



PROFILE-BASED 3D-AIDED FACE RECOGNITION*

B. Efray, E. Ismailov, S. Shah, and I. A. Kakadiaris

Department of Computer Science
University of Houston
Houston, TX, 77204, USA
<http://www.cs.uh.edu>

Technical Report Number UH-CS-10-08

September 15, 2010

Keywords: Profile-based face recognition, Active Shape Model, facial data, 3D, multi-frame recognition, multi-modal vision

Abstract

This paper presents a framework for fully automatic face recognition based on a silhouetted face profile. Previous research has demonstrated the high discriminative potential of this biometric. However, for the successful employment of this characteristic, one is confronted with many challenges, such as the sensitivity of a profile's geometry to face rotation and the difficulty of accurate profile extraction from images. Compared to traditional approaches in profile-based recognition, our approach is not limited to only standard side-view faces. We propose to explore the feature space of profiles under various rotations with the aid of a 3D face model. In the enrollment mode, 3D data of subjects are acquired and used to create profiles under different rotations. The features extracted from these profiles are used to train a classifier. In the identification mode, the profiles are extracted from side view images using a modified Active Shape Model approach and its metadata is matched with gallery metadata. We validate the accuracy of the extractor and the robustness of classification algorithms using data from publicly available databases.



*This work was funded in part by the Central Intelligence Agency under the DISA ENCORE contract DCA200-02-D-5014 with Unisys Corporation serving as the primary on behalf of the Government.

PROFILE-BASED 3D-AIDED FACE RECOGNITION*

B. Efray, E. Ismailov, S. Shah, and I. A. Kakadiaris

Abstract

This paper presents a framework for fully automatic face recognition based on a silhouetted face profile. Previous research has demonstrated the high discriminative potential of this biometric. However, for the successful employment of this characteristic, one is confronted with many challenges, such as the sensitivity of a profile’s geometry to face rotation and the difficulty of accurate profile extraction from images. Compared to traditional approaches in profile-based recognition, our approach is not limited to only standard side-view faces. We propose to explore the feature space of profiles under various rotations with the aid of a 3D face model. In the enrollment mode, 3D data of subjects are acquired and used to create profiles under different rotations. The features extracted from these profiles are used to train a classifier. In the identification mode, the profiles are extracted from side view images using a modified Active Shape Model approach and its metadata is matched with gallery metadata. We validate the accuracy of the extractor and the robustness of classification algorithms using data from publicly available databases.

Index Terms

Profile-based face recognition, Active Shape Model, facial data, 3D, multi-frame recognition, multi-modal vision

I. INTRODUCTION

The development of face recognition algorithms has been an active area of research since the rise of the Computer Vision field. Numerous algorithms have been proposed [1], [2], and several commercial products are available and widely used for computer-aided identification. However, most of these methods and products are designed to work with near-frontal face images. The face profile provides complementary shape structure that is not visible in a frontal view, but is often available from other views. According to a study by Davidenko [3], silhouetted face profiles play an important role in human perception for the task of identity and gender recognition. Use of the face profile for biometrics is especially attractive for scenarios where only side view images are available. Figure 1 depicts exemplar images where the silhouetted profile contains more reliable information than texture, which makes profile information especially valuable for recognition.

Profile-based recognition presents several challenges (e.g., pose variation, occlusions, and facial expressions). The face profile contains only a small, compact amount of the face shape information; therefore, it is very sensitive to a loss in precision, which is unavoidable due to the limited accuracy of sensors and the discrete nature of digital representations.

Until recently, research in profile-based recognition was based on comparison of *standard profiles* – the contours of side view images with yaw very close to -90° . Research in 3D-3D face recognition has indicated that the profile information contains highly discriminative information [4], [5], [6], where the term “profile” is often associated with the facial area along the symmetry axis of the 3D face model. However, neither approach is capable of accurate modeling of a silhouetted face profile, as observed in a 2D image because (i) the face is not perfectly symmetric, (ii) the face is almost never at yaw equal to -90° with respect to the sensor, and (iii) if the distance between camera and object is not sufficiently large, perspective projection needs to be considered (based on imaging sensor parameters). In this paper, the term “profile” always indicates the silhouette of nearly side view head images for clarity of presentation.

Recent advances in 3D surface reconstruction methods based on both hardware and software solutions [7] have made 3D face shape information readily available. This information has been employed with 2D images in order to overcome challenges such as lighting, pose variations, and facial expressions [8], [9]. An accurate 3D model provides

*This work was funded in part by the Central Intelligence Agency under the DISA ENCORE contract DCA200-02-D-5014 with Unisys Corporation serving as the primary on behalf of the Government.



Fig. 1: Examples of side view images where reliable color texture is not available: binary image, camouflage, backlighting, and near infrared imaging.

