3D Assisted Face Recognition: Dealing With Expression Variations

Nesli Erdogmus, *Member, IEEE*, Jean-Luc Dugelay, *Fellow Member, IEEE*

***Abstract*—One of the most critical sources of variation in face recognition is facial expressions, especially in the frequent case where only a single sample per person is available for enrollment. Methods that improve the accuracy in the presence of such variations are still required for a reliable authentication system. In this paper, we address this problem with an analysis-by-synthesis based scheme, in which a number of synthetic face images with different expressions are produced. For this purpose, an animatable 3D model is generated for each user based on 17 automatically located landmark points. The contribution of these additional images in terms of the recognition performance is evaluated with 3 different techniques (PCA, LDA and LBP) on FRGC and BOSPHORUS 3D face databases. Significant improvements are achieved in face recognition accuracies, for each database and algorithm.**

***Index Terms*—Face recognition, facial expressions, 3D facial animation**

I. Introduction

Recognition of humans has become a substantial topic today as the need for security applications grows continuously. Biometry enables reliable and efficient identity management systems by exploiting physical and behavioral characteristics of the subjects which are permanent, universal and easy to access. The motivation to improve the security systems based on single or multiple biometric traits rather than passwords and tokens emanates from the fact that controlling a person's identity is less precarious than controlling what he/she possesses or knows. Additionally, biometry-based procedures obviate the need to remember a PIN number or carry a badge.

Each having their own limitations, numerous biometric systems exist that utilize various human characteristics such as iris, voice, face, fingerprint, gait or DNA. "Superiority" among those traits is not a realistic concept when it is parted from the application scenario. The system constraints and requirements should be taken into account as well as the purposes of use-context that include technical, social and ethical factors [1]. For instance, while fingerprint is the most wide-spread biometric trait from a commercial point of view [2] (mainly due to a long history in forensics), it mostly requires user collaboration. Similarly, iris recognition, which is very accurate, highly depends on the image quality and also requires significant cooperation from the subjects.

Face recognition stands out with its favorable reconcilement between accessibility and reliability. It allows identification at relatively long distances for unaware subjects that do not have to cooperate. Like other biometric traits, the face recognition problem can also be briefly interpreted as identification or verification of one or more persons by matching the extracted patterns from a 2D or 3D still image or a video with the templates previously stored in a database.

Despite the fact that face recognition has been drawing a never-ending interest for decades and major advances were achieved, the intra-class variation problems due to various factors in real-world scenarios such as illumination, pose, expression, occlusion and age still remain a challenge. In Face Recognition Vendor Test (FRVT) 2002, it was demonstrated that using 2D intensity or color images, a recognition rate higher than 90% could be achieved under controlled conditions [3]. However, with the introduction of aforementioned variations, the performances deteriorated. The obtained results led to acceleration of studies on alternative modalities, especially the three-dimensional (3D) face, since it -by its nature- seems like a logical way to evade pose and illumination problems.

As 3D sensing technologies advance and the acquisition devices become more accurate and less expensive, the utilization of range data instead of / together with the texture data broadens. Consequently, in FRVT 2006 [4], an order-of-magnitude improvement in recognition performance was achieved over the FRVT 2002 for 3D face images, with FRR of 0.01 at a FAR of 0.001. Similar results were also achieved with high resolution still images under controlled illumination. Still, for the uncontrolled scenario, performances again drop.

Even though 3D face recognition has a better potential than its 2D counterpart, in practice it is not straightforward to obtain an accurate 3D image. Systems that extract shape information from 2D images, e.g. passive stereo approach, rely on the knowledge of extrinsic parameters of the scene and intrinsic parameters of the camera to obtain a certain degree of accuracy. On the other hand, with active sensors like laser scanners, a scan can take several seconds. It requires the

N. Erdogmus is with the Biometrics Group at IDIAP Research Institute, Centre du Parc Rue Marconi 19, PO Box 592, CH-1920 Martigny, Switzerland. (phone: +41 27 721 77 18; e-mail: nesli.erdogmus@idiap.ch)

J.-L. Dugelay is with the Multimedia and Communications Department at EURECOM, 2229 Route des Crêtes , BP 193 F-06560 Sophia-Antipolis, France. (phone: +33 49 300 81 41; e-mail: jean-luc.dugelay@eurecom.fr) .

This work was supported in part by the ANR under the project ANR-07-SESU-004.

Fig. 2. Synthesis examples: (a) input intensity image and accordingly synthesized face images under 8 different lighting conditions, 8 different pose variants and 6 different expressions [6] (b) Images rotated (left to right) by angles 5°, 10°, 25°, 35°; (top to bottom) illuminated under conditions where ($\alpha = 0°$) and ($\alpha = 30°, \tau = 120°$) [5]

subject to remain still during the process and furthermore reconstruction of depth is limited to short range. While these aspects are inconsequential in the well-controlled enrollment phase, they complicate the capture of the probe image. By taking this fact into consideration, we believe that a system, in which a combined enrollment is realized by both 2D and 3D information whereas the target images are still 2D images, is the optimal strategy to fuse advantageous features of the two modes.

With the assumption of this asymmetrical scenario, in this paper, we address the problem of expressions in 2D probe images. Our aim is to facilitate recognition by simulating facial expressions on 3D models of each subject. With regard to the causes of the intra-class variations, synthesis of facial images under various pose and illumination conditions using 3D face models is straightforward since these variations are external. However, this does not hold for expressions which alter the facial surface characteristics in addition to appearance. In order to achieve realistic facial expression simulations, we propose an automatic procedure to generate MPEG-4 compliant animatable face models from the 2.5D facial scans (range images) of the enrolled subjects based on a set of automatically detected feature points. Using a facial animation engine, different expressions are simulated for each person and the synthesized images are used as additional gallery samples for the recognition task. It is important to emphasize that synthetic sample augmentation is carried out during enrollment only once for each subject.

A. *Existing Work on 3D Assisted 2D Face Recognition*

Possible solutions for illumination and view point invariance in 2D face recognition are limited due to the 3D nature of the problem. By incorporating the 3D shape data of the face, researchers aim to improve 2D recognition performances in the presence of such variations. On the other hand, acquisition of facial models using 3D scanners can be problematic in the operational mode, especially under the scenario of uncooperative persons to be recognized. Mainly due to these two factors, the idea of 3D shape assisted 2D face recognition emerged, for which 3D shape can be reconstructed based on the captured 2D images or it can be acquired asymmetrically during the enrollment phase.

Methods based-on 2D images as their unique modality try and extract the 3D shape from the available data. In [5], a shape-from-shading (SFS) based method is proposed to generate synthetic facial images under different rotations and illuminations. In [6], a 3D generic face model is aligned onto a given frontal image using 115 feature vertices and different images are synthesized with variant pose, illumination and expression. Example synthetized faces from [5] and [6] are given in Fig. 2. A similar scheme is also presented in [7], where a personalized 3D face is reconstructed from a single frontal face image with neutral expression and normal illumination using 83 automatically located feature points (Fig. 1).

Fig. 1. An example study [7] is illustrated. Detected feature points and 3D reconstruction result is given in the top row. At the bottom, the synthesized pose and illumination images are compared with their originals and some generated expressions are presented.

Another study [8] presents a combination of an edge model and color region model for face recognition, after synthetic images with varying pose are created via a deformable 3D model. In [9], a 3D morphable model is used to generate 3D face models from three input images of the enrolled subjects. Similar to previously mentioned studies, the generated 3D models are utilized to augment the 2D training set with synthetic images rendered under varying pose and illumination conditions.

Lately, a study in which 3D model reconstruction is achieved by applying the 3D Generic Elastic Model approach is published [10]. Instead of enlarging the training set, they choose to estimate the pose of the test query and render the constructed 3D models at different poses within a limited search space about the estimated pose.

Differently from these mentioned methods, an important class of approaches that relies on 3D morphable models (3DMM) uses 3D data as an intermediate step for 2D face recognition. Again a generic 3D face model is morphed to match 2D images, but instead of synthesizing new facial images under different conditions, model coefficients obtained after fitting are used for recognition as feature vectors to be compared.

