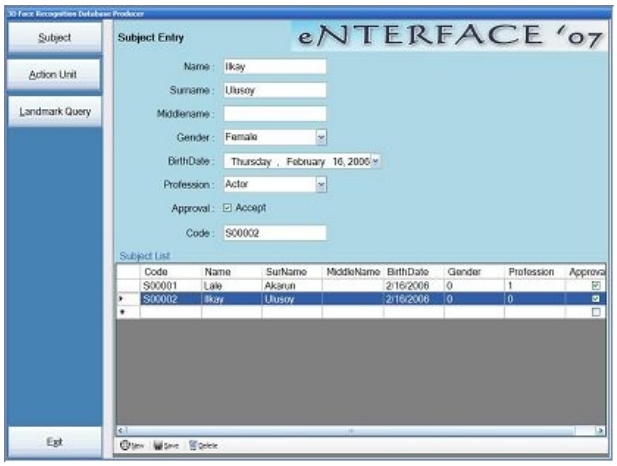

Figure 8: *User can insert new subjects to the database from this screen; editing or deleting previously saved subjects is also done from here.*

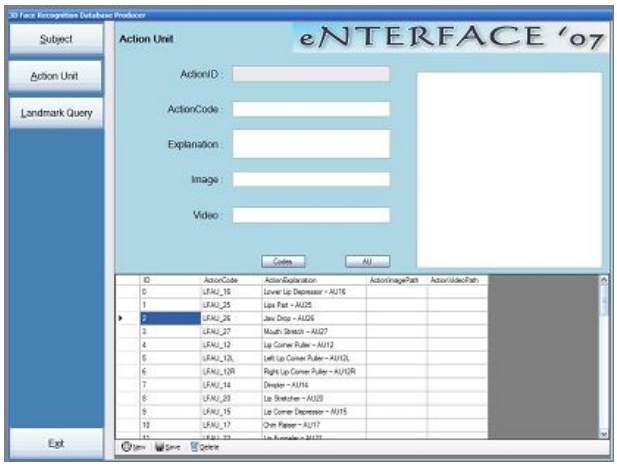

Figure 9: *User can insert the action units used during posing from this screen. Each action unit, its code and explanation has to be inserted. This also lets user to relate his/her action units to be used in their landmarking study to general structure. This part is vital for researchers for choosing their own database structure to be used in their landmarking algorithms.*

In order to write a quick code in limited time we decided to use .net 2.0 framework and C# as the programming language. We couldn't try the cross platform compatibility of the code using Mono in Linux, but we've implemented the code in cross platform manner. The LandMark querying part of the software is finished for the use of participants during Manual - Automatic Landmarking comparisons. For landmarking users have no chance but add the destination directory names as input to their software. After using our software user will be able to query the system by gender, occlusion or any structure they want and save the results to a document which can be input to their implementations. After having such an input document they can easily test their algorithms on the data they exactly want. While querying and finding the data directory paths is working properly, exporting this structure to an input file will be implemented with the acknowledgements coming from the users who want to test their landmarking algorithms on the database. Some screen shots of the software is shown in Fig. 8, 9 and 10.

As a conclusion, the proposed database structure does not need any third-party component or querying structure. Since it is XML based, it is human readable, and can also be used in many acquisition database structures owing to its easy and generic implementation. We hope the structure we've proposed will evolve and become a database structure for similar research applications.

4. DATA ACQUISITION

Facial data are acquired using Inspeck Mega Capturor II 3D, which is a commercial structured-light based 3D digitizer device [6]. It is able to capture a face in less than a second. Our acquisition setup is displayed in Fig. 11. Subjects are made to sit at a distance of about 1.5 meters away from the 3D digitizer. To obtain a homogenous lighting directed on the face we use a 1000W halogen lamp without any other ambient illumination. However, due to the strong lighting of this lamp and the device's projector, usually specular reflections occur on the face. This does not only affect the texture image of the face but can also cause noise in the 3D data. To prevent it, we apply a special powder to the subject's face which does not change the skin color. Moreover, during acquisition, each subject wears a band to keep his/her hair above the forehead. This also simplifies the segmentation of face before 3D reconstruction.

The on-board software in the scanner is used for acquisition and 3D model reconstruction. Inspeck's 3D digitizer is based on an optical principle, which combines Active Optical Triangulation and Phase Shifted Moir (Fig. 12). The surface of the object must reflect light back to the 3D digitizer. Hence transparent surfaces (like glass) can not be scanned. Certain fringe patterns, that are shifted versions of each others, are projected on the 3D surface of an object, and as a result these patterns are deformed according to object surface. In the Inspeck's system, a fringe pattern is in the form of straight vertical lines. All the 3D information is extracted from these deformed patterns. During acquisition we employed usually four fringe patterns. However, sometimes we preferred to use three to diminish the effect of motion for some subjects.

Once the raw data are acquired, three stages of processing (called PreProcessing, Processing and PostProcessing in the de-

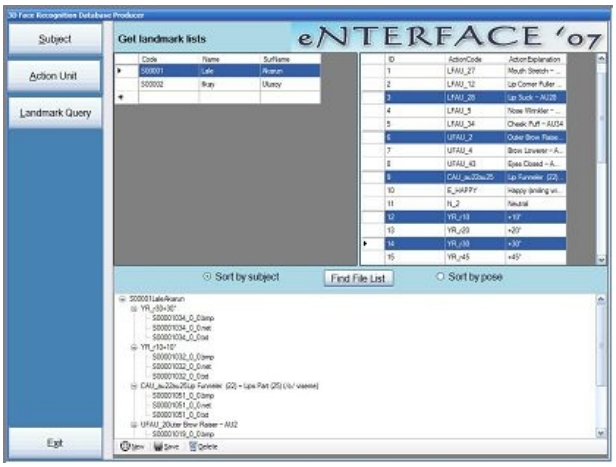

Figure 10: *Querying of the image data of manual landmarks and sorting due to subject or pose. User can query the data paths related to subject and their poses to be used in landmarking. Software user can select the subjects and the action units he/she wants to relate during landmarking. This query returns the user paths of the data they want to use.*

Figure 11: *Acquisition setup. At the left 3D Digitizer, which is mounted on a tripod, is seen. A mirror, an LCD display and a 1000W halogen lamp are placed below, above and behind of it respectively. A seat is positioned about 1.5 meters away from the 3D Digitizer. On the floor, just below the seat, yellow straps are placed according to rotation angles.*

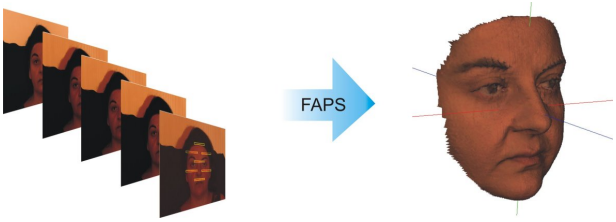

Figure 12: *3D model reconstruction. On the left are the raw data containing texture and 4 frames with shifted fringe pattern. The last picture shows parallax points. After processing by Inspeck Faps software we obtain the 3D model.*

vice's jargon) are carried out to obtain the 3D information (Fig. 13). In the PreProcessing stage, first the phase function is calculated (top right image in Fig. 13). Consequently the depth information is measured by using some reference points, called parallax points (top left image in Fig. 13). Most of the time we allowed the Inspeck software to detect these parallax points automatically; however, we reverted to manual marking in case of doubt. Next, an Interest Area is selected for the face image. We set this zone manually by creating one or more polygons that define it (top middle image in Fig. 13). A clearly defined Interest Area is very helpful to remove the background clutter. Thus we can cancel out extra streaks, which come from the digitizing of the background area.