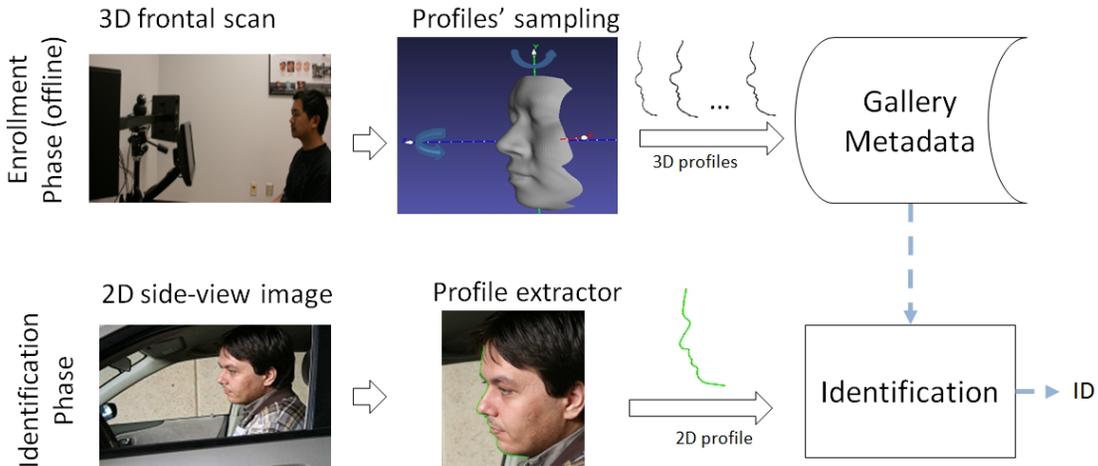


Fig. 2: Enrollment and identification phases of the proposed integrated profile-based face recognition system.

shape information (not necessarily on the symmetry axis) that is highly valuable for modeling of the profile shape, therefore by employing the tools from 3D face analysis one can improve performance of profile-based recognition.

In this work, we specifically address the challenges of profile contour extraction from an image and those presented by recognition under pose variation. The central idea of our approach is to use 3D face models to explore the feature space of a profile under various rotations. An accurate 3D model embeds information about possible profile shapes in the probe 2D images, which allows flexibility and control over the training data. We suggest that sufficient sampling in the pose space, which corresponds to nearly side-view face images, provides robustness for a recognition task. Specifically, we propose to generate various profiles using rotations of a 3D face model. The profiles are used to train a classifier for profile-based identification.

Two different types of profiles are therefore being employed in our system: (i) *3D profiles* – those generated synthetically through 3D face models to be used as training data, and (ii) *2D profiles* – those extracted from 2D images of side-view faces.

The schematic illustration of the profile-based face recognition system (URxD-PV) is depicted in Fig. 2 and includes *Enrollment* and *Identification* phases. We provide algorithmic solutions for the entire 3D-aided profile-based recognition framework including profile modeling, landmark detection, shape extraction, and classification.

During the enrollment phase, the exact geometric method for 3D profile computation is applied on a triangular mesh, which is preregistered to a common reference model [9]. Classifier-specific features are extracted from these profiles and constitute the gallery metadata. As part of the identification phase, we present an algorithm for 2D profile extraction from a side-view image. It is based on an adaptation of the Active Shape Model [10], which does not require a manually labeled set of exemplar images in the modeling stage.

We compare two different approaches for profile identification: the first one based on matching 3D and 2D profiles represented explicitly using modified Hausdorff distance, the second method employing a hierarchy of SVM classifiers applied on rotation-, translation-, and scale-invariant features extracted from profiles.

The contributions of our work include (i) an adaptive and accurate algorithm for profile extraction, (ii) a robust

method for the extraction of fiducial points, and (iii) robust classifier for a range of poses. The system is fully automatic and can be used in applications related to access control.

The rest of this paper is organized as follows: Section II reviews previous work, Section III describes methods used for profile representation and modeling using 3D data. Section IV introduces the procedure for the profile extraction from 2D images. Section V concentrates on identification approaches. Section VI discusses components of a fully automatic face recognition system based on our methodology. Section VII presents a performance evaluation using extensive and publicly available databases that contain single and multi-frame probe datasets from unconstrained and constrained environments and multiple modalities. Section VIII provides a summary of the main contributions.

II. RELATED WORK

The use of face profile for identification had attracted research interest even before the arrival of the associated computer technologies [14]. This interest has increased over the last decade due to exploration of 3D-aided face recognition. The methods for recognition using the profile curve can normally be classified into one of two categories: *landmark-based* methods [15], [16], [17], [18] or *global* methods [11], [19], [20], [21]. Landmark-based methods rely on the attributes associated with a set of fiducial points, and recognition uses similarity metrics based on those attributes. Global methods consider each profile as a geometrical object and introduce a similarity metric between homogeneous objects: all regions of a profile are treated equally.

Harmon *et al.* [15] defined 17 fiducial points; after aligning two profiles based on selected landmarks, the matching was achieved by measuring the Euclidean distance of the feature vectors derived from the outlines. A 96% recognition rate was reported. Wu *et al.* [16] used a B-spline to find six landmarks and extracted 24 features from the resulting segments. Liposcak and Loncaric [17] used scale-space filtering to locate 12 landmarks and extracted 21 distances based on those landmarks. The Euclidean distance between the vectors of features was used for the identification.

Bhanu and Zhou [20] proposed curvature-based matching using a dynamic warping algorithm. They reported a recognition rate of almost 90% on the University of Bern Database of 30 subjects. Gao and Leung [22] introduced a method to encode profiles as attributed strings and developed an algorithm for attributed string matching. They reported nearly 100% recognition rate on the same database. Pan *et al.* [19] proposed to use metrics for the comparison of probability density functions on properly rotated and normalized profile curves. Gao *et al.* [21], [23] proposed new formulations of the Hausdorff distance. Initially, it was extended to match two sets of lines, while later, it was based on weighting points by their significance. In both cases, they applied their distance metric to measure the similarity of face profiles.

As we mentioned earlier, all these methods are designed for standard profiles only and use 2D images as gallery. Kakadiaris *et al.* [11] introduced the use of a 3D face model for the generation of profiles under different poses for the gallery. Modified directional Hausdorff distance of the probe profile to the gallery profile was used for identification. In addition, four different profiles under various rotation angles were used in order to introduce robustness to pose.

An important step in the implementation of a fully automatic system suitable for unconstrained scenarios is developing an accurate profile extractor. The majority of profile-based identification research does not sufficiently address this issue: instead they use manual extraction [11], [19] or very basic thresholding methods based on the assumption of indoor controlled illumination and uniform background [24], [17], [20].