In [11], a morphable model that is based on a vector space representation of faces is constructed from 3D scans of 100 males and 100 females. After dense correspondences are established between the scans, Principal Component Analysis (PCA) is performed on the shape and texture vectors resulting in two orthogonal bases formed by 90 eigenvectors. During

the fitting process, the shape and texture coefficients together with illumination parameters are estimated iteratively to bring the morphable model as close as possible to the query image. Finally, two faces are compared by the set of coefficients that represent shape and texture using Mahalanobis distance. Later in [12], non-rigid ICP registration is proposed to replace the general optimization and this approach is proven to be more robust against missing data.

Methods that require 3D acquisition during the enrollment phase utilize the accurately captured shape data instead of extracting it from 2D image(s). In many approaches of this kind, the 3D face model is again utilized to generate synthetic images of the enrolled subjects with different poses or under different illumination conditions, in order to span their all possible appearances. The probe images in 2D are then compared with the synthetically augmented gallery for better match.

As one of the earliest examples of these studies [13], Chang et al. give an interesting approach to the use of 3D and 2D images. Taking only range maps as the gallery images, facial surface normals are decomposed into three components and their weighted sums are utilized to simulate 2D intensity images with different illuminations. The proposed recognition is reported to be less sensitive to the illumination variance. Later in [14], a similar approach is adopted in the sense of 3D-2D face matching. Exploiting the Canonical Correlation Analysis (CCA) to learn the mapping between range and texture LBP faces, similarity scores are computed and fused with the pure 2D counterpart.

In [15], authors develop a multi-modal system in which 3D models in the gallery are used to synthesize new appearance samples with pose and illumination variations for discriminant subspace analysis in order to improve the performance of the 2D modality. In a similar way, again a multi-model system is proposed in [16] in which synthetic views depicting various head poses and illumination conditions are generated and used to train Embedded Hidden Markov Models.

Another type of methodology in the reverse direction tries to remove existing variations in the probe images by pose normalization and/or illumination compensation.

In [17], a system is proposed in which 3D models of the users are acquired during enrollment and utilized for illumination correction by separating the effect of light and albedo in the 2D test images. Similarly, in [18], a multi-modal system is presented for which the 3D component of the face is utilized to compensate pose and illumination in its 2D counterpart. After estimating the pose using the nose tip and the facial symmetry line in the range image, it is normalized to be frontal. Next, scene illumination is recovered from the depth and color image pair and the input image is re-lighted. Two compensations are reported to improve the 2D face recognition performance.

Employing a combined enrollment with 2D and 3D data eliminates the risk of faulty face reconstruction. This potential scenario was also proposed during the Face Recognition Grand Challenge (FRGC), where enrolled images are 3D and the target images are still 2D images [19] but no baseline was provided. Reinforcing this trend, in their analysis [20], Husken et al. deduce that combining both modalities on different algorithmic levels is a promising approach to compensate for malfunctions of each of them separately.

Accordingly, in our study, acquired 2D and 3D face data are blended together in multiple steps of enrollment. Based on the assumption of a controlled environment, the subjects are enrolled with a neutral face and under ambient light, by both 2D and 3D sensors. 17 feature points are detected automatically and utilized to produce an MPEG-4 compliant animatable model for each subject by warping a generic head using Thin Plate Spline (TPS) method.

The main contribution of this paper revolves around this last stage, in which the rigid facial surface with no semantics is transformed in to a highly realistic animatable model. Once this model is obtained, it is animated using a compatible animation engine for various expressions. The efficacy of the generated synthetic face images are evaluated on 3 different face recognition systems.

Looking back at the existing works in this domain, we observe that 3D data is mostly utilized to compensate pose and illumination changes. In fact, the facial expressions are included only in [6] and [7]. In both studies; they are handled together with other variations without any particular analysis on expression simulations. In [6], the experiments are conducted on a small database of 10 subjects, for which the gallery set is augmented by synthesizing images under different pose, illumination and expression conditions. On the other hand in [7], a larger database is utilized but the impact of the generated images is only analyzed in terms of pose variations.

The remainder of this paper starts with the overview of the proposed system in Section II where the workflow is explained in detail. In Section II-A, the preprocessing of 2D and 3D data is described and in Section II-B, the automatic landmarking method is presented. Section II-C and II-D include the procedure to obtain an animatable model for an enrolled face sample and its simulation, followed by Section III which presents the utilized datasets with experimental results. Finally, the paper is concluded in Section IV.

II. Proposed System

In the proposed system, the enrollment is assumed to be done in both 2D and 3D for each subject under a controlled environment – frontal face images with a neutral expression and under ambient illumination.

The obtained 3D shape of the facial surface together with the registered texture is preprocessed, firstly to extract the face region. On the extracted facial surface, scanner-induced holes and spikes are cleaned and a bilateral smoothing filter is employed to remove white noise while preserving the edges.

After the hole and noise free face model (texture and shape) is obtained, 17 feature points are automatically detected using either shape, texture or both, according to the regional properties of the face [21]. These detected points are then

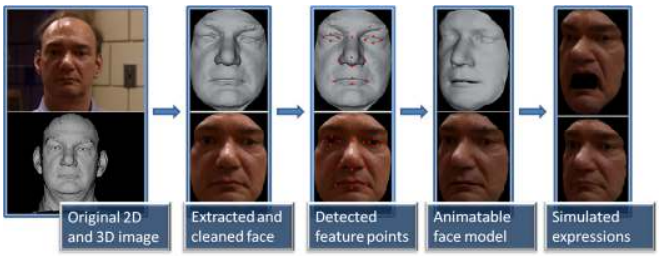


Fig. 3. The enrollment procedure is illustrated on an example.

utilized to warp a generic animatable face model so that it completely transforms into the target face. The generic model with manually labeled 71 MPEG-4 points is suitable to simulate facial actions and expressions via an animation engine that is in accordance with MPEG-4 Face and Body Animation (FBA) specifications.

Finally, in order to simulate the facial expressions on the obtained animatable model, an animation engine, called visage|life™[1] is utilized. Multiple expression-infused face images are generated for each subject to enhance face recognition performance. The whole system is illustrated in Fig. 3.

A. *Data Preprocessing*

3D scanner outputs are mostly noisy. The purposes of the preprocessing step can be listed as:

1. to extract the face region (same in 2D and 3D images);
2. to eliminate spikes/holes introduced by the sensor;
3. to smooth the 3D surface.

Firstly, adopting the method proposed in [22], the nose tip is detected: For each row, the position with the maximum z value is found and then for each column, the number of these positions is counted to create a histogram. The peak of this histogram is chosen as the column for the position of the vertical midline, and the maximum point of this contour is identified as the nose tip. Using a sphere of radius 80mm and centered 10mm away from the nose tip in +z direction, the facial surface is cropped.

Next, the existing spikes are removed by thresholding. Spikes are frequent with laser scanners, especially in the eye region. After the vertices that are detected as spikes are deleted, they leave holes on the surface. Together with other already existing holes (again usually around the eyes and eyebrows), they are filled by applying linear interpolation.

Once the complete surface is obtained, a bilateral smoothing filter [23] is employed to remove white noise while preserving the edges. This way, the facial surface is smoothed but the details hidden in high frequency components are maintained.

B. *Automatic Landmarking*

Bearing in mind that subject cooperation is required during the enrollment, we base our system on the assumption of a well-controlled acquisition environment in which subjects are registered with frontal and neutral face images. In accordance with our scenario, we aim to extract a subset (17 points) of MPEG-4 Facial Definition Parameters (FDPs) to be utilized for the alignment of the faces with the animatable generic model. For the extraction of the points, 2D and/or 3D data are used according to the distinctive information they carry in that particular facial region.

[1]Visage Technologies – The Character Animation Company, visage|life http://www.visagetechnologies.com/products_life.html.

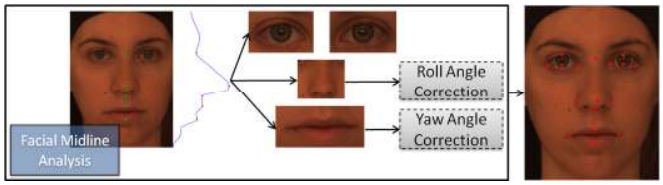


Fig. 5. The outline of the landmarking process and the detected feature points as the outcome

Firstly, facial midline (vertical profile) analysis is done and 5 fiducial points on that midline are detected. Based on that information; face is split into sub-regions for the coarse localization of eyes, nose and lips. After that, further analysis is done inside these extracted sub-regions to detect the points of interest. For those regions with non-informative texture (like nose), 3D data is analyzed. On the other hand for the regions with noisy surface and/or distinctive color information (like eyes), 2D data is utilized. As a result, 17 facial interest points are detected in total, consisting of 4 points for each eye, 5 points for the nose and 4 points for the lips (Fig. 5). The steps are detailed in the following:

1) *Vertical Profile Analysis*

The analysis done on the vertical profile constitutes the backbone of the whole system. It starts with the extraction of the facial midline and for this purpose; the nose tip is detected as explained previously. The nose tip position allows us to search for the eyes in the upper half of the face in order to approximately locate irises, so that the roll angle of the face can be corrected before any further processing.