After segmenting the face, phase unwrapping operation is performed to obtain 3D data in the Processing stage. However, during phase unwrapping some discontinuity errors, which are erroneous depth levels associated with discontinuous surfaces, may occur. The verified discontinuities are fixed by sequential push or pull operations.

Finally in the Post-Processing stage, the so-far processed data, currently in machine units, is converted to a data set in geometric parameters (millimeters). To convert the data into milimeters we determine a reference from a set of candidate reference (parallax) points. Generally, the point which has the highest confidence value is selected and the geometric parameters of the other points in the image are determined accordingly. Also, in this stage we sometimes perform simple filtering operations to correct slight optical distortions. Whenever serious distortions occurred we preferred to re-capture the data.

The result is exported in the form of ".net" file, which is a specific file format for the data-processing software of Inspeck Inc. [6]. The output file stores the 3D coordinates of the vertices together with color information as a related separate texture output file in ".bmp" format.

Since there might be problems for some of the scans, FAPS preprocessing has been done simultaneously with data acquisition. If a problem is observed on the data then that scan is repeated on the spot. Some commonly occurring problems are as follows (Fig. 14):

- *Movement during acquisition:* The Inspeck hardware is adjusted to capture four pictures that later constitute the raw data for 3D face images and one more picture of texture image. During acquisition, the subject must stay still (better hold the breath), so that the pictures are consistent with each other. Otherwise the wavelike defects occur globally around the face. In this case the acquisition is repeated.
- *Hair occlusion:* Fringes projected on hair becomes undetectable, hence the corresponding parts on the 3D image is noisy. Hair occluded parts are included in the face segmentation, since avoiding them is not possible. However, with smoothing operation we can obtain less noisy data.
- *Hand occlusion and open mouth:* In the case of hand occlusion and open mouth, depth level changes sharply over the edges, causing a discontinuous fringe pattern. In the 3D image, hand and tongue are constructed at wrong depth levels.
- *Facial hair and eyes:* Beards, eyebrows and eyelashes cause small fluctuations. Also eyes can not be kept constant because of high light intensity, resulting in a spiky surface on the eyes. These problems are ignored, not additional filtering processes are applied.
- *Eyeglasses:* Fringe pattern can not be reflected back from transparent surfaces. Eyeglasses seem to deform the area they cover on the face. The resulting noisy surface is included in the face segmentation.

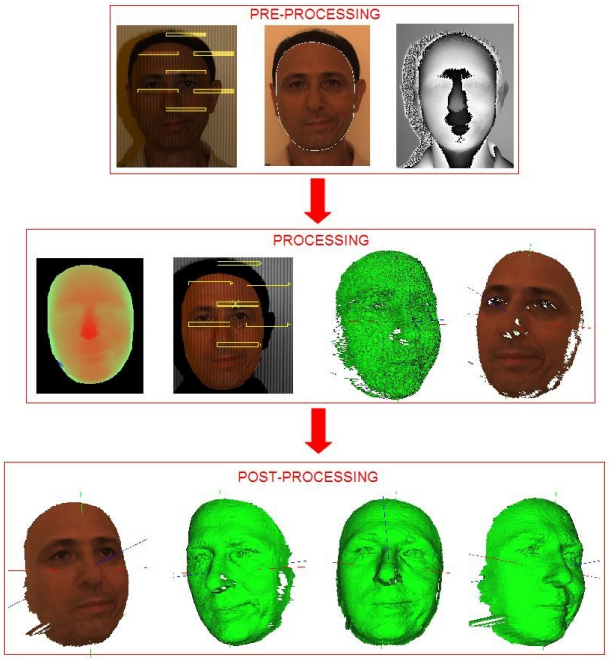

Figure 13: *Reconstruction by FAPS. Pre-processing (top row), processing (middle row) and post-processing (bottom row) stages. Top row: Parallax points, polygon to define the region of interest, phase image. Middle row: Depth image, parallax points, 3D shape, 3D texture. Bottom row: A scanned phase from different view points after post-processing such as hole filling.*

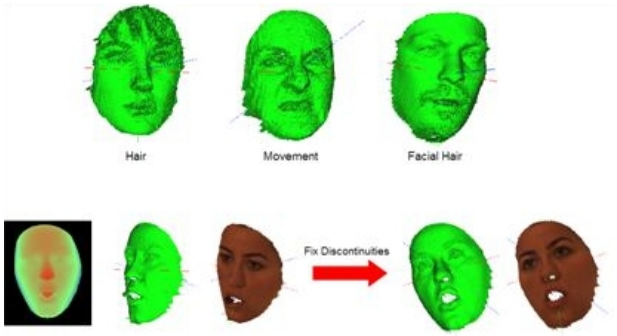

Figure 14: *Commonly occurring problems during image acquisition and face reconstruction. At top of the figure, noise due to hair, movement, and facial hair is seen on the face scans. At the bottom left, a mistake in the depth level of the tongue, and at the right, its correction is displayed.*

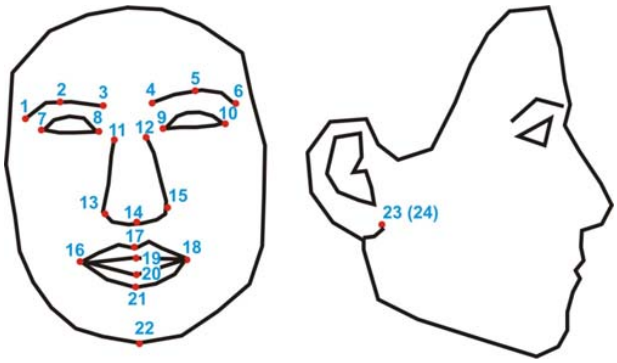

Figure 15: *Manually marked landmark points.*

5. MANUAL DATA LANDMARKING

All of our face images are landmarked manually in order to be used in various training and testing methods as ground truth. On each face scan, 24 points are marked provided that they are visible in the given scan. The landmark points are listed in Table 4 and shown in Fig. 15.

1. Outer left eye brow	2. Middle of the left eye brow
3. Inner left eye brow	4. Inner right eye brow
5. Middle of the right eye brow	6. Outer right eye brow
7. Outer left eye corner	8. Inner left eye corner
9. Inner right eye corner	10. Outer right eye corner
11. Nose saddle left	12. Nose saddle right
13. Left nose peak	14. Nose tip
15. Right nose peak	16. Left mouth corner
17. Upper lip outer middle	18. Right mouth corner
19. Upper lip inner middle	20. Lower lip inner middle
21. Lower lip outer middle	22. Chin middle
23. Left ear lobe	24. Right ear lobe

Table 4: *Landmark points manually labeled on each scan.*

Landmarking software is implemented as a MATLAB GUI. This software is designed so that all landmark points can be marked guided by a reference template image for the landmarks in Fig. 15. This scheme helps us avoid errors which can occur when marking many landmark points over lots of images. The 3D coordinates corresponding to the 2D marked points on the texture image is extracted automatically using this software. Here we make the assumption that 2D image and 3D data are registered although they are not 100% registered.

In the end of manual landmarking, we obtain 2D and 3D locations of 24 facial fiducial points for every scan of every subject. The landmark data structure is stored in a MATLAB "mat" file, one for each subject. The same data is also saved in "text" file format in order not to force the users to MATLAB.