Much more efficient methods have been applied for near-frontal face extraction and feature localization. Among the most powerful are the methods based on the Active Shape Model (ASM), originally proposed by Cootes *et al.* [10]. These methods are based on recovering parameters of a statistical shape model, when a local minimum of the matching energy is found based on a search in local neighborhoods of the shape points. During the last decade, numerous modifications for the ASM have been proposed. Gu and Kanade [25] successfully incorporated the model into their 3D alignment algorithm, where view-based patches are used to model appearance of shape points for the local search, and an Expectation Maximization algorithm is applied to deform and fit the model to the image iteratively. The training stage employs about 50,000 images. Mahoor and Abdel-Mottaleb [26] incorporated color information and an additional alignment step for accurate frontal face landmarking. The final goal for most of these algorithms is the alignment, therefore the shape is mostly defined by sparse set of common face landmarks

visible on the frontal view, enforced by only a few additional points. For the contour extraction task, points should be dense in order to approximate the curve accurately. Another known shortcoming of the ASM approach is the sensitivity for initialization, which is especially critical for ridge-like shapes.

III. PROFILE REPRESENTATION AND MODELING

We start this section with a description of the mathematical model used to represent the profile shape and then describe the procedure of the profile sampling in the pose space using a 3D model.

A. Profile Representation

In our approach, we treat the profile as an open curve. The profile curve \mathcal{C} may be described by a pair of arc-length parameterized 1D functions $Y_{\mathcal{C}}(l)$ and $X_{\mathcal{C}}(l)$, where $l \in [0, 1]$. A set of k landmarks is defined by their coordinates on a parametric curve: $\{0 = \nu^1 < \dots < \nu^k = 1\}$. The set contains both anatomical landmarks (e.g., “chin”) and pseudo-landmarks (e.g., “middle of the nose”). The complete set of landmarks is illustrated in Fig. 5(a). We approximate functions $Y_{\mathcal{C}}(l)$ and $X_{\mathcal{C}}(l)$ by a finite set of points and obtain an equivalent n -points shape model as follows:

$$\mathbf{v} = [x_1, y_1, x_2, y_2, \dots, x_n, y_n]^T \in \mathbb{R}^{2n}. \quad (1)$$

The positions of the points are obtained through uniform arc-length sampling of the curve between a predefined subset of the landmarks. The sampling pattern is consistent for all profiles and, therefore, coordinates of these landmarks always preserve their indices.

B. Generating 3D profiles

Let’s assume that a 3D model has been aligned to a common reference model and does not contain any artifacts (e.g., holes or unconnected components). In order to model the variations in the geometry of profiles caused by head rotations, we generate profiles corresponding to a range of rotation angles. Our method is based on geometric transformations and the perspective projection of a 3D surface with respect to camera parameters. This method preserves information about the vertices on the profile using floating point precision (rather than pixel precision) and has linear run-time complexity. Algorithm 1 describes the proposed method and uses following notations:

$\{f_i\}$ is the set of faces of the triangular polygon mesh and $\{\mathbf{n}_i\}$ their corresponding normal vectors.

$e(i, j)$ is the edge shared by faces f_i, f_j , which can also be defined by its vertices $e \equiv (v_k, v_l)$. Each vertex is associated with coordinates in the 2D projection plane.

\mathbf{o} is the vector that defines the position of a 3D object with respect to the observer.

The resulting profile contains a chain of line segments. We use spline interpolation to introduce smoothness and perform uniform resampling of the spline curve by arc-length.

C. Landmark Detection

The main principles of landmark extraction are well established for standard profiles [17], [24]. Those methods are usually driven by the extremum locations and the curvature of the function, and they are subject to anthropometric standards of the face. For example, for an upward facing profile, the tip-of-nose is the extremum point of $Y_{\mathcal{C}}(l)$ with $l \in [0.4, 0.7]$. However, the accuracy of such rules is very low when applied to general profiles. In this paper, we use both landmark information from standard profiles and a 3D face surface to accomplish the localization of landmarks. Specifically, we propose to utilize the 3D surface in order to extrapolate information about landmarks on a standard profile to any other profile obtained with known rotation angles. The method is described in Algorithm 2.

The final step in Algorithm 2 is a refinement step and is based on the elaboration of general landmark definition. The position of the landmarks, whether located by an automatic procedure or a human operator, has a certain amount of ambiguity. For instance, the position of a landmark that is defined by the curvature value is equivocal in the region with constant curvature (“ball-shaped”). Consider the curve in Fig. 4(a): it has nearly the same curvature for the region between points C and D (both of these points correspond to the zeros of second derivative), whereas point E corresponds to the zero of the first derivative and is sensitive to rotations. In order to reduce such ambiguity, we define a refinement operator. This operator finds the most protruding (or intruding) point in the region.

Algorithm 1 Generating a 3D profile

Input: $\{n_i\}, \{v_i\}, \{e_i\}$ – sets of normals, vertices and edges, \mathbf{o} – observer vector

Output: Profile \mathcal{C}

1: Set $\mathcal{C} = \emptyset$. Compute a critical set of edges \mathcal{E} :

$$\mathcal{E} = \{e(i, j) : \text{sign}(\langle \mathbf{n}_i, \mathbf{o} \rangle) \neq \text{sign}(\langle \mathbf{n}_j, \mathbf{o} \rangle)\}.$$

2: Compute all the intersections between line segments defined by \mathcal{E} in the projection plane using the plane sweep algorithm [27].