For coarse iris extraction, the non-skin region is found by removing the pixels with the most frequent chrominance values present in the upper half of the face in YCbCr space. Edge maps are constructed for the non-skin region using Canny edge detector by iteratively adjusting the threshold until a descriptive map is obtained. Subsequently, Hough transform is applied to the edge map to detect circles. For each detected circle, an overlapping score is calculated by the ratio

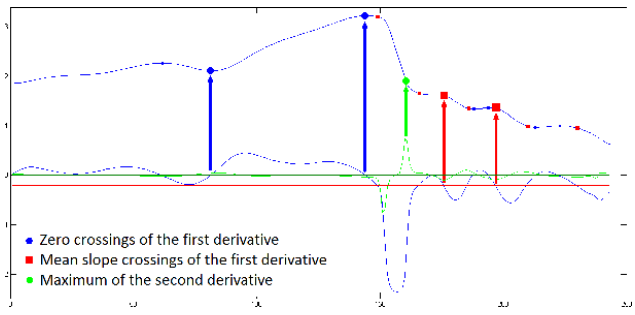


Fig. 4. A sample profile curve and its first (blue) and second (green) derivative curves. The arrows show the five detected interest points among the candidates.

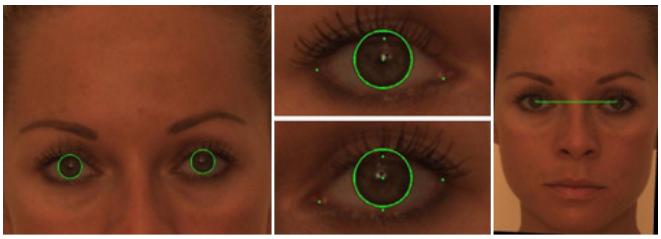


Fig. 6. a) Initial detection of iris circles in the upper half of the face b) Result of the detailed analysis for each eye c) Re-rotated face to align iris centers

of the detected portion of the circle to the whole circle parameter. After the detected iris candidates are grouped as right and left according to their relative positions to the profile line, the one with the maximum total overlapping score is selected among the compatible pairs [24]. Next, the 2D and 3D images of the face are rotated in order to align the detected iris centers on the same horizontal line. Thereby, our assumption for vertical profile is better assured.

After correcting the roll angle of the face, the nose tip location is recalculated; the vertical midline of the face is extracted and smoothed by median filtering. This curve compounds of bulges (forehead, nose, upper and lower lips and chin) and bores in between. Even though, those shapes do not fully expose the location of the facial interest points, they can be highly informative. For this reason, the peaks and nadirs in the profile curve are found with the help of the zero-crossings of the curve's first and second derivatives, except the end of the nose which is located as the maximum of the second derivative.

For the points on the lips, since this region is usually inclined due to the spherical shape of the facial surface, the "mean slope" is calculated between the end of the nose and the bottom of the face. A set of "mean slope-crossings" are calculated by subtracting the calculated mean slope from the first derivatives. The curves for the profile and the first and the second derivatives are depicted in Fig. 4.

Now those landmarks are known, the face can be broken into more meaningful sub-images for locating or refining the locations of points of interest.

2) Eye Regions

The 3D surface around the eyes tends to be noisy because of the reflective properties of the sclera, the pupil and the eyelashes. On the other hand, its texture carries highly descriptive information about the shape of the eye. For that reason, 2D data is preferred and utilized to detect the points of interest around the eyes, namely the iris center, the inner and outer eye corners and the upper and the lower borders of the iris.

For this purpose, the method proposed in our previous work [24] is adopted. After applying an averaging filter with a rectangular kernel and the noise and the horizontal edges are suppressed, the vertical edges are detected with a Sobel operator. Using the vertical edge image, the irises are once again detected by Hough transform, this time more accurately.

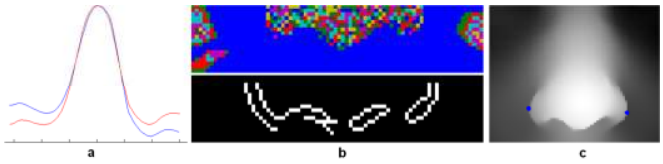


Fig. 7. a) The horizontal profile of the nose tip before (blue) and after correction (red) b) The minimum curvature and the corresponding edge map c) The depth map of the nose region with detected points marked

For the detection of eye corners, horizontal edges that belong to the eyelids are detected as described in [24] and two polynomials are fitted for lower and upper eyelids. The inner (near the nose bridge) and outer eye corners are determined as the intersection points of the two fitted polynomials.

After this step, the image is rotated one last time in the 2D image plane, if necessary, to horizontally align the two iris centers (Fig. 6).

3) Nose Region

Contrary to the eye region, nose region is extremely distinctive in surface but quite plain in texture. For this reason, we choose to proceed in 3D.

To start with, the yaw angle of the face is corrected in 3D. For this purpose, the horizontal curve passing through the nose tip is examined. Ideally, the area under this curve should be equally separated by a vertical line passing through its maximum (assuming the nose is symmetrical). With this assumption, the curve is iteratively rotated to minimize the difference between these two partitions under the curve. Once the angle is determined, the whole surface is rotated, so that a "more frontal" face is obtained.

After this adjustment, the minimum curvature is calculated for each point in the nose region. Then, edge detection is applied on the minimum curvature map. Those edges reveal the position of the points of interest on both sides of the nose tip (Fig. 7). The other three points; on the nose bridge, on the nose tip and at the bottom of the nose are already found during the vertical profile analysis.

4) Lips Region

Numerous studies are proposed to extract the lip contour based on deformable [25] or active contour models [26] [27], working only with 2D images. However with our system, a much simpler method can be easily adopted because we have good estimates of two points of interest (upper and lower mid points) on the lip, thanks to the analysis performed on the vertical midline of the face.

Since we work on faces with neutral expressions, the mouth is assumed to be closed. A closed mouth always yields to a darker line between the two lips. Based on this knowledge, the contact point of the lips is found by applying a vertical projection analysis. For this purpose, the row R with the minimum projection value P in the lip image is computed as:

$$R = \arg min_r \; P(r) = \arg min_r \sum_{pixel\; p \in r} I(p) \qquad (1)$$

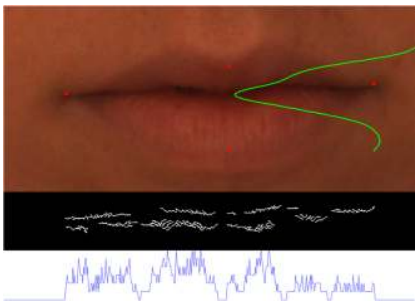


Fig. 8. A lip image is given with the detected points. The green line shows the horizontal projection result. At the bottom, the calculated edge map and its vertical projection is given.

where I is the grayscale intensity. Afterwards, horizontal edges are detected in a narrow window around *R*, and horizontally projected in a similar fashion to detect left and right corners of the lips (Fig. 8).

Once the 17 points are obtained, they are used to align the generic face to the target face for the construction of the animatable model.

C. Constructing the Animatable Face Models

In order to construct an animatable face model for each enrolled subject, a mesh warping algorithm based on the findings in [28] is proposed. A generic face model, with holes for the eyes and an open mouth is strongly deformed to fit the facial models in the database, using the TPS method. 17 points to be automatically detected together with the rest of the FDP points for MPEG-4 compliant animations are annotated for the generic face (Fig. 9).

MPEG-4 specifications and the mathematical background of the TPS method will be briefly explained before going into details about the proposed animatable face construction method.

1) MPEG-4 Specifications and Facial Animation Object Profiles

MPEG-4 is an ISO/IEC standard developed by Moving Picture Experts Group which is a result of efforts of hundreds of researchers and engineers from all over the world. Mainly defining a system for decoding audiovisual objects, MPEG-4 also includes a definition for the coded representation of animatable synthetic heads. In other words, independent of the model, it enables coding of graphics models and compressed transmission of related animation parameters.