6. AUTOMATIC LANDMARKING

Face detection, normalization or registration should be performed before any face recognition algorithm. ICP (Iterative Closest Point) [10] and TPS (Thin Plate Spline) [11] are the mostly used 3D to 3D registration methods. However, they perform well only when the models being registered are very close to each other since these methods are sensitive to initialization conditions. Thus, a rough registration should be performed before the fine registration by ICP or TPS methods.

Colbry, Stockman and Jain [12] proposed a rough registration method based on 3D anchor points detected from the face data before application of ICP. They detected facial features based on their shape indexes, intensity values and spatial relations where they used both 3D shape and 2D image outputs of the scanner in order to locate anchor points. Similarly, Lu and Jain [13], Chang et. al. [14] and Boehnen and Russ [15] used 3D and 2D information together in order to extract facial anchor points. In most of these methods, detection usually starts from a single point and then followed by an order of other points. For example in [13], first nose was needed to be detected so that the eye and mouth corners could be detected next. Nose tip was detected as follows: First, pose was quantized into some number of angles. Then, the point with the maximum projection

value along the corresponding pose direction was found. Finally, based on the nose profile, the best candidate for a nose tip was selected. After locating the nose tip, other points are located by considering the spatial relationships among the points, shape indexes for these points obtained from 3D data and cornerness for these points determined from the intensity image. In [14], eye cavities were detected first by investigating the mean and Gaussian curvature values and by using some heuristics. Then, nose tip was detected as being a peak which has a spatial relationship with the eye pits. In [12], facial points are detected based on their mean and Gaussian curvature values independently and, again, to resolve between multiple detections, spatial relations between the anchor points were considered. This was done by a relaxation algorithm where some rules were used for anchor points and relations between them in order to eliminate incorrect sets of anchor points.

In this project we applied three different methods for automatic landmarking. There are 24 goal landmark points each of which is marked manually on the ENTERFACE07 database. Although seven of the landmark points are considered to be the most fiducial ones, each method deals with its own set of points. The three competitor methods are briefly described below and their test results are presented in Section 8 (Testing and Evaluation Section).

6.1. Statistical Automatic Landmarking (SAL)

This method exploits the face depth maps obtained by 3D scans. In the training stage, an incremental algorithm is used to model patches around facial features as mixture models. These mixture models are, in turn, used to derive likelihoods for landmarks during testing [16]. The likelihood tests yield the seven fiducial feature points (inner and outer eye corners, nose tip and mouth corners) and the remaining 15 landmarks are estimated based on the initial seven fiducial ones by back-projection. For enhanced reliability, the seven landmarks are first estimated on a coarse scale, and then refined using a local search [17]. Furthermore, incorrect localizations are detected and fixed with the GOLLUM Algorithm [18]. GOLLUM Algorithm uses smaller number of points to check all point localizations. The assumption is that these points contain statistically discriminatory local information, and can be independently localized. An example result is shown in Fig. 16.

In a second tier application of the GOLLUM algorithm, the locations of 15 additional feature points are estimated. These points are based on the coordinates of the seven fiducial ones. The initial seven points contain richer statistically discriminatory local information and can thus be independently localized, whereas the remaining 15 points have much weaker image evidence around them, hence they need the structural information of the former ones. In Fig. 17 all of these landmark points are shown.

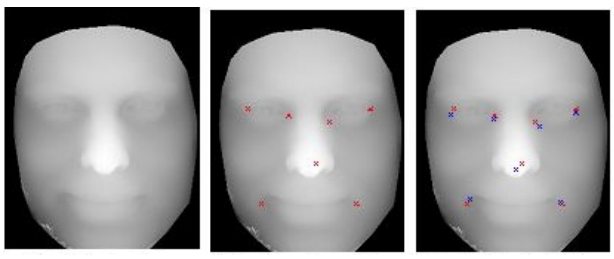

Figure 16: *Depth map of the face image (a), statistically located feature points (b) and corrected feature points on the face (c).*

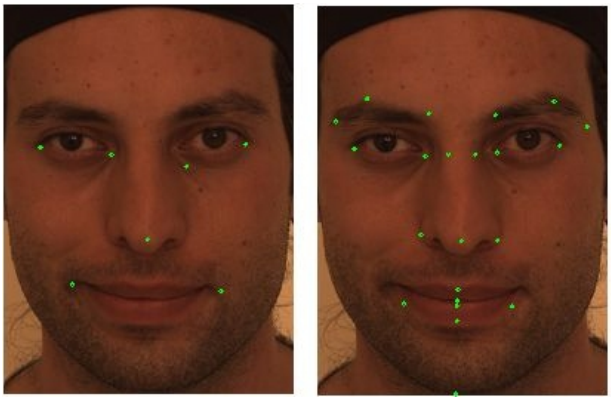

Figure 17: *A total of 22 feature points is landmarked (b) by using automatically landmarked 7 feature points (a).*

6.2. Robust 2D Landmarking (RL)

This method is quite similar to the method described in previous subsection, in that it is also a two-tier method which captures first the seven fiducial landmarks. It also proceeds from an initial coarse evaluation on 80x60 images to a refined version of 640x480 resolution images. The fiducial ones are the four eye corners, the tip of the nose, and the two mouth corners [2]. This method differs from the one in previous subsection as follows: we use templates for each landmark consisting of a subset of DCT coefficients and we obtain scores (likelihoods) employing SVM classifiers. In the coarse localization, 8x8 DCT blocks are extracted and trained separately for each facial landmark. We permit initially four candidates corresponding to the highest peaks in the matching score map of each landmark. This is in order not to miss any feature points. The 4x7 = 28 landmark candidates are reduced to a final seven based on an exhaustive graph search technique. Two methods are proposed for this task: Probabilistic Graph Model-I and Probabilistic Graph Model-II. In PGM-I, we first find a subset of the most reliable landmarks based on the anthropomorphic data learned during a training session and then estimate the position of the missing landmarks. On the other hand, PGM-II is designed for simpler scenarios, without variety in illumination and poses. It systematically searches for the seven feature landmark among the four peaks of the similarity score map. Facial feature landmarks are determined according to companion features, called the support set for that feature. Once the coarse-stage landmarks are located we proceed with the refinement stage on the original high-resolution image. The search proceeds with a 21x21 window from around the coarse localization points. Each configuration formed by reliable landmarks is set to origin, scaled to a fixed length and rotated. Since this procedure makes the configuration independent of the pose and scale of the face, this algorithm achieves high accuracy even under various poses and for this reason is called robust 2D landmarking. In our future work, we aim to apply this method to 3D scans and extend our algorithm to cover the remaining 15 features as the method in the previous subsection.

6.3. Transformation Invariant 3D Feature Detection (TIFD)

In transform and scale invariant feature detection algorithm, 3D fundamental elements such as peaks, pits and saddles are extracted with their scale information from the 3D scan data. With a predefined graph structure based on a group of fundamental elements and their special relationships, we can construct a topology of these elements which could then be used to define the

object. And this graph structure can be used for object detection, registration and recognition purposes.

The method finds the mean (H) and Gaussian (K) curvatures [10, 13] on the surface. Afterwards using Gaussian pyramiding, the H and K values are computed for the higher scales. In all of the scale levels, fundamental elements (peaks, pits and saddles) are found using H and K values as in [10, 13]. In Fig. 18, the higher levels of these fundamental elements are depicted with color (blue: peak, cyan: pit, red: saddle ridge, yellow: saddle valley) and are shown on the 3D surface. We name this 3D volume a "UVS space" where "u" and "v" are surface coordinates and "s" is for scale. By extracting the connected components inside this UVS volume, we obtain the fundamental elements with their position and scale information. The details of the algorithm are given in [19] and [20].