3: **for all** intersecting edges $(v_1, v_2), (v_3, v_4)$ with intersection p **do**

4: Split $(v_1, v_2), (v_3, v_4)$ into 4 edges $(v_1, p), (p, v_2), (v_3, p), (p, v_4)$. Denote the modified set of edges by $\tilde{\mathcal{E}}$ and set of corresponding endpoint vertices by $\tilde{\mathcal{V}}$.

5: **end for**

6: **for all** $v \in \tilde{\mathcal{V}}$ **do**

7: Create circular list \mathcal{L}_v of edges with endpoint in v ordered clockwise.

8: **end for**

9: Find u, a – vertices with lowest and highest Y coordinate in $\tilde{\mathcal{V}}$.

10: Find edge $(u, v) \in \mathcal{L}_u$ inducing the smallest angle with vector in direction $[-1, 0]$.

11: **repeat**

12: $\mathcal{C} \leftarrow \mathcal{C} \cup (u, v)$.

13: Let (v, w) to succeed (u, v) in \mathcal{L}_v .

14: $u \leftarrow v, v \leftarrow w$.

15: **until** $v \neq a$

Algorithm 2 Detecting landmarks on a rotated profile

Input: \mathcal{G} – 3D surface, $\mathcal{C}_0, \mathcal{C}$ – standard and rotated profiles, $\{\nu_0^1, \dots, \nu_0^k\}$ – landmarks on \mathcal{C}_0 , rotation and projection parameters.

Output: $\{\nu^1, \dots, \nu^k\}$ – landmarks on \mathcal{C}

1: Back-project landmarks from \mathcal{C}_0 onto a center line of \mathcal{G} (Fig. 3(a)).

2: **for all** landmarks i with coordinate $(x_i, y_i, z_i) \in \mathcal{G}$ **do**

3: Define a subset of points in ϵ_i -neighborhood of (x_i, y_i, z_i) :

$$\mathcal{N}_i = \{(x, y, z) \in \mathcal{G} \mid \| (x, y, z)^T - (x_i, y_i, z_i)^T \| \leq \epsilon_i\} \text{ (Fig. 3(b)).}$$

4: Project \mathcal{N}_i onto the profile plane using rotation and projection parameters, thus obtaining a set of 2D points $\tilde{\mathcal{N}}_i = \{(\tilde{x}, \tilde{y})\}$ located from one side of \mathcal{C} .

5: Find coordinate ν^i on profile \mathcal{C} with minimal distance to $\tilde{\mathcal{N}}_i$ (Fig. 3(c)).

6: **end for**

7: Refine all landmark coordinates based on neighbor landmarks depicted in Fig. 4(b) and Equation (2) (Fig. 3(d)).

Let i, j be the indices of two landmarks on a profile's curve and ν^i, ν^j denote their corresponding arc-length coordinates. For any coordinate ν , the function $H(\nu^i, \nu^j, \nu)$ defines the signed distance between the vector connecting the points of ν^i, ν^j and the point corresponding to ν . The sign of $H(\nu^i, \nu^j, \nu)$ denotes whether the point lies on the left or the right side of the vector. The refined coordinate is given by:

$$\nu^* = \underset{\{\nu \mid \nu^1 \leq l_1 < \nu < l_2 \leq \nu^2\}}{\text{argmax}} (\gamma \cdot H(l_1, l_2, \nu)). \quad (2)$$

The parameter γ is defined to be 1 (or -1) according to the convexity (or concavity) of the profile curve in the neighborhood of the landmark. Figure 4(a) illustrates the effect of this operator, which results in point F . We apply this refinement operator to all anatomical landmarks, except the chin in relation to their neighboring landmarks. The ‘‘chin’’ landmark is refined differently: by finding the most distant point ‘visible’ from the tip-of-nose (Figure

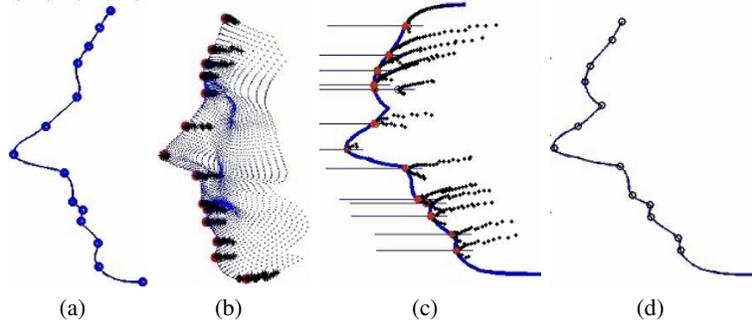


Fig. 3: Profile landmarks' localization. Depiction of (a) landmarks on the standard profile, (b) landmarks projected on a 3D model (red markers) and extensions of landmarks to 3D regions (black markers), (c) projection of landmarks on a rotated (not standard) profile, and (d) result of the refinement procedure.

4(b)). This simple and fast operator may be efficiently applied for landmark refinement on any profile, even after rough initialization (Fig. 4(c)).