The facial animation object profiles defined under MPEG-4 are often classified under three groups [29] [30] [31]:

- Simple facial animation object profile: The decoder receives only the animation information and the encoder has no knowledge of the model to be animated.
- Calibration facial animation object profile: The decoder also receives information on the shape of the face and calibrates a proprietary model accordingly prior to animation.
- Predictable facial animation object profile: The full model description is transmitted. The encoder is capable of completely predicting the animation produced by the decoder.

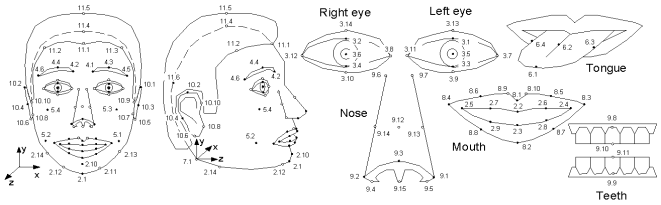


Fig. 9. MPEG-4 Facial Definition Parameters

The profile most conformable to our approach is the second one: calibration facial animation object profile, since we are aiming to calibrate the "generic" model according to the samples in our database. Our system generates an animatable model by using 17 of 71 MPEG-4 specified face definition parameters (FDP) which are annotated automatically. The rest of the points are only marked on the generic model for animation.

In Fig. 9, the positions of the MPEG-4 FDP points are given. Most of these points are necessary for an MPEG-4 compliant animation system, except for the ones on the tongue, the teeth and the ears, depending on the animation tool structure.

2) Thin Plate Spline Warping

As the name indicates, the TPS method is based on a physical analogy to how a thin sheet of metal bends under a force exerted on the constraint points. The TPS method was made popular by Fred L. Bookstein in 1989 in the context of biomedical image analysis [32].

For the 3D surfaces S and T, and a set of corresponding points (point pairs) on each surface, P_i and M_i respectively, the TPS algorithm computes an interpolation function f(x,y) to compute T', which approximates T by warping S:

$$T' = \{ (x', y', z') \mid \forall (x, y, z) \in S, \quad x' = x;\ y' = y;\ z' = z + f(x, y) \} \tag{2}$$

$$f(x, y) = a_1 + a_x x + a_y y + \sum w_i U(|P_i - (x, y)|) \tag{3}$$

with U(.), the kernel function, expressed as:

$$U(r) = r^2 \ln r^2,\ r = \sqrt{x^2 + y^2} \tag{4}$$

In the interpolation function f(x,y), the w_i, $i \in \{1,2,...n\}$ are the weights. As given in (3), the interpolation function consists of two distinct parts: An affine part ($a_1+a_x x+a_y y$) which accounts for the affine transformation necessary for the surface to match the constraint points and a warping part ($\Sigma w_i U(|P_i-(x,y)|)$).

3) The Method

TPS is commonly used to establish registration in non-rigidly deformed surface patches, like two different facial surfaces [33]. The deformation of the registered models is minimal since only few point pairs are utilized. As more

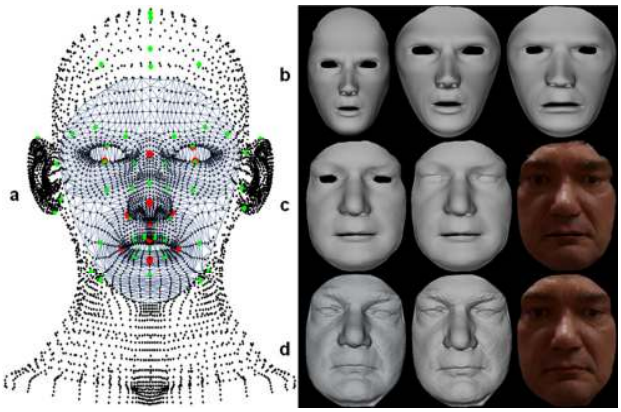


Fig. 10. a) The point cloud for the generic head model is given with MPEG-4 points marked. Red points are the 17 feature points that are also located on the target faces. Only the manually cropped face region (in blue) of this head model is utilized for warping. b) Generic face before the warping procedure and after anisotropic scaling and coarse warping (using the 17 point pairs only). c) Generic face after fine warping (using half of the points on the generic face), after creating the eyeballs and after being textured. d) The target face, before alignment, after alignment without and with texture.

control points are added to the TPS warping, the amount of deformation increases and the face becomes more and more similar to the target surface. By exploiting this fact, in this paper we propose to strongly deform a generic face to fit target faces in the gallery.

On the other hand, before applying the TPS warping, we have to make sure that the target face and the generic face models are well-aligned. Each step of the process to obtain the animatable model is detailed in the following subsections.

a) Rescaling and alignment:

In the first step, the generic model is rescaled in three directions:

- X: The x-distances between the two outer eye corners of the generic model and target are equalized.
- Y: The y-distances between the outer left eye corner and the lower mid-point of the lips are equalized.
- Z: The z-distances between the outer left eye corner and the nose tip are equalized.

Next, a rigid transformation is calculated based on the 17 point pairs. Using the two sets of landmarks, the best fit mapping is computed in a least squares sense, where the squared distance between the point sets is minimized. The calculated transformation is applied on the target face. This step corrects the existing deviations from the frontal pose of the enrolled face (if any) while better aligning the two surfaces.

b) Coarse Warping

The generic model is warped coarsely for better alignment to the target face. By taking the 17 feature point pairs as the source and target landmarks, a non-linear warp is calculated, which moves any point on the mesh around a source landmark closer to the corresponding target landmark. The points in between are interpolated smoothly using the Thin Plate Spline algorithm.

c) Fine Warping

At this stage, the two surfaces are very well aligned and hence, we can assume that for all points on the generic model, the corresponding pair on the target model is the one that is the closest. Based on this assumption, for every second point on the generic face, the matching vertex on the target face is found and used in TPS calculation. This way, half of the points on the generic model conform into the target face counterparts and maintain their smoothness.

Finally, two spheres for the two eyes are created based on the 4 feature points detected around each eye and the texture is copied onto the obtained animatable model. The proposed method for animatable model generation is illustrated on a sample model in Fig. 10.

Fig. 11. The 12 simulated expressions on a sample face

D. Expression Simulations

Once the animatable face models are generated, they are animated with 12 different expressions that are pre-defined in the visagelife™ animation tool and the images of the resulting faces are rendered. The simulated expression are presented on an example face in Fig. 11

III. Performance Evaluation

In order to evaluate the proposed system two databases are utilized: FRGC [19] and Bosphorus [34] 3D face databases.

The 3D data for FRGC experiments is divided into two partitions for training and testing. The training set consists of 943 3D scans and controlled and uncontrolled still images, whereas the validation partition contains 4007 3D and 2D images collected from 466 individuals with different facial expressions. On the other hand, in the Bosphorus database, there are 105 subjects in various poses, expressions and occlusion conditions, with the total number of 4666 face scans.

Contrary to the assumed scenario, in the FRGC database, the images are taken in uncontrolled illumination conditions. For this reason, automatic landmarking is tested only on the Bosphorus database for which the subjects were recorded in a highly controlled environment and a 1000W halogen lamp was used in a dark room to obtain homogeneous lighting for good quality texture images.

3 key techniques are adopted in order to observe the effect of gallery augmentation: PCA, LDA [35] and LBP [36]. The evaluations are done for both verification and identification scenarios.