For the ENTERFACE'07 project, we have created a 3D topology for the human face as formed by one peak, two pits and one saddle which represent the nose, the eye pits and the nose saddle, respectively. A four node graphical model is constructed where each fundamental element is a node in the graph and spatial relationships between the fundamental elements are carried by the edges between the nodes. The spatial relations between a node couples are the position difference between the nodes normalized by scale, normal difference between the nodes and scale difference. These values are modeled statistically from the training scans. Then during testing, a graph search is accomplished on the topology obtained from the test scans. As the topology for a human face is formed as having two eye pits, one nose saddle and one nose peak, among the 1peak-2pits-1saddle combinations, the most probable combination is selected as the correct model.

Thus, using this restricted topology, we can detect two inner eye corners, middle of nose saddle and top of the nos. These differ from the labeled landmark points but the latter can be extrapolated easily. Since the location of face is found in the 3D scan data as well as landmark points, pose estimation and registration can be performed easily afterwards.

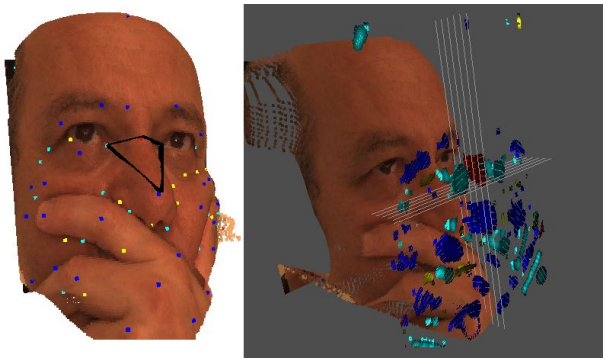

Figure 18: *Left: 3D surface structure centers colored by their types (blue: peak, cyan: pit, red: saddle ridge, yellow: saddle valley) and the graph representing the face, Right: UVS volume shown over the 3D data. Each level represents a layer in the scale space. Connected components are found inside this volume in order to extract the fundamental elements, namely peaks (blue), pits (cyan), saddle ridges(red) and saddle valleys (yellow).*

7. AFM BASED ICP REGISTRATION

In any 3D face processing, be it face recognition or expression understanding, the registration of faces is a crucial step. ICP (Iterative Closest Point) [10] and TPS (Thin Plate Spline) [11] are the most frequently used 3D-to-3D registration methods. However, these methods perform satisfactorily only when the models to be registered are very close to each other since these methods are initialization dependent. Thus, a coarse registration is a must before any further fine registration via ICP or TPS methods. And this coarse registration is usually obtained by automatic landmarking defined in the previous section.

Even though Iterative Closest Point algorithm has a high computational cost, its ease of implementation and accurate end-results make it a preferred method in the registration of 3D data. The ICP method finds a dense correspondence between a test surface and a gallery surface by transforming the test surface to fit the gallery while minimizing the mean square error [10]. The transformation is rigid, consisting of a translation and a rotation. Because this transform with six degrees of freedom is nonlinear, the error minimization problem has to be solved by means of iterative methods.

The most accurate results obtained so far in the literature, make use of a one-to-all registration in which a test shape is aligned separately to each of the faces in the gallery and the test face is compared to each gallery face separately. To overcome the computational cost of one-to-all registration, in [21] it was proposed to use an average face model (AFM). In the registration with AFM, a single registration of a test face is adequate to align it with all the pre-registered faces in the gallery. Motivated by this work on registration via AFM, in [22] it was proposed to use multiple AFMs for different facial categories to increase the accuracy of registration. Multiple AFMs can be achieved either by manually grouping the faces into different categories or by clustering them into classes in an unsupervised manner. The latter approach is a more accurate one, in which the discriminating facial properties existing in the dataset are acquired automatically. Although this method was not used during eNTERFACE'07, it is particularly suited to the dataset because of its many inherent variations. We will apply it in the future.

8. TESTING AND EVALUATION

The database of 3D face scans have been processed with conditioning operations. In addition, they contain ground-truth landmarks; hence they can potentially be used for testing in various tasks such as face detection, automatic facial feature detection, registration, face recognition and facial expression estimation. During the eNTEFACE'07 project we have limited ourselves to testing automatic landmarking and registration algorithms due to time limitations. Work on other tasks will be completed in the near future.

8.1. Landmarking Tests

For the comparative evaluation of landmark localizers, we have used a normalized distance. We divided the localization error, measured as Euclidean distance between the manually marked and automatically estimated landmark positions by the inter-ocular distance. A landmark is considered correctly detected if its deviation from the true landmark position is less than a given threshold, called the acceptance threshold. This threshold itself is typically a percentage of the inter-ocular distance. Curves obtained in this manner are used to present the success of the algorithms (Fig. 19, 21 and 22).

For automatic landmarking tests, we chose 66 subjects that possessed multiple frontal and neutral poses. The same subjects were used for both training and testing. More specifically, we trained our algorithms using one neutral frontal pose from each subject, and then tested it using alternate neutral frontal poses,

which were not used for training. In addition, we also tested the landmarking on three difficult categories: one from action units (open mouth, AU27), one from rotations (+20° yaw) and one from occlusion scans (mouth occlusion).

8.1.1. Statistical Automatic Landmarking (SAL)

In the training step of Statistical Automatic Landmarking, neutral poses of 40 subjects, in our 3D Face & Expression Database, have been used. To evaluate the success of the Statistical Automatic Landmarking algorithm, a test set has been prepared. The testing set consists of:

- 20 subjects for neutral pose
- 20 subjects for neutral pose with glasses,
- 20 subjects for neural pose with eye or mouth occlusion.

According to test results, it is seen that results with the neutral pose with occlusion is better (Fig. 19). Because of the statistical nature of the algorithm, initial landmarking step is processed similarly for all poses. In other words, occlusion in the pose does not affect statistical initial landmarking. After statistical initial landmarking, fine localization process corrects the misplaced feature points by using the depth map. In this step, a pose with occlusion can be corrected more efficiently than a neutral pose. Because there are several local minimas around the mouth in the neutral pose and this decreases the efficiency of the fine localization algorithm. Namely, mouth occlusion (by hand) provides to get rid off local minimas around mouth. As a result, we can say that if the pose and the pose angles are stable, this approach gives good results for the poses with occlusions.

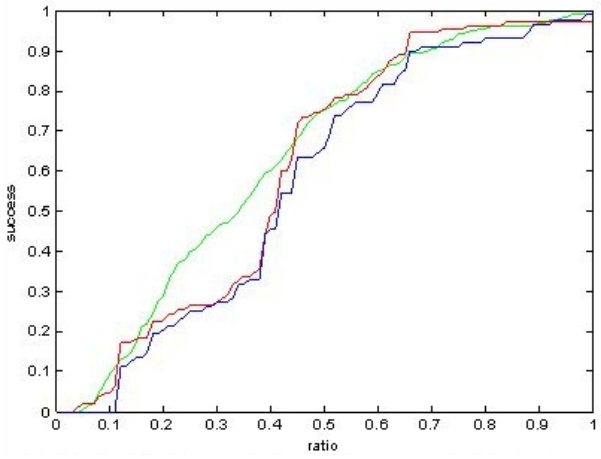

Figure 19: *Statistical Automatic Landmarking test results. Results for neutral pose, with mouth occlusion and with eye glasses occlusion is shown in as green, red and blue curves respectively.*

8.1.2. Robust 2D Landmarking (RL)

In this section, we presented results of our automatic landmarking algorithm tested on eNTERFACE Database. We have used 68 samples with neutral poses as training set. For each sample, DCT coefficients of manually landmarked points were computed and then, we trained an SVM classifier with these coefficients. In order to verify the robustness of proposed algorithm, we tested our classifier on samples with different poses. The testing set consists of:

- 12 subjects for rotation (+10° yaw rotation),
- 28 subjects for facial action unit (mouth stretch),
- 28 subjects for eye or mouth occlusion.