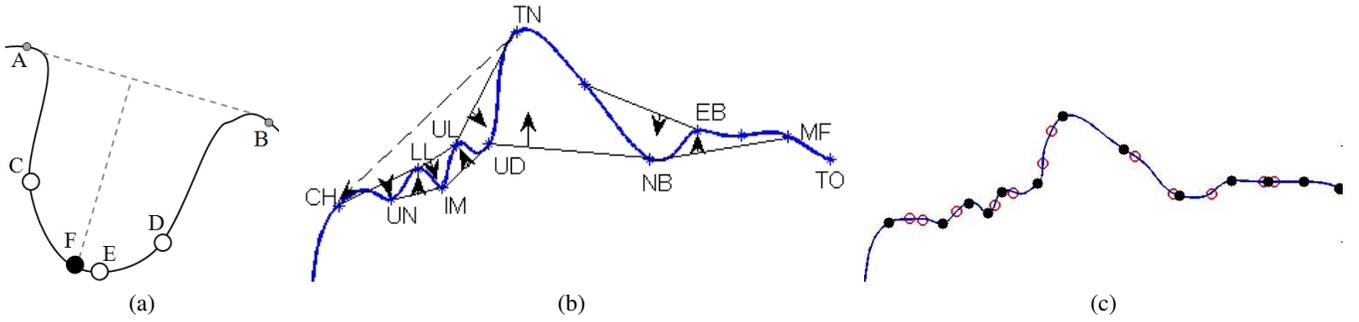


Fig. 4: Landmark refinement. Depiction of (a) geometric entities related to Equation (2); (b) neighbor relations used for the refinement of landmarks; and (c) example of refinement output (discs) compared to input before refinement (circles).

IV. PROFILE EXTRACTION FROM 2D IMAGES

The procedure of profile detection from 2D images is applied to a region of interest in the image (result of Face Detection). The extraction procedure is guided by specific pixel features and based on a statistical shape model, both of which are explained in the next sections.

A. Computation of Features to Guide Shape Extraction

This step allows the incorporation of any available knowledge about the appearance of the shape in the probe image. In an attempt to avoid the application being biased by a specific training set, we simplify our knowledge about the face profile using two basic observations: (i) profile points are mostly located on the separation line between face region and background region, and (ii) profile points are formed by edges.

Let $B_e(x, y)$ be the binary map of the image edges (Fig. 6(b)) and let $B_s(x, y)$ be the binary function corresponding to the left boundary of the skin segmentation result (Fig. 6(a)). The information embedded in B_s typically has low precision and will be used for a coarse scale search, while the edges in $B_e(x, y)$ reflect fine scale details, but may contain a considerable number of edges not related to the profile. Our variant of the edge detector is based on the incorporation of a multiscale normalized cut algorithm [28] for the edge detecting operator, whereas the facial segmentation relies on a mean-shift algorithm [29] and incorporates knowledge about color, shape, size, and position of the face in the region of interest. The details about these operators may be found in [30].

We estimate the principle orientation at every edge pixel by convolving the edge map with eight structuring elements M_c , each representing one of the principal orientations $c = 1, \dots, 8$:

$$\Upsilon_e(x, y) = \underset{c}{\operatorname{argmax}} (B_e * M_c)(x, y).$$

Then, we apply a distance transform mapping to acquire the final set of pixel-related features used by the ASM framework. Namely, we associate each pixel (x, y) with $G_s(x, y)$ and $\{G_e^c | c = 1, \dots, 8\}$, where G_s represents for the distance to the nearest pixel on the skin contour and G_e^c represents the distance to the nearest edge pixel with the principle orientation c .

B. Statistical Shape Model

In this section, first we review the standard linear deformable model also known as a *Point Distribution Model* (PDM) and then we introduce our generalization for it, used for profile extraction. Using a data set of m exemplar 3D profiles $\{\mathbf{v}_1, \dots, \mathbf{v}_m\}$, the profile \mathbf{v} is represented using Principal Component Analysis (PCA) as follows:

$$\mathbf{v} = \bar{\mathbf{v}} + \sum_{i=1}^{2n} b_i \phi_i. \quad (3)$$

$\bar{\mathbf{v}}$ is the mean shape and ϕ_i are the eigenvectors of the covariance matrix Σ :

$$\Sigma = \frac{1}{m-1} \sum_{i=1}^m (\mathbf{v}_i - \bar{\mathbf{v}})(\mathbf{v}_i - \bar{\mathbf{v}})^T. \quad (4)$$

To compensate for the shape variance related to rotation, translation and scale parameters, the profiles are aligned in advance with a normalized reference shape using generalized Procrustes analysis [31]. The eigenvectors are sorted according to descending order of corresponding eigenvalues $\lambda_1 \geq \dots \geq \lambda_{2n}$. In practice, in (3) one needs to retain only h modes ($h < 2n$) in the representation $\mathbf{b} = \mathbf{b}(\mathbf{v}) = [b_1, \dots, b_h]^T$ to explain a given proportion of the variance in the training set:

$$\mathbf{v} \approx \mathbf{v}(\mathbf{b}) = \bar{\mathbf{v}} + \sum_{i=1}^h b_i \phi_i. \quad (5)$$

Due to the orthogonal properties of the decomposition, the vector of parameters \mathbf{b} may be found by simple projection:

$$\mathbf{b} = \Phi^T (\mathbf{v} - \bar{\mathbf{v}}), \quad (6)$$

which guarantees the least squares solution

$$\min (\mathbf{v} - \bar{\mathbf{v}} - \Phi \mathbf{b})^T (\mathbf{v} - \bar{\mathbf{v}} - \Phi \mathbf{b}), \quad (7)$$

where Φ is a matrix with eigenvectors as columns.