Verification is a binary (1:1) classification problem with two possible types of errors: False acceptance where two samples from two different persons are classified as a match

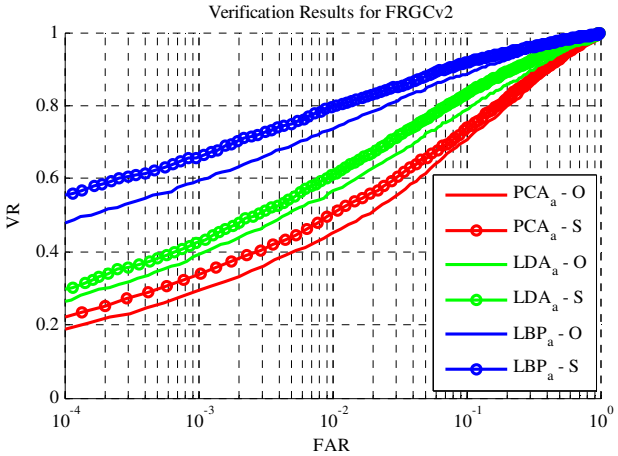


Fig. 12. ROC curves for PCA, LDA and LBP methods applied to the whole FRGCv2 database using the original (O) and the synthesized (S) galleries

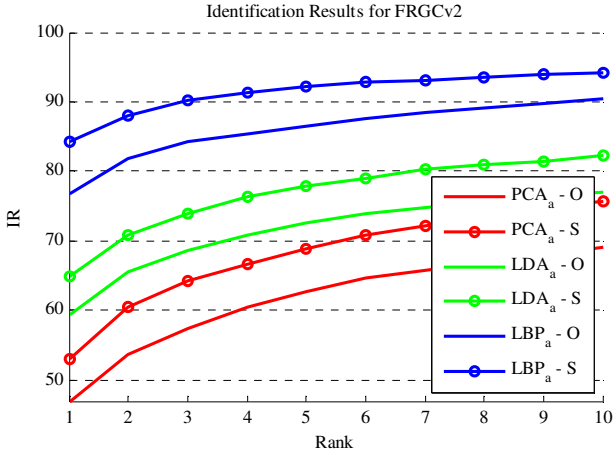


Fig. 13. CMC curves for PCA, LDA and LBP methods applied to the whole FRGCv2 database using the original (O) and the synthesized (S) galleries

and false rejection where two samples from the same person are classified as a non-match. The percentages of occurrences for these errors are referred as False Acceptance Rate (FAR) and False Rejection Rate (FRR), respectively. These two rates are inversely related and can be adjusted by changing the verification score threshold. In our experiments, the verification performances are reported at two different thresholds:

$$\tau_1 = \{\tau : FAR(\tau) = 0.001\}$$

$$\tau_2 = \arg min_\tau |FAR(\tau) - FRR(\tau)|$$

Verification rate (VR) at τ_1: Percentage of correctly accepted clients (1-FRR) at 0.1% FAR.

Equal error rate (EER) at τ_2: Error rate at which FAR and FRR are equal.

Finally, Receiver Operating Characteristics (ROC) curves are also presented to visualize verification performances with varying thresholds.

Identification, on the other hand, is a 1:N matching, where probe images are matched with enrolled subjects in the gallery. In this scenario, rank-1 identification rates (IR) are given, which is the percentage of correctly classified probe images. Additionally, Cumulative Match Characteristic (CMC) curves are also presented which plot the probability of correct identification with varying candidate list size.

A. *Experiments with FRGC Database*

In the validation set of the FRGC database, 466 persons are recorded with different facial expressions. According to the proposed system, the enrollment process is realized with a single neutral 3D+2D image for each person. Hence, the persons with at least one neutral image in the database are selected, resulting in a gallery of 400 persons and the rest of the images that belong to the enrolled subjects (3522 images) are taken as the probe set.

Firstly, 3D face data is cropped and processed as detailed in Section III and in order to produce the synthetic face images with expressions, an animatable model for each enrolled person is constructed based on 17 manually labeled landmarks by Szeptycki et al [37]. The obtained faces are animated for 12 expressions and their images are stored in the simulation gallery.

TABLE I
FACE RECOGNITION ACCURACIES WITH THE ORIGINAL (O) AND THE SYTHESIZED (S) GALLERIES

Method	VR	EER	IR
PCA - O	29,27%	20,15%	47,02%
PCA – S	33,84%	18,85%	52,98%
LDA – O	39,41%	16,23%	59,51%
LDA – S	42,33%	13,91%	64,96%
LBP – O	59,63%	10,83%	76,89%
LBP – S	65,82%	8,82%	84,21%

Next, all images in the training, gallery and probe sets are preprocessed by applying cropping, geometrical normalization with respect to eye positions (64x80 pixels) and histogram equalization.

2D subset of the FRGC training set is used to train PCA and LDA classifiers for both FRGC and Bosphorus experiments. Considering the cumulative sum of the eigenvalues, the most significant 97 eigenvectors (90%) are used for representation. The first three eigenvectors are ignored in order to achieve a certain degree of illumination insensitivity [35]. Finally, the recognition is achieved by the minimum cosine distance between the test and gallery images.

Verification rates at 0.1% FAR, equal error rates and rank-1 identification rates for PCA, LDA and LBP methods with and without using the simulated images are given in TABLE I. Additionally in Fig. 12 and Fig. 13, ROC and CMC curves are given for all methods using the original and the synthesized galleries.

According to the obtained results, LBP is the most positively affected method with 6.2% increase in VR and 7.3% increase in IR. If we look at the two recognition scenarios, for all three methods it is observed that higher gain is obtained for identification.

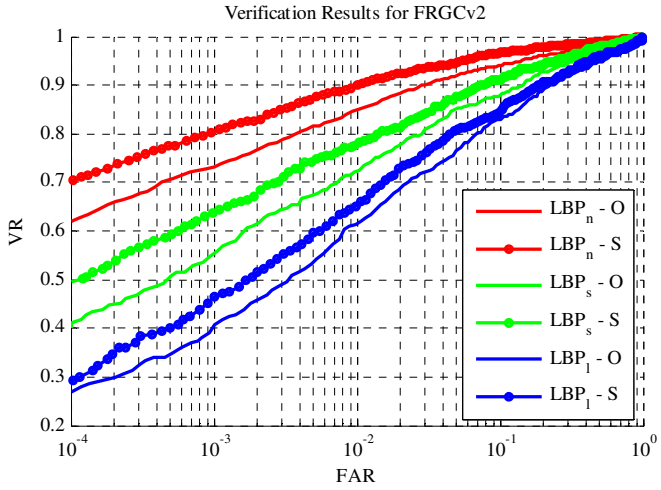


Fig. 14. ROC curves for the LBP method for 3 subsets: Neutral (n), small expressions (s) and large expressions (l) with (S) and without (O) simulations

TABLE II
RECOGNITION RATES FOR LBP METHOD WITH THE ORIGINAL (o) AND THE SYNTHESIZED (s) GALLERIES FOR EACH SUBSET

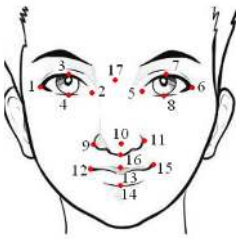
Set	VR	EER	IR
Neutral-o	73,28%	6,96%	83,81%
Neutral-s	80,16%	5,26%	90,98%
Small-o	54,89%	11,30%	73,23%
Small-s	63,59%	9,20%	80,32%
Large-o	40,61%	14,34%	61,05%
Large-s	46,41%	12,90%	69,06%

For further analysis, the probe images are broken down to three subsets according to their categories of facial expressions: Neutral, small and large [38]. Small expressions include moderate smiles and talking gestures. On the other hand, large contains unnatural expressions like blown cheeks. This arrangement results in a neutral probe set of size 2051 (58.2%), small probe set of size 747 (21.2%) and large probe set of size 724 (20.6%).

This approach is adopted because for face recognition the magnitude of the shape change is mostly much more important than its type. It is important to see the impact of the database augmentation for expression sets of different strengths. In Fig. 14, recognition results using LBP for each category are presented and more detailed breakdown of the results is given in TABLE II.

As expected, the recognition performances deteriorate with stronger expressions. Under the verification scenario, the subset with small expressions is observed to benefit most from gallery augmentation with expression simulated images. It acquires an increase of 8.7% in VR values. This might be due to the fact that the expressions simulated for this study mainly include moderate facial deformations. On the other hand, for the identification case, the improvements are similar for all 3 subsets while the one with large expressions receives the maximum increase (8.0%).

TABLE III
SUCCESS RATES FOR THE DETECTED POINTS AT 10% ERROR THRESHOLD



Points	Success Rates	Points	Success Rates
1	93.11%	**10**	93.65%
2	95.66%	**11**	96.99%
3	93.32%	**12**	88.96%
4	96.21%	**13**	95.32%
5	95.30%	**14**	94.65%
6	93.85%	**15**	87.96%
7	93.01%	**16**	100.00%
8	96.44%	**17**	78.60%
9	91.64%		

B. *Experiments with Bosphorus Database*

Among 4666 facial scans in the database, one neutral face scan for each individual is chosen as the single gallery sample. The emotions subset of the database which includes posed expressions of the six basic emotions (anger, disgust, fear, happiness, sadness and surprise) and the remaining neutral scans compose a test set of 647 images. Since the facial surfaces are already cropped and cleaned in the database, only smoothing is applied.