In Fig. 20, output of the algorithm is shown for some representing poses. As expected, the best results are obtained for neutral poses. In Fig. 21, performance of the feature localizer versus acceptance threshold is given for three testing set. According to results, the proposed method is more robust to facial expressions and occlusion rather than rotation.

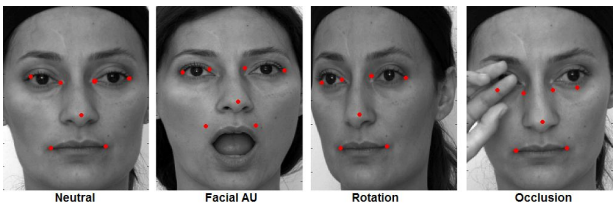

Figure 20: *Landmarking outcomes of the automatic feature localizer for neutral pose, yaw, rotation, mouth stretch and eye occlusion, respectively from left to right.*

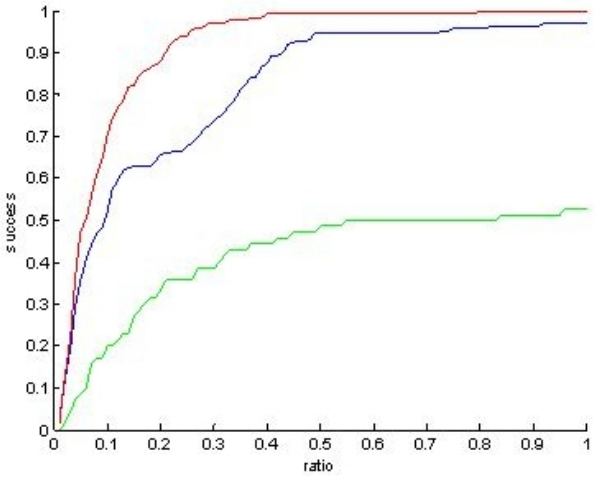

Figure 21: *Performance of 2D Landmarking. Red: pose with expression, Blue: mouth or eye occlusion, Green: rotation.*

8.1.3. Transformation Invariant 3D Feature Detection (TIFD)

In order to test the performance of the TIFD method on the eNTERFACE'07 scans we have made some of experiments. Beforehand, we have trained the system using 66 neutral scans taken from 66 different people. By training, we mean extraction of a statistical model for the relations between the nodes, namely the nose peak, the nose saddle and the eye pits by training a Gaussian Mixture Model [20]. During testing, the method searches for a quarted of a peak, a saddle and two pits which are statistically similar in terms of relative scale and relative orientation to the trained model of neutral scans.

For testing, we have used three groups of scans. The first group included the scans with the expression of mouth opening. Since opening the mouth does not change the locations of the four nodes or does not occlude them, the success rates were very high. We have used 25 models of different people having this expression. In Fig. 22, red curve depicts the success rate for this experiment. By success rate, we mean the average success among all feature points. If the difference between the four points of the detected quartet and their four originally marked points are below a threshold, the success is "one" for that model;

otherwise it is zero. The success rate is calculated among 25 models as an average and taking different error thresholds for success, the curve is depicted.

The second model group included the faces with poses where faces are rotated by 20° around the y-axis (yaw rotation). In this case, the right eye pit gets occluded. However TIFD algorithm still detects a pit near that region, but as center of the eye pit is occluded, the localization slightly fails. As expected, the success rates for the posed faces deteriorate due to this phenomenon. We have used 24 posed models of different people. In Fig. 22, green curve depicts the success rate for this experiment.

Finally we have tested the algorithm on mouth occlusion. Mouth occlusion profoundly distorts the geometry of the facial surface. On the other hand, since we have searched for the nose peak, nose saddle and the eye pits and since these regions are not occluded; the success rates for landmarking on these types of scans were very high. We have used 19 mouth occluded models of different people. In Fig. 22, blue curve depicts the success rate for this experiment.

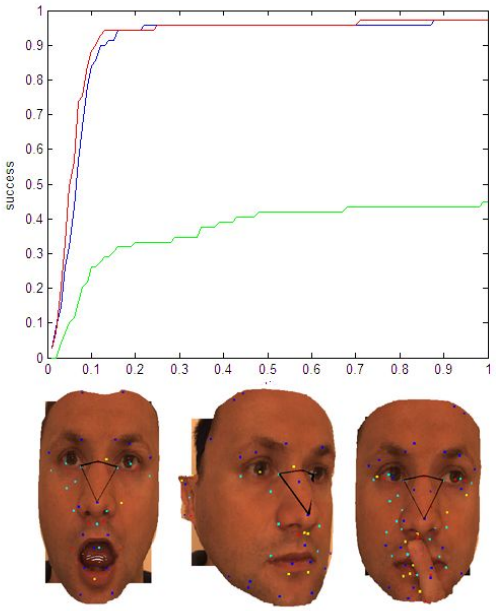

Figure 22: *Performance 3D feature detection results with (TIFD). Plot red: action unit (open mouth, AU27), green: one from rotations (+20° yaw) and one from blue: occlusion scans (mouth occlusion).*

8.2. Registration Tests

The ICP registration based on the use of an AFM can be summarized as in Fig. 23, where a test face is registered to an AFM before a comparison with the gallery faces that have already been aligned with the model.

The depth image of the AFM generated from the first neutral samples belonging to each subject is given in Fig. 24 (a). The construction algorithm proposed by [22] is based on initial alignment by Procrustes analysis and fine registration by TPS warping.

To explore the effect of pose and facial action variations on 3D face recognition, four different face subsets were used for registration and classification simulations. One subset consisted of faces with a pose variation of 10 degrees. Another one was

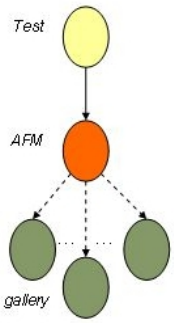

Figure 23: *The AFM-based registration approach.*

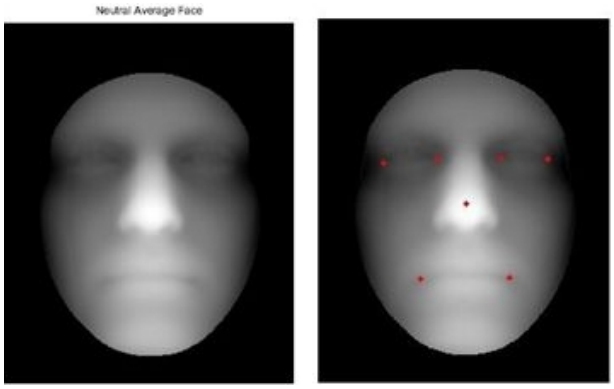

Figure 24: *The AFM constructed from neutral samples of FR database. (a) The depth image for the AFM, (b) the AFM with the calculated landmarks. These landmarks will be used for initialization in the registration phase.*

the subset of faces with "jaw drop" facial action unit. The other face subset had faces with eye(s) occluded. To compare the effects of these different facial variations, a subset of neutral faces was also used in simulations. The gallery set consisted also of neutral faces (but different samples from the neutral test subset). The set of landmarks used for the initial registration by Procrustes analysis can be seen in Fig. 24 (b). The initialization of faces with eye occlusion was handled with fewer landmark points that existed in each face (excluding the eye corners for the eye that was occluded by hand). The seven landmarks (inner and outer eye corners, the nose tip and the mouth corners) were also used for initialization before TPS warping that is used for AFM construction.