Equations (6) and (3) are used to replace the n -point shape by its representation in PCA basis (analysis) and vice-versa (synthesis). It is common practice to apply (6) followed by (5) to obtain an approximation of the n -point shape represented by h variation modes; such procedure is useful for *shape reconstruction* from noisy observations. In certain cases, we want to associate varying validity weights w_i for each point of the shape \mathbf{v} . If $W = \operatorname{diag}(w_1^2, \dots, w_{2n}^2)$, then the weighted least square variant of (7) is:

$$\min (\mathbf{v} - \bar{\mathbf{v}} - \Phi \mathbf{b})^T W (\mathbf{v} - \bar{\mathbf{v}} - \Phi \mathbf{b}), \quad (8)$$

and its solution is known to be:

$$\mathbf{b} = (\Phi^T W \Phi)^{-1} \Phi^T W (\mathbf{v} - \bar{\mathbf{v}}). \quad (9)$$

One apparent disadvantage of the parametric representation in (5) when compared with the n -point-shape model is the absence of local information. The information about a small region of the profile is reflected in all b_i coefficients. In order to isolate the mutual influence of distinct regions of the profile, we define a subdivision $\mathcal{S} = \{s_1, \dots, s_d\}$

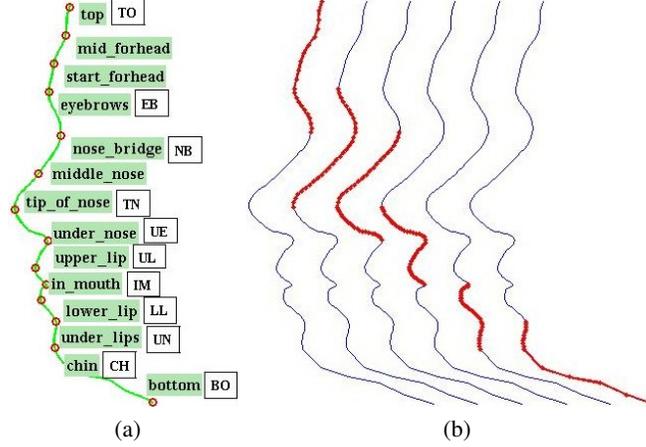


Fig. 5: Depiction of (a) profile landmarks and their names and (b) partition of the profile in sectors with overlapping regions.

of the profile shape in d sectors. Each sector s_j is defined by two indices, denoting the starting and the ending index of the points in the n -point shape. In the most general case, sectors may have overlapping regions (Fig. 5(b)). The n -point shape representation is obtained from the sector-based deformable model by applying Eq. (5) independently to each region:

$$\mathbf{v} \approx \mathbf{v}(\mathbf{b}(\mathcal{S})) = \sum_{j=1}^d \left(\bar{\mathbf{v}}_j + \sum_{i=1}^{h_j} b_{ij} \phi_{ij} \right) \circ \boldsymbol{\eta}, \quad (10)$$

where, the subscript j relates the variable to the j^{th} sector. The dimensionality of vectors \mathbf{v} , $\bar{\mathbf{v}}$ and ϕ is the same as in (5), with zero entries outside the sector borders. The column vector $\boldsymbol{\eta} \in \mathbb{R}^{2n}$ contains regularization weights for points that belong to overlapping regions. Equation (5) is a special case of (10) with only one sector. The flexibility of a sector-based representation also allows excluding fragments.

To summarize, the gallery metadata includes parameters of a standard linear deformable model: $\hat{\mathbf{v}}, \Phi, \{\lambda_i\}$ and its sector-based variant: $\{\hat{\mathbf{v}}_j\}, \{\Phi_j\}, \{\lambda_{ij}\}, \mathcal{S}$. These components together represent two levels of hierarchical information.

C. Shape Extraction

The profile extraction is based on the Active Shape Model paradigm developed by Cootes *et al.* [10]. It uses the following iterative approach to improve the fit of a given instance of the n -point shape to an image:

- 1) Find the best match for each point (x_i, y_i) in its neighborhood (we name this step ‘Local Search’);
- 2) Update parameters to best fit the shape to these points (we name this step ‘Coefficient Estimation’); and
- 3) Repeat until convergence.

The local search neighborhood is usually defined along the normal to the profile curve through the point under inspection, as illustrated in Fig. 6(c). The range of search is gradually reduced through iterations to ensure faster convergence.

In order to address occlusions, we define a set of weighting parameters $w_i \in [0, 1]$, $i = 1, \dots, n$ as the local quality of profile point i . Ideally this value approaches 0 if the point is occluded or, alternatively, it approaches 1 if the point is reliable.

In the classical ASM framework, one needs to have a manually labeled set of training images. This dataset is used for the construction of the PDM. The same dataset is employed to design the features to guide the local search, typically by computing a pixel’s likelihood belonging to a shape.

In our case, no such labeled training set is available. Therefore, the main difference of our ASM framework from the classical approach is the fact that we only use available 3D shape information to guide the search of 2D

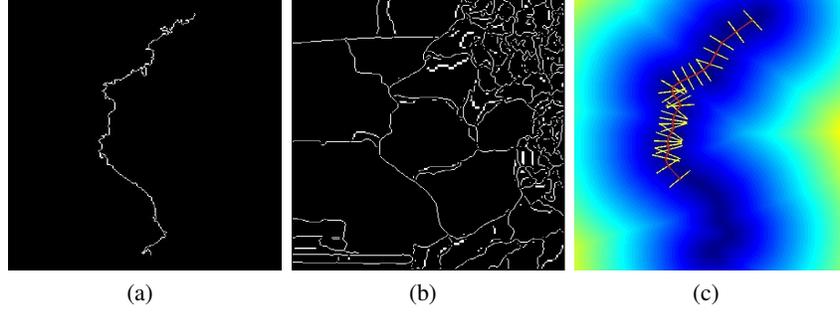


Fig. 6: Depiction of (a) left contour of the face region (B_s); (b) detected edges (B_e); and (c) the profile shape and search regions overlaid on the of distance map G_s .

profiles in the images: the PDM is based on principle component analysis of the 3D profile (see Section IV-B) and the local search is guided by features defined in Section IV-A.