Since the dataset is collected under a highly controlled environment, the automatic landmarking is applied.

1) *Automatic Landmarking*

For each of the 105 subjects, 17 feature points are located automatically. The localization error is taken to be the ratio of the Euclidean distance between the automatically detected and manually labeled points to the inter-pupillary distance (IPD – distance between the center of the pupils of the two eyes).

The success rates computed for the error threshold of 10% are given in TABLE III. The least accurate results are obtained for the point 17 at the nose bridge. In addition to the high ambiguity in the location of that point, this is also due to the difficulty in marking it on 2D.

The best result is achieved with point 16, for which the evaluation is done on the vertical profiles of each face since its exact location is not always obvious in color images.

For all points detected on all faces (1785 points in total), the success rate is 82.91% for 5% error threshold and 93.07% for 10% error threshold.

2) *Face Recognition*

In order to observe the effects of the automatic landmarking errors on the performance improvements, identification tests are conducted with both automatically detected and manually marked feature points. For 105 persons in the database, animatable models are generated and 12 expressions are simulated to create the synthetic gallery images. All images in both gallery and probe sets undergo the same preprocessing steps as in the FRGC experiments.

Using a single neutral image per person, the rank-1 identification rate for PCA is found to be 63.83%. With the addition of synthetic gallery images, generated by using the manually marked feature points, the rate rises to 70.79%. Whereas, the increase using synthesized gallery images that

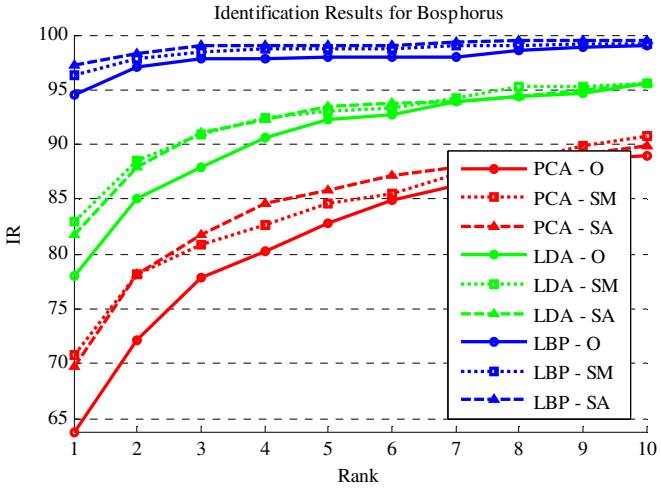


Fig. 15. Recognition rates from rank-1 to rank-10 with PCA for the three experiments: without simulation, with simulation using manually marked landmarks and with simulation using automatically detected landmarks

are created based on automatically detected landmarks is a bit lower (69.71%). A similar trend is also observed for the LDA method.

On the other hand, the rank-1 identification rate with LBP method is observed to be slightly higher with the points that are detected automatically when compared to the ones that are manually marked. The CMC curves for all methods and galleries are given in Fig. 15. Additionally, recognition rates are presented in TABLE IV in more detail.

These results are very impressive in the sense that despite the errors in the automatic detection of landmarks, using these points for simulated image generation leads to a comparable amelioration in results with the ones created using manually marked points. The reason behind this result lies in the fine warping step in the animatable model generation process. At this stage, the landmarks are left aside and the correspondences are created by pairing every second point on the generic model to the closest vertex in the face scan. Hence, the errors in coarse warping due to the inaccuracies in automatic landmarking are partially corrected.

TABLE IV
FACE RECOGNITION ACCURACIES WITH THE ORIGINAL (O) AND THE SYTHESIZED GALLERIES WITH MANUALLY MARKED (SM) AND AUTOMATICALLY DETECTED (SA) LANDMARKS

Method	VR	EER	IR
PCA - O	37,71%	15,93%	63,83%
PCA – SM	40,80%	15,24%	70,79%
PCA – SA	38,95%	14,85%	69,71%
LDA – O	49,92%	10,35%	78,05%
LDA – SM	53,94%	9,97%	83,00%
LDA – SA	55,18%	9,81%	81,76%
LBP – O	67,54%	8,14%	94,59%
LBP – SM	68,62%	7,83%	96,29%
LBP – SA	69,09%	8,075	97,22%

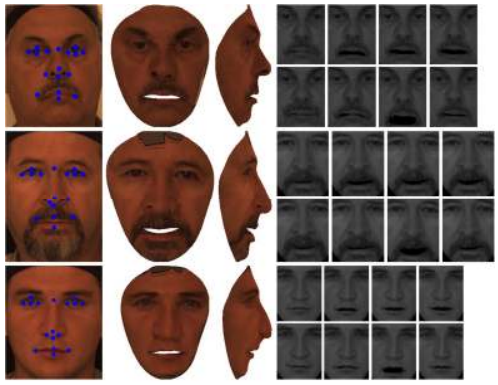


Fig. 16. Three examples with high errors in automatic detection of landmark points, the resulting animatable model after coarse warping (frontal and profile) and final images with simulated expressions after fine warping

Several example faces that are corrected in the fine warping stage are given in Fig. 16 with the resulting simulation results.

Similar to the results obtained with the FRGC database, gallery augmentation via expression simulations is found to be more advantageous for the close-set identification scenario. It shows that the simulated images added to the gallery can be more similar to the probe images than the enrollment samples. This increases the chances of finding a match in lower ranks.

C. *Discussions*

In this section, we will compare the obtained results with previous studies and present a short analysis on the impact of generated animatable model accuracies on the improvement of recognition rates.

As presented in Section I, the idea of enriching a 2D gallery with synthetically generated images is previously employed. However, this application is mostly limited to pose and illumination variations; mainly because it is much more straightforward to render 3D face models under different illumination conditions and with different poses. In fact, to the best of our knowledge, there are only two studies which attempt to simulate different expressions [6] [7].

In [6], identification experiments are conducted on a database of 10 subjects using a PCA-based method. According to the results, an increase of about 6% is obtained in IR using synthetized images of not only various expressions but also different pose and illuminations (Fig. 17a). This improvement is consistent with the outcomes of our experiments on FRGC and Bosphorus. On the other hand, the database utilized in [6] is too small to make a strong conclusion. The most important drawback of the approach proposed in this study is that 115 landmarks are required for facial variation synthesis. The landmarks are labeled manually for the experiments but in practice, it would be highly challenging to manually annotate the gallery images for a large number of clients or to locate these points automatically.

Similar to [6], different pose and illumination conditions are also simulated in [7] in addition to expression variations. On the other hand, the virtual faces are synthesized with different pose-illumination-expression (PIE) combinations.

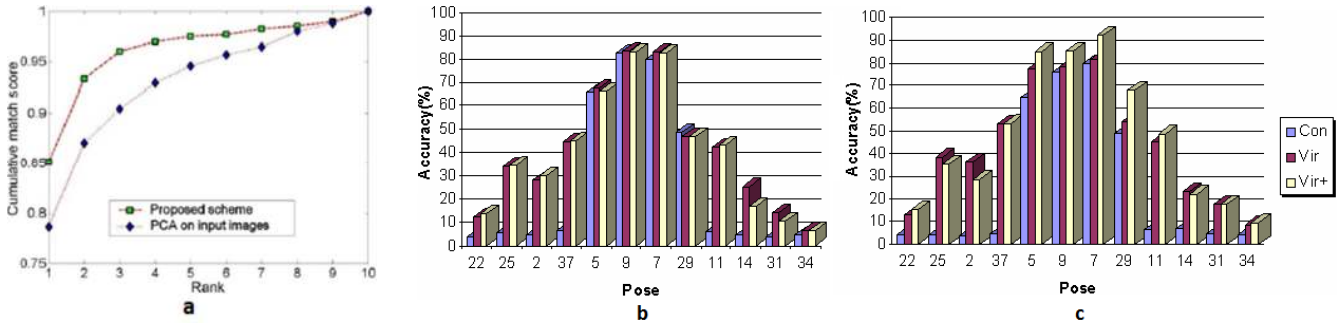


Fig. 17. (a) Performance of PCA-based method with and without data augmentation by synthesis [6] (b) Recognition accuracy comparison between face recognition with/without virtual face using PCA [7] (c) Recognition accuracy comparison between face recognition with/without virtual face using LDA [7]

Identification experiments are conducted on the CMU-PIE database [39] using PCA and LDA methods. The results are presented in Fig. 17b and Fig. 17c, respectively. While the conventional method uses only the frontal and neutral faces (Con), the proposed method includes the generated virtual images (Vir). They also tested another configuration (Vir+) where a priori knowledge of pose is assumed and only pose specific virtual images are utilized.