Faces in each test subset and the gallery were registered to the AFM by ICP. The registered faces were also cropped according to the average model. A classification simulation was performed based on point set distance (PSD), which is an approximation to the volume difference between faces after registration. Each test face was classified as the closest gallery face. The recognition rates for various ranks are plotted in Fig. 25. In Table 5, the rank-1 recognition rates for each face subset are given.

It is claimed that by using 3D data instead of 2D, pose variations can be eliminated. The registration is the key method to overcome the problems that will arise from pose changes. As seen from these results, a small pose variation can be eliminated by ICP registration. The slight drop in recognition performance is due to more noisy data when acquiring data with pose rotation: The nose occludes one side of the face; and the resulting holes are patched by post-processing, resulting in noisy data.

In the case of eye occlusion, the initial alignment was handled with fewer landmark points that were available in each face. Nevertheless the registration was successful. The low recognition rates were caused by the use of PSD for classification. It would be more appropriate to use a parts-based representation that eliminates the affect of the occluding hand. In the case of mouth opening, the registration was affected a little, while the movement of landmarks belonging to the mouth area also caused a transform for the whole face. Also the rigid structure of ICP is effective in this case. The test face will be aligned as much as possible to the AFM, but the open-mouth form will be kept. Therefore in PSD calculations, the mouth area will augment the distance values, causing the ill-effect on classification results. As in occlusion, a parts-based representation would have been more appropriate.

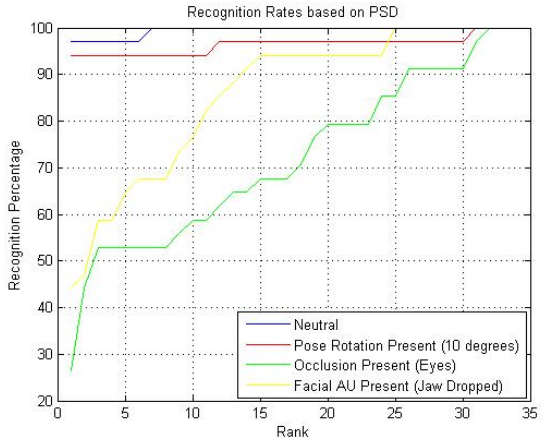

Figure 25: *The recognition rates for various ranks and for different subset simulations.*

	Neutral	Pose Rotation (10°)	Occlusion (Eye)	Facial Action (Jaw Drop)
Rank-1	97.06	94.12	26.47	44.12
EER	5.75	8.73	44.52	23.84

Table 5: *The Rank-1 recognition rates and EER values for each face subset*

9. CONCLUSION AND FUTURE WORK

As future work, we would like to investigate automatic methods for refining and validating the quality of hand-labeled landmarks. As with all human-produced annotations, the landmarks positioned by the labelers are subject to variation. This raises the issue of inter-rater agreement across different raters. We would like to perform appropriate inter-rater comparison and develop automated methods for refining rater input.

The 3D database created during this month-long project will be a very valuable resource for our future work on expression analysis. Including many AUs from FACS and also the six emotions, this database encompasses a rich set of facial deformations. Existing research has focused on recognizing prototypical expressions (e.g., the so called six basic emotions). Poses containing various action unit combinations will help us extend the scope of automatic facial expression analysis. As future work, we intend to explore how wider range of expressions can be recognized using 3D surface data.

The present face database will be put to three different tests:

- 2D-3D Expression Understandings: 3D information can also be combined with the corresponding 2D texture to improve recognition. Moreover, this database can be very useful for developing algorithms that use 2D images for expression recognition. Most existing 2D systems perform well with frontal face images but fail under considerable out of plane head rotations. We would like to find ways of exploiting 3D information to develop pose independent systems. Consequently, we expect to advance the state of art in automatic expression recognition using our database.
- Automatic 2D-3D Landmarking: Facial feature point localization is an intermediate step for registration, face recognition and facial expression analysis. Since correct localization directly affects the performance of the face analysis algorithms, landmarking should be robust with respect to changes in illumination, pose and occlusions. Although there have been advances in automatic 2D/3D landmarking algorithms, there are still open problems. For example, existing algorithms generally work well with frontal and neutral poses; but fail in poses with rotation and facial expressions. We envision the 3D database to be a useful resource in the course of developing and testing 3D-aided 2D or 3D automatic landmarking algorithms that can deal with variations due to pose, illumination and occlusions.
- Registration: This database; including different poses, expressions and occlusions for each subject, is a good test case for registration of faces. The facial surfaces corresponding to different facial groups can be tested to examine the effect of registration of each facial surface difference. Also the multiple AFM-based registration approach from [22] can be adapted to the different groups in this database.

10. ACKNOWLEDGEMENTS

We would like to thank many participants of eNTERFACE'07 Workshop, actors and others who voluntarily let their faces to be scanned. We appreciate their help and patience, since they spent their valuable time during data acquisition process which takes at least half an hour. We would like to thank Niyazi Ölmez for his invaluable help. Without him, we could not find the opportunity of arranging many actors for our database collection. Also, his recommendations for our data capturing setup were very useful, and his personal assistance during acquisition was very valuable. Finally, we would like to thank organizers of eNTERFACE'07 for making collection of such a database possible.

The acquisition hardware in this project was made available through TUBITAK grant 104E080: 3D Face Recognition.

11. REFERENCES

[1] F. L. Bookstein, "The measurement of biological shape and shape change", *Lecture Notes Biomathematics*, vol. 24, 1978. 88

[2] H. Çınar Akakın and B. Sankur, "Robust 2D/3D Face Landmarking", tech. rep., 3DTV CON, 2007. 88, 94

[3] J. Wang, L. Yin, X. Wei, and Y. Sun, "3D Facial Expression Recognition Based on Primitive Surface Feature Distribution", in *IEEE International Conference on Computer*

Vision and Pattern Recognition (CVPR 2006), (New York, NY), IEEE Computer Society, June 17-22 2006. 88

[4] L. Yin, X. Wei, Y. Sun, J. Wang, and M. J. Rosato, "A 3D Facial Expression Database For Facial Behavior Research", in *7th International Conference on Automatic Face and Gesture Recognition (FGR06)*, pp. 211–216, April 10-12 2006. 88

[5] P. Ekman and W. V. Friesen, *Facial Action Coding System: A Technique for the Measurement of Facial Movement*. Palo Alto: Consulting Psychologists Press, 1978. 88

[6] "InSpeck". http://www.inspeck.com. 88, 91, 92

[7] P. Ekman and W. V. Friesen, "Constants Across Cultures in the Face and Emotion", *Journal of Personality and Social Psychology*, vol. 17, no. 2, pp. 124–129, 1971. 88

[8] C. Wallraven, D. Cunningham, M. Breidt, and H. Bülthoff, "View dependence of complex versus simple facial motions", in *Proceedings of the First Symposium on Applied Perception in Graphics and Visualization* (H. H. Bülthoff and H. Rushmeier, eds.), vol. 181, ACM SIGGRAPH, 2004. 88