Algorithm 3 Profile Extractor

Input: 2D image, profile PDM, profile sector-PDM

Output: Profile \mathbf{v}

- 1: Compute G_s and $\{G_e^c | k = 1, \dots, 8\}$.
- 2: Compute the set of tip-of-nose candidates (see Fig. 7).
- 3: **for all** tip-of-nose candidates **do**
- 4: Set iteration $t = 0$, initial weights $w_i = 1, \forall i \in 1, \dots, n$, shape $\mathbf{b}^{(t)} = 0$ (mean shape). Translate profile to initial position using the nose-tip.
- 5: Perform **ASM Iterations** in coarse search mode.
- 6: **end for**
- 7: Select the best quality profile and discard the remaining profiles.
- 8: Perform **ASM Iterations** in fine scale search mode.

ASM Iterations

Input: $G_s, \{G_e^c | c = 1, \dots, 8\}, \mathbf{b}^{(0)}$, search mode, profile sector-PDM

Output: $\mathbf{v}, \mathbf{b}^{(t)}, \{w_i\}$

- 1: **repeat**
 - 2: *Local search:* For each point i find best match (\hat{x}_i, \hat{y}_i) using (11).
 - 3: *Coefficient Estimation:* Compute parameters \mathbf{b} of the shape model using $\{(\hat{x}_i, \hat{y}_i), w_i^{(t)}\}$ and PDM using (9).
 - 4: *Profile quality estimation:* Estimate profile quality $Q(\mathbf{b})$ using (13).
 - 5: *Shrinkage:* If $Q(\mathbf{b}) > 3$ then adjust profile parameters toward the mean profile $\mathbf{b}^{(t+1)} = S(\mathbf{b})$ using (14).
 - 6: Reconstruct n -point shape using (10).
 - 7: *Re-weighting:* Compute new weights $\{w_i^{(t+1)}\}$ using (15).
 - 8: $t \leftarrow t + 1$.
 - 9: **until** Convergence of $\mathbf{b}^{(t)}$
-

Figure 8 depicts the shape propagation of the extractor. The process for the profile extraction is described in Algorithm 3, and it relies on the following considerations.

Coarse and Fine Search Modes: The main ASM iteration is first executed for the coarse scale search and then for the fine scale search with slightly different settings. The main differences between modes are:

- 1) Employ G_s for coarse search and G_e^k for fine search.
- 2) Use simplified profile shape (with half the number of points) for coarse search.
- 3) Apply profile sector-PDM-based shape reconstruction every second step in the fine search mode (but not in the coarse search mode).

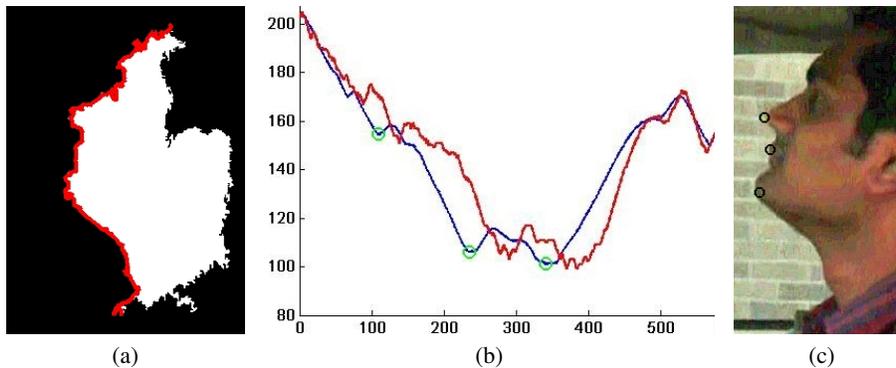


Fig. 7: Detecting candidates for the tip-of-nose landmark. Depiction of (a) result of segmentation and left contour (pink), (b) $X(i)$ function of the left contour (red) and its smoothed version (blue) with corresponding local minima (green circles), and (c) candidates for tip-of-nose landmarks back-projected on original image.

Our experiments showed that alternating between sector-based and non-sector based models is extremely beneficial for preserving fine detail reconstruction without violating proportions of the whole face.

Multiple Initializations: The ASM algorithm is known to be sensitive to initialization; to address this issue, we initialize the coarse search in several positions according to the candidate locations of the nose-tip. The candidates for the nose-tip landmark are defined as local extrema of the smoothed face skin contour (see Fig. 7).

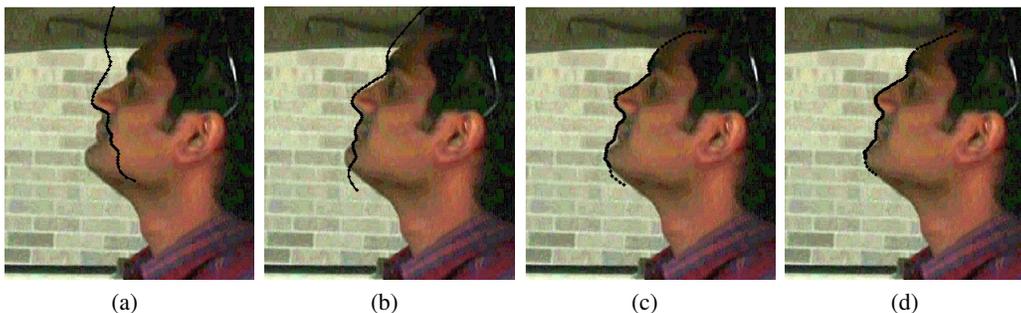


Fig. 8: Propagation of profile search. Depiction of (a) initial profile; (b) after two iterations; (b) after 5 iterations; and (c) final result.