For a meaningful comparison with our results, the accuracy improvement for probe images in near-frontal poses (5, 9 and 7) is considered using all virtual images (Vir). The plots show that with PCA, the achieved increase in accuracy is about 3% (average of the three pose bins), whereas it is about 5% for LDA. Although the utilized database has a smaller set of clients (68 subjects) with respect to FRGC and Bosphorus databases and the number of synthesized images is much higher, the obtained improvements in identification rates are comparable to ours. The results are summarized in TABLE V (The values for [6] and [7] are derived approximately from the plots, since the exact results are not given.)

Fitting a morphable model to 2D and 3D images to handle expression variations is also fairly common but instead of enlarging the gallery with simulated expressions, these methods propose to model identity and expression related deformations on the facial surface separately to facilitate face or expression recognition (e.g. [40]). It is important to see that this is completely different than the proposed approach.

TABLE V
COMPARISON WITH IMPROVEMENTS IN RANK-1 IDENTIFICATION RATES OF PREVIOUSLY PROPOSED METHODS

Method	Database	No. of subjects	ΔIR
PCA [6]	private	10	~6%
PCA [7]	CMU_PIE	68	~3%
LDA [7]	CMU_PIE	68	~5%
PCA	FRGC	400	5,96%
LDA	FRGC	400	5,45%
LBP	FRGC	400	7,33%
PCA	Bosphorus	105	5,87%
LDA	Bosphorus	105	3,71%
LBP	Bosphorus	105	2,63%

Lastly, we would like to touch on the subject of fitting quality for animatable model generation and its impact on recognition performances. This discussion is also related to the realism of the models and expression simulations but for the time being, this analysis is not viable, since we have used a single method and tool for animations.

On the other hand, it is possible to assess the quality of the obtained animatable models for each subject and observe its correlation to subject-specific recognition rates. Here, we define the quality as the similarity to original face data obtained during enrollment.

In order to measure the quality, similarities of 3D animatable models and their rendered images to the original enrolled samples are measured. For this purpose, Bosphorus database is utilized where LBP and ICP methods are employed to compare 2D and 3D data, respectively. For 2D images, the χ^2 distance between the two LBP histograms is computed. For 3D face models, the final error after ICP alignment is taken as the distance metric. Finally, for each subject, rank-1 identification rates are found with and without using simulations.

The effect of the simulations on identification performances is calculated as the amount of increase in IR values. For a better evaluation, the subjects having 100% identification rate both with and without using the synthetized images are excluded from the analysis. For the remaining 91 subjects, the differences between IR values and 2D and 3D distances are plotted in Fig. 18 after being normalized to [0,1].

The results show no strong connections between the two distances and the increase in recognition performances. In fact, the correlation coefficients between the two quality measurements and IR difference are found to be remarkably low; 0.22 and 0.14 for 2D and 3D respectively.

IV. CONCLUSIONS

Based on the assumption of a fully-controlled environment for enrollment, a face recognition framework is proposed in which the widely-encountered single sample problem for identification of faces with expressions is targeted by augmenting the dataset with synthesized images. Several expressions are simulated for each enrolled person on an animatable model which is specifically generated based on the

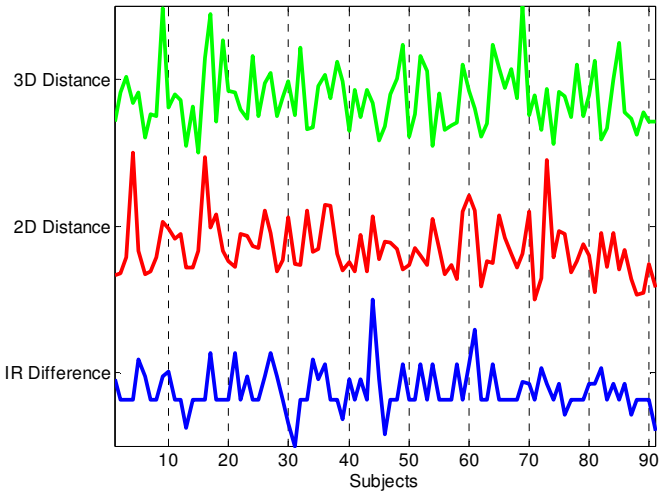


Fig. 18. Subject-specific identification rate differences are plot together with the 2D and 3D distances between the enrollment data and the generated animatable model of each subject

3D face scan of that subject.

For the animatable model generation, a generic model for which MPEG-4 FDPs are located manually is utilized. Based on only 17 common points on both the generic and the target models, the generic model is first coarsely warped using the TPS method. Then, assuming the surfaces are close enough, new and denser point correspondences are formed by pairs with minimum distance and fine warping is applied. Finally, the texture is copied.

A sub-procedure on automatic detection of those 17 landmarks is presented utilizing both 2D and 3D facial data.

For the simulation of facial expressions on the generated models, an animation engine, called visagelife™ is utilized. The facial images with expressions constitute a synthetic gallery, of which the contribution to the face recognition performance is evaluated on a PCA-based implementation.

The experiments are conducted on two large and well-accepted databases; FRGC and Bosphorus 3D face database. The experiment results reveal that introduction of realistically synthesized face images with expressions improves the performance of the identification system. Additionally, based on the evaluations on Bosphorus database, the imprecisions introduced by the proposed automatic landmarking algorithm has no adverse effect on the success rates thanks to the corrective property of the warping phase.

ACKNOWLEDGMENT

This research has been supported by the ANR under the National French project ANR-07-SESU-004 and the FP7 European TABULA RASA Project (257289).

REFERENCES

[1] L. D. Introna and H. Nissenbaum, "Facial recognition technology: A survey of policy and implementation issues," Report of the Center for Catastrophe Preparedness and Response, NYU, New York, 2009.

[2] A. F. Abate, M. Nappi, D. Riccio and G. Sabatino, "2D and 3D face recognition: A survey," *Pattern Recognition Letters,* vol. 28, no. 14, pp. 1885-1906, 2007.

[3] J. P. Phillips, P. Grother, R. J. Michaels, D. M. Blackburn, E. Tabassi and M. Bone, "FRVT 2002 Evaluation Report," 2003.

[4] J. P. Phillips, T. W. Scruggs, A. J. O'Toole, P. J. Flynn, K. W. Bowyer, C. L. Schott and M. Sharpe, "FRVT 2006 and ICE 2006 Large-Scale Experimental Results," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 32, no. 5, pp. 831-846, 2009.

[5] W. Y. Zhao and R. Chelappa, "SFS Based View Synthesis for Robust Face Recognition," in *IEEE International Conference on Automatic Face and Gesture Recognition*, 2000.

[6] X. Lu, R.-L. Hsu, A. K. Jain, B. Kamgar-Parsi and B. Kamgar-Parsi, "Face Recognition with 3D Model-Based Synthesis," in *International Conference on Biometric Authentication*, 2004.

[7] Y. Hu, D. Jiang, S. Yan, L. Zhang and H. Zhang, "Automatic 3D Reconstruction for Face Recognition," in *IEEE International Conference on Automatic Face and Gesture Recognition*, 2004.

[8] M. W. Lee and S. Ranganath, "Pose-Invariant Face Recognition Using A 3D Deformable Model," *Journal of Pattern Recognition,* vol. 36, no. 8, pp. 1835-1846, 2003.

[9] J. Huang, B. Heisele and V. Blanz, "Component-based Face Recognition with 3D Morphable Models," in *International Conference on Audio- and Video-Based Biometric Person Authentication*, 2003.

[10] U. Prabhu, J. Heo and M. Savvides, "Unconstrained Pose-Invariant Face Recognition Using 3D Generic Elastic Models," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 33, no. 10, pp. 1952-1961, 2011.

[11] V. Blanz and T. Vetter, "Face Recognition Based on Face Recognition Based on Fitting a 3D Morphable Model," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 25, no. 9, pp. 1063 - 1074 , 2003.