[9] M. Kleiner, C. Wallraven, and H. Bülthoff, "The MPI VideoLab - A system for high quality synchronous recording of video and audio from multiple viewpoints", Tech. Rep. 123, MPI-Technical Reports, May 2004. 88

[10] 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, 1992. 93, 95

[11] F. L. Bookstein, "Principal warps: thin-plate splines and the decomposition of deformations", *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 11, pp. 567–585, 1989. 93, 95

[12] D. Colbry, G. Stockman, and A. Jain, "Detection of Anchor Points for 3D Face verification", in *Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2005. 93, 94

[13] X. Lu and A. K. Jain, "Automatic feature extraction for multiview 3D face recognition", in *Proc. 7th International Conference on Automated Face and Gesture Recognition*, 2006. 93, 95

[14] K. Chang, K. W. Bowyer, and P. J. Flynn, "Multiple nose region matching for 3D face recognition under varying facial expression", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, no. 10, 2006. 93, 94

[15] C. Boehnen and T. Russ, "A fast multi-modal approach to facial feature detection", in *Proc. 7. th. IEEE Workshop on Applications of Computer Vision*, pp. 135–142, 2005. 93

[16] A. Salah and E. Alpaydin, "Incremental Mixtures of Factor Analyzers", in *Int. Conf. on Pattern Recognition*, vol. 1, pp. 276–279, 2004. 94

[17] A. A. Salah, R. J. Tena, M. Hamouz, and L. Akarun, "Fully Automatic Dense Registration of 3D Faces: A Comparison of Paradigms", *submitted to IEEE Trans. PAMI*, 2007. 94

[18] A. A. Salah, H. Çınar, L. Akarun, and B. Sankur, "Robust Facial Landmarking for Registration", *Annals of Telecommunications*, vol. 62, pp. 1608–1633, January 2007. 94

[19] E. Akagündüz and İlkay Ulusoy, "Extraction of 3D Transform and Scale Invariant Patches from Range Scans", in *2nd Beyond Patches Workshop "Patches Everywhere" Workshop in conjunction with CVPR 2007*, 2007. 95

[20] E. Akagündüz and İlkay Ulusoy, "3D Object Representation Using Transform and Scale Invariant 3D Features", in *Workshop on 3D Representation for Recognition (3dRR-07), ICCV 2007 (Accepted)*, 2007. 95, 96

[21] M. O. Irfanoğlu, B. Gökberk, and L. Akarun, "3D Shape-Based Face Recognition Using Automatically Registered Facial Surfaces", in *Proc. Inf. Conf. on Pattern Recognition*, vol. 4, pp. 183–186, 2004. 95

[22] A. A. Salah, N. Alyüz, and L. Akarun, "Alternative face models for 3D face registration", in *SPIE Conf. on Electronic Imaging, Vision Geometry*, (San Jose), 2007. 95, 97, 98

12. APPENDICES

12.1. Database design

- Database:
  (Since the whole system may be composed of different databases we'll have a structure called database.)
  - DBID: Primary Key that will be given by system
  - DBName Folder - Mapping to the initial folder structure of the database
  - StartDate - If the different databases evolve this may be needed for date retrievals
  - EndDate
  - SessionNumber
- Subject:
  - SubjectID: Primary Key that'll be given by system
  - Name:
  - Middle Name:
  - Surname:
  - Birth date:
  - Gender:
  - Profession: Actor or a normal human
  - Approval for Publication:
- SessionList:
  - List ID: Primary key.
  - List of the sessions of the database. It will be situated under the database folder and will hold all the session list structure necessary to undertake session architectures.
  - Folder record path of related session.
- Session:
  - SessionID: Primary Key that'll be given by system
  - SubjectID: Refers to the subject table. (referential integrity a subject have many sessions / 1-n mapping)
  - Facial Hair: Yes / No If facial hair then Facial Hair situation Beard, Mustache, beard + mustache Facial Hair Degree Degree: 1 (Low) - 2 (Normal) - 3 (very Much)
  - Hair: 0 (No hair) - 1 (Normal) - 2 (Very Much)
  - Earing: Yes/No
  - Session Folder: Maps to the version folder where all the session data are located.

  - SessionDate
  - SessionCode
  - RecordFolder: Name of the folder holding the records of the session.
- Action Unit:
  - AUID - primary key
  - Code - Code of the action unit
  - Explanation
  - Image Path
  - Video Path
- Records:
  - RecordID: primary key
  - SessionID Referential integrity for finding the owner session.
  - AUID (The ID of the record that comes from Action Unit Table)
  - recordfile name of the corresponding record file
  - Applicable or Not? The stored 3D image is not proper.

12.2. XML Based Subject Table Example

```
<?xml version="1.0" standalone="yes"?>
<NewDataSet>
  <xs:schema id="NewDataSet" xmlns="" xmlns:xs="http
      ://www.w3.org/2001/XMLSchema" xmlns:msdata="urn
      :schemas-microsoft-com:xml-msdata">
  <xs:element name="NewDataSet" msdata:IsDataSet="true
      " msdata:UseCurrentLocale="true">
  <xs:complexType>
  <xs:choice minOccurs="0" maxOccurs="unbounded">
  <xs:element name="Subject">
  <xs:complexType>
  <xs:sequence>
  <xs:element name="ID" type="xs:short" minOccurs="0"
      />
  <xs:element name="Code" type="xs:string" minOccurs
      ="0" />
  <xs:element name="Name" type="xs:string" minOccurs
      ="0" />
  <xs:element name="SurName" type="xs:string"
      minOccurs="0" />
  <xs:element name="MiddleName" type="xs:string"
      minOccurs="0" />
  <xs:element name="BirthDate" type="xs:dateTime"
      minOccurs="0" />
  <xs:element name="Gender" type="xs:short" minOccurs
      ="0" />
  <xs:element name="Profession" type="xs:short"
      minOccurs="0" />
  <xs:element name="ApprovalCondition" type="xs:
      boolean" minOccurs="0" />
  </xs:sequence>
  </xs:complexType>
  </xs:element>
  </xs:choice>
  </xs:complexType>
  </xs:element>
  </xs:schema>
  <Subject>
    <ID>0</ID>
    <Code>S00001</Code>
    <Name>Lale</Name>
    <SurName>Akarun</SurName>
    <MiddleName />
    <BirthDate>2006-02-16T00:00:00+02:00</BirthDate>
    <Gender>0</Gender>
    <Profession>1</Profession>
    <ApprovalCondition>true</ApprovalCondition>
  </Subject>
  <Subject>
    <ID>1</ID>
    <Code>S00002</Code>
    <Name>Ilkay</Name>
    <SurName>Ulusoy</SurName>
    <MiddleName />
    <BirthDate>2006-02-16T00:00:00+02:00</BirthDate>
    <Gender>0</Gender>
    <Profession>0</Profession>
    <ApprovalCondition>true</ApprovalCondition>
  </Subject>
</NewDataSet>
```

13. BIOGRAPHIES

Arman Savran was born in Turkey, in 1981. He received the B.Sc. degree in Electronic and Communication Engineering from the İstanbul Technical University, İstanbul, Turkey, in 2002, and the M.Sc. degree in Electrical and Electronics Engineering from the Boğaziçi University, İstanbul, Turkey, in 2004. He worked at Sestek Inc., İstanbul, Turkey, between 2004 and 2006, where he did research on speech and text driven synthesis of facial animation. He is currently a Ph.D student at Electrical and Electronics Engineering Department of Boğaziçi University, BUSIM Laboratory, and. his research work is on analysis of facial expressions and 3D face modeling.
Email: arman.savran@boun.edu.tr

Oya Çeliktutan received the B.E degree in Electronic Engineering from Uludağ University, Bursa, Turkey, in 2005. She is now a teaching assistant and pursuing the M.S. degree in signal processing and communications at the department Electrical and Electronics Engineering at Boğaziçi University, İstanbul, Turkey. Her research interests include multimedia forensics, machine learning and pattern recognition, computer vision and image processing. Currently, she is working on automatic 2D/3D facial feature extraction.
Email: oyaxceliktutan@yahoo.com

Aydın Akyol was born in Afyon, Turkey in 1979. He received his B.Eng. degree at İstanbul Technical University, Computer Engineering Department, in 2001 and MSc. degree at Sabanci University, Computer Science Department, in 2003. After his graduation he worked as an engineer at R&D department of Verifone Inc. for 3 years. Currently he is pursuing the Ph.D. degree at İstanbul Technical University Vision Laboratory. His research interests include inverse problems in Computer Vision and Image Processing.
Email: akyol@su.sabanciuniv.edu

Jana Trojanová was born in Ústí nad Labem, Czech Republic in 1982. She received her M.Sc degree at the University of West Bohemia (UWB) in Pilsen in 2006. She is currently a PhD student at the UWB and a research engineer at project MUSSLAP. Her main research interest is in data mining, machine learning, artificial intelligence, pattern recognition and computer vision. The topic of the PhD thesis is Emotion detection in audio-visual expression of the person.
Email: jeskynka.jana@seznam.cz

Hamdi Dibeklioğlu was born in Denizli, Turkey in 1983. He received his B.Sc. degree at Yeditepe University Computer Engineering Department, in June 2006. He is currently a research assistant and a M.Sc. student at Boğaziçi University Computer Engineering Department Media Laboratory. His research interests include computer vision, pattern recognition and intelligent human-computer interactions. He works on his thesis with Professor Lale Akarun on Part Based 3D Face Recognition.
Email: hamdi.dibeklioglu@boun.edu.tr

Semih Esenlik was born in Tekirdag, Turkey in 1984. He is currently an undergraduate student in Bogazici University, Turkey. He is a senior student in Electrical & Electronics Engineering Department. His specialization option is Telecommunication Engineering.
Email: semihese@yahoo.com

Nesli Bozkurt was born in Izmir, Turkey in 1982. She received her B.Sc. degree at M.E.T.U. Electrical and Electronics Engineering Department, in June 2005. She worked as a research assistant during her M.Sc. education at M.E.T.U E.E.E. Computer Vision and Intelligent Systems Laboratory for more than a year. She is currently employed for a software company and in the mean time; she works on her thesis with Asst. Prof. İlkay Ulusoy on 3D Scan Data Analysis and Improvement technologies.
Email: e124410@metu.edu.tr

Cem Demirkır was born in Gölcük, Turkey in 1970. He received his B. Sc. at İ.T.Ü. Electronics and Telecommunications Engineering Department in 1991, M.Sc. degree at M.E.T.U Electrical and Electronics Engineering Department in 2000. He is currently a research assistant at Turkish Air Force Academy and a Ph.D student at Boğaziçi University, BUSIM laboratory. He works with Prof.Dr. Bülent SANKUR on image/video Processing and biometrics.
Email: cemd@boun.edu.tr

Erdem Akagündüz was born in Izmir, Turkey in 1979. He received his B.Sc. and M.Sc. degrees at M.E.T.U Electrical and Electronics Engineering Department, in June 2001 and January 2004. He is currently a research assistant and a Ph.D. student at M.E.T.U E.E.E. Computer Vision and Intelligent Systems Laboratory. He works with Asst. Prof. İlkay Ulusoy on 3D Computer Vision, 3D Pattern Recognition and 3D Facial Modeling.
Email: erdema@metu.edu.tr

Kerem Çalışkan was born in Izmit, Turkey in 1978. He received his B.Sc. and M.Sc. degrees at M.E.T.U Computer Engineering Department, in June 2000 and January 2005. He is currently a Ph.D student at M.E.T.U. Informatics Institute - Medical Informatics department. He works with Asst. Prof. Didem Akcay. He owns an R&D company InfoDif whose specialization is in RFID and Signal Processing applications. His main interest areas are Real Time Signal Processing and Medical Imaging Devices.
Email: kcaliskan@infodif.com

Neşe Alyüz was born in İstanbul, Turkey in 1982. She received her B. Sc. Degree at Computer Engineering Department, I.T.U, İstanbul, Turkey in January 2005. She is currently an M. Sc. student in Computer Engineering Department, Boğaziçi University, İstanbul, Turkey. She works with Prof. Lale Akarun on 3D Face Registration and Recognition techniques.
Email: nese.alyuz@boun.edu.tr

Bülent Sankur has received his B.S. degree in Electrical Engineering at Robert College, İstanbul, and completed his graduate studies at Rensselaer Polytechnic Institute, New York, USA. Since then he has been at Boğaziçi (Bosporus) University in the Department of Electric and Electronic Engineering. His research interests are in the areas of Digital Signal Processing, Image and Video Compression, Biometry, Cognition and Multimedia Systems. He has established a Signal and Image Processing laboratory and has been publishing 150 journal and conference articles in these areas. Dr. Sankur has held visiting positions at University of Ottawa, Technical University of Delft, and Ecole Nationale Supérieure des Télécommications, Paris. He also served as a consultant in several private and government institutions. He is presently in the editorial boards of three journals on signal processing and a member of EURASIP Adcom. He was the chairman of ICT'96: International Conference on Telecommunications and EUSIPCO'05: The European Conference on Signal Processing as well as technical chairman of ICASSP'00.
Email: bulent.sankur@boun.edu.tr

İlkay Ulusoy was born in Ankara, Turkey in 1972. She received her B.Sc. degree at Electrical and Electronics Engineering Department, M.E.T.U., Ankara, Turkey in 1994, M.Sc. degree at the Ohio state University in 1996 and Ph.D. degree at Electrical and Electronics Engineering Department, M.E.T.U. in 2003. She did research at the Computer Science Department of the University of York, UK and Microsoft Research Cambridge, UK. She has been a faculty member at the Electrical and Electronics Engineering Department, M.E.T.U. since 2003. Her main research interests are computer vision, pattern recognition and graphical models.
Email: ilkay@metu.edu.tr

Lale Akarun received the B.S. and M.S. degrees in electrical engineering from Boğaziçi University, Istanbul, Turkey, in 1984 and 1986, respectively, and the Ph.D. degree from Polytechnic University, Brooklyn, NY, in 1992. From 1993 to 1995, she was Assistant Professor of electrical engineering at Boğaziçi University, where she is now Professor of computer engineering. Her research areas are face recognition, modeling and animation of human activity and gesture analysis. She has worked on the organization committees of IEEE NSIP99, EUSIPCO 2005, and eNTERFACE 2007. She is a senior member of the IEEE.
Email: akarun@boun.edu.tr

Tevfik Metin Sezgin graduated summa cum laude with Honors from Syracuse University in 1999. He received his MS in 2001 and his PhD in 2006, both from Massachusetts Institute of Technology. He is currently a Postdoctoral Research Associate in the Rainbow group at the University of Cambridge Computer Laboratory. His research interests include affective interfaces, intelligent human-computer interfaces, multimodal sensor fusion, and HCI applications of machine learning.
Email: metin.sezgin@cl.cam.ac.uk