[12] B. Amberg, S. Romdhani and T. Vetter, "Optimal Step Nonrigid ICP Algorithms for Surface Registration," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2007.

[13] S. Chang, M. Rioux and J. Domey, "Face Recognition with Range Images and Intensity Images," *Optical Engineering,* vol. 36, no. 4, pp. 1106-1112, 1997.

[14] D. Huang, M. Ardabilian, Y. Wang and L. Chen, "Automatic Asymmetric 3D-2D Face Recognition," in *Three-dimensional face recognition using geometric model*, 2010.

[15] X. Lu and A. K. Jain, "Matching 2.5D Face Scans to 3D Models," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 28, no. 1, pp. 31-43, 2006.

[16] F. Tsalakanidou, S. Malassiotis and M. G. Strintzis, "Integration of 2D and 3D Images for Enhanced Face Authentication," in *IEEE International Conference on Automatic Face and Gesture Recognition*, 2004.

[17] J. Kittler, A. Hilton, M. Hamouz and J. Illingworth, "3D Assisted Face Recognition: A Survey of 3D Imaging, Modelling and Recognition Approaches," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition - Workshops* , 2005.

[18] S. Malassiotis and M. G. Strintzis, "Pose and Illumination Compensation for 3D Face Recognition," in *International Conference on Image Processing (ICIP)*, 2004.

[19] J. P. Phillips, P. J. Flynn, T. Scruggs, K. W. Bowyer, J. Chang, K. Hoffman, J. Marques, J. Min and W. Worek, "Overview of the Face Recognition Grand Challenge," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2005.

[20] M. Husken, M. Brauckmann, S. Gehlen and C. v. d. Malsburg, "Strategies and Benefits of Fusion of 2D and 3D Face Recognition," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition - Workshops*, 2005.

[21] N. Erdogmus and J.-L. Dugelay, "Automatic Extraction of Facial Interest Points Based On 2D and 3D Data," in *Electronic Imaging Conference on 3D Image Processing (3DIP) and Applications*, San Francisco, California, 2011.

[22] X. Lu and A. K. Jain, "Multimodal Facial Feature Extraction for Automatic 3D Face Recognition," 2005.

[23] N. Bozkurt, U. Halici, I. Ulusoy and E. Akagunduz, "3D data processing for enhancement of face scanner data," in *Signal Processing and Communications Applications Conference, SIU*, Antalya, 2009.

[24] N. Erdogmus and J.-L. Dugelay, "An Efficient Iris and Eye Corners Extraction Method," in *Joint IAPR International Workshop on Structural, Syntactic, and Statistical Pattern Recognition*, Cesme, 2010.

[25] A. W. C. Liew, S. H. Leung and W. H. Lau, "Lip Contour Extraction Using a Deformable Model," in *International Conference onImage Processing, Proceedings*, 2000.

[26] Y. Yang, X. Wang, Qian Y. and S. Lin, "Accurate and real-time lip contour extraction based on constrained contour growing," in *Joint Conferences on Pervasive Computing (JCPC)*, 2009.

[27] X. Liu, Y.-M. Cheung, M. Li and H. Liu, "A Lip Contour Extraction Method Using Localized Active Contour Model with Automatic Parameter Selection," in *International Conference on Pattern Recognition*, 2010.

[28] J. R. Tena, M. Hamouz, A. Hilton and J. Illingworth, "A Validated Method for Dense Non-rigid 3D Face Registration," in *IEEE International Conference on Video and Signal Based Surveillance*, 2006.

[29] F. Lavagetto and R. Pockaj, "The Facial Animation Engine: Toward A High-Level Interface For The Design Of Mpeg-4 Compliant Animated Faces," *IEEE Transactions on Circuits and Systems for Video Technology,* vol. 9, no. 2, pp. 277-289, 1999.

[30] V. Kuriakin, T. Firsova, E. Martinova, O. Mindlina and V. Zhislina, "Mpeg-4 Compliant 3d Face Animation," in *International Conference on Computer Graphics*, 2001.

[31] I. S. Pandzic and R. Forchheimer, Mpeg-4 Facial Animation: The Standard, Implementation And Applications, New York: John Wiley & Sons, 2003.

[32] F. L. Bookstein, "Principal Warps: Thin-Plate Splines and the Decomposition of Deformations," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 11, no. 6, pp. 567-585, 1989.

[33] X. Lu and A. K. Jain, "Deformation Analysis for 3D Face Matching," in *IEEE Workshops on Application of Computer Vision*, 2005.

[34] A. Savran, N. Alyuz, H. Dibeklioglu, O. Celiktutan, B. Gokberk, B. Sankur and L. Akarun, "Bosphorus Database For 3D Face Analysis," in *COST Workshop on Biometrics and Identity Management*, 2008.

[35] P. N. Belhumeur, J. P. Hespanha and D. J. Kriegman, "Eigenfaces Vs. Fisherfaces: Recognition Using Class Specific Linear Projection," *EEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 19, no. 7, pp. 711-720, 1997.

[36] T. Ojala, M. Pietikainen and T. Maenpaa, "Multiresolution Gray-scale and Rotation Invariant Texture Classification with Local Binary Patterns," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 7, 2002.

[37] P. Szeptycki, M. Ardabilian and L. Chen, "A Coarse-To-Fine Curvature Analysis-Based Rotation Invariant 3d Face Landmarking," in *International Conference on Biometrics: Theory Applications and Systems*, 2009.

[38] T. Maurer, D. Guigonis, I. Maslov, B. Pesenti, A. Tsaregorodtsev, D. West and G. Medioni, "Performance of Geometrix ActiveID 3D Face Recognition Engine on the FRGC Data," in *Computer Vision and Pattern Recognition - Workshops*, 2005.

[39] T. Sim, S. Baker and M. Bsat, "The CMU pose, illumination, and expression (PIE) database," in *Automatic Face and Gesture Recognition*, 2002.

[40] I. Mpiperis, S. Malassiotis and M. G. Strintzis, "Bilinear elastically deformable models with application to 3d face and facial expression recognition," in *Automatic Face and Gesture Recognition*, 2008.

Nesli Erdogmus graduated from Electrical and Electronics Enginering Department of Middle East Technical University (METU), Ankara, Turkey in 2005. She received her master's degree from the same department in 2008. She completed her Ph.D. at Eurecom, Sophia-Antipolis, France, in the Multimedia Communication Department under the supervision of Prof. Jean-Luc Dugelay in 2012.

During her master studies, she worked as a Research Assistant in Electrical and Electronics Department of METU, in the Computer Vision and Intelligent Systems Laboratory. Her industrial experiences include Test Engineering at AYESAS and Systems Engineering at SDT. While pursuing her Ph.D. degree, her research was part of an ANR project, FAR 3D.

Currently, she continues her research on utilization of 3D data in face recognition as a post-doc at Idiap Research Institute.

Jean-Luc Dugelay (M'76–SM'81-F'11) obtained his PhD in Information Technology from the University of Rennes in 1992. His thesis work was undertaken at CCETT (France Télécom Research) at Rennes between 1989 and 1992. He then joined EURECOM in Sophia Antipolis where he is now a Professor in the Department of Multimedia Communications. His current work focuses in the domain of multimedia image processing, in particular activities in security (image forensics, biometrics and video surveillance, mini drones), and facial image processing.

He has authored or co-authored over 235 publications in journals and conference proceedings, 1 book on 3D object processing published by Wiley, 4 book chapters and 3 international patents. His research group is involved in several national and European projects. He has delivered several tutorials on digital watermarking, biometrics and compression at major international conferences such as ACM Multimedia and IEEE ICASSP. He participated in numerous scientific events as member of scientific technical committees, invited speakers or session chair. He is a fellow member of IEEE and an elected member of the EURASIP BoG.

Jean-Luc Dugelay is (or was) associate editor of several international journals (IEEE Trans. On IP, IEEE Trans. On MM) and is the founding editor-in-chief of the EURASIP journal on Image and Video Processing (SpringerOpen). Jean-Luc DUGELAY is co-author of several conference articles that received an IEEE award in 2011 and 2012. He co-organized the 4th IEEE International Conference on Multimedia Signal Processing held in Cannes, 2001 and the Multimodal User Authentication held in Santa Barbara, 2003. In 2015, he will serve as general co-chair of IEEE ICIP and EURASIP EUSIPCO.