Local Search: Let $\mathcal{R}_i \subset \mathbb{R}^2$ represent the local search domain for the model point i . The best matching coordinate (\hat{x}_i, \hat{y}_i) for the point i with current coordinate (x_i, y_i) is

$$(\hat{x}_i, \hat{y}_i) = \underset{(x,y) \in \mathcal{R}_i}{\operatorname{argmin}} G_s(x, y). \quad (11)$$

For the fine search, the principle orientation $k(i)$ of the point i on the shape is estimated and $G_e^{k(i)}$ is employed in (11) instead of G_s .

Profile Quality Estimation: The projection of local search results onto the parametric space does not always generate a plausible profile. We wish to estimate how far the parametric representation \mathbf{b} is from a distribution learned from training examples, which is assumed to be distributed as a diagonal Gaussian:

$$\beta \sim \mathcal{N}(\mathbf{0}, \Lambda), \Lambda = \operatorname{diag}(\lambda_1, \lambda_2, \dots, \lambda_t). \quad (12)$$

According to the properties of PCA, variances λ_i are identical to eigenvalues (see Section IV-B). The quality of any profile with parametric representation \mathbf{b} is based on the Mahalanobis distance which is related to the log-probability

then \mathbf{b} is drawn from the same distribution [32]. Thus, it is given by:

$$\log P(\beta = \mathbf{b}) \propto Q(\mathbf{b}) = \mathbf{b}\Lambda^{-1}\mathbf{b}^T. \quad (13)$$

Shrinkage: In order to correct non-plausible profiles, parameters \mathbf{b} are shrunk toward the mean shape along each principal direction. We use a similar strategy as proposed by Kanade and Gu [25] to correct by a larger amount in the directions that correspond to lower eigenvalues. However, unlike the method in [25], this operator is triggered only if the quality of the profile exceeds a predefined threshold of τ . The shrinkage result $S(b_i)$ of parameter b_i is given by

$$S(b_i) = b_i \cdot \gamma_i + (1 - \gamma_i) \frac{\tau\sqrt{\lambda_i}}{b_i}; \gamma_i = \frac{1}{i^2}; i = 1 \cdots h. \quad (14)$$

Re-weighting: The distance between the i^{th} point position resulting from local search (\hat{x}_i, \hat{y}_i) and the resulting position after shrinkage $(S(\hat{x}_i), S(\hat{y}_i))$, results in a fitting error $e_i = \|(\hat{x}_i, \hat{y}_i)^T - (S(\hat{x}_i), S(\hat{y}_i))^T\|$. Usually, outliers correspond to larger fitting errors and, therefore, weights w_i are adjusted accordingly:

$$w_i^{(t+1)} = \frac{1}{3}w_i^{(t)} + \frac{2}{3}\left(1 - \frac{e_i^2}{\sum_{k=1}^n e_k^2}\right)\rho, \quad (15)$$

where the parameter ρ is chosen so that all weights will sum to n .

V. IDENTIFICATION

In order to identify the extracted profile as one of the subjects in the gallery, we consider one of two general approaches: *Matching* or *Classification*. Both these approaches and the methods for multiple-frame identification are discussed in following sections.

A. Hausdorff Distance-based Matching

In the first approach, the matching scores between the probe profile and every profile in the gallery are computed. The decision is made according to the nearest neighbor rule. Identification through matching does not require any training stage, and is robust to outliers in the gallery and noisy probe data, but may be time consuming for large galleries, and has limited accuracy. The chosen matching score should possess some attributes of the distance metric. We propose to employ a modified Hausdorff distance for such distance metrics.

For two finite point sets $\mathcal{M} = \{m_1, \dots, m_n\}$ and $\mathcal{T} = \{t_1, \dots, t_n\}$ with associated weights $\{w_1^{\mathcal{M}}, \dots, w_n^{\mathcal{M}}\}$ and $\{w_1^{\mathcal{T}}, \dots, w_n^{\mathcal{T}}\}$, the distance is defined as:

$$\frac{1}{n} \max \left(h_{\mathcal{M}} \sum_{m_i \in \mathcal{M}} \min_{t_j \in \mathcal{T}} \|m_i - t_j\| \sqrt{w_i^{\mathcal{T}} w_j^{\mathcal{M}}}, h_{\mathcal{T}} \sum_{t_i \in \mathcal{T}} \min_{m_j \in \mathcal{M}} \|t_i - m_j\| \sqrt{w_j^{\mathcal{M}} w_i^{\mathcal{T}}} \right),$$

where \mathcal{M} and \mathcal{T} are probe and gallery n -point shapes and $h_{\mathcal{M}}$ and $h_{\mathcal{T}}$ are normalization factors of the distance between UN and NB landmarks (see Fig. 5(a)) to eliminate scale influence. The set of weights for a probe profile reflects the accuracy of a shape extractor (all equal 1 for manually extracted profiles). The set of weights for a gallery profile reflects a prior knowledge about discriminative properties of various regions.

B. SVM-based Classification

The classification approach takes advantage of machine learning techniques to define class boundaries in the feature space. Such a classifier may require an extensive training stage, but is expected to provide better recognition if both gallery and probe sets are drawn from the same distributions.

We define five types of features based on the properties of the profile p between two landmarks i and j : (i) Euclidean distance between landmarks, (ii) arc-length between landmarks, (iii) mean curvature of the region between landmarks, (iv) L_2 -norm of curvature along the contour between landmarks (proportional to *bending energy*), and (v) L_2 -norm of angular chain coding for the region between landmarks under the assumption of uniform discretization. All of these features are already translation- and rotation-invariant. Normalizing the distances with respect to the size of the profile provides scale-invariance as well. Currently, we use forward selection and 2-fold cross validation for the attribute selection based on the training examples and the *Support Vector Machine* (SVM) classifier.

This assumption however, is not true in our case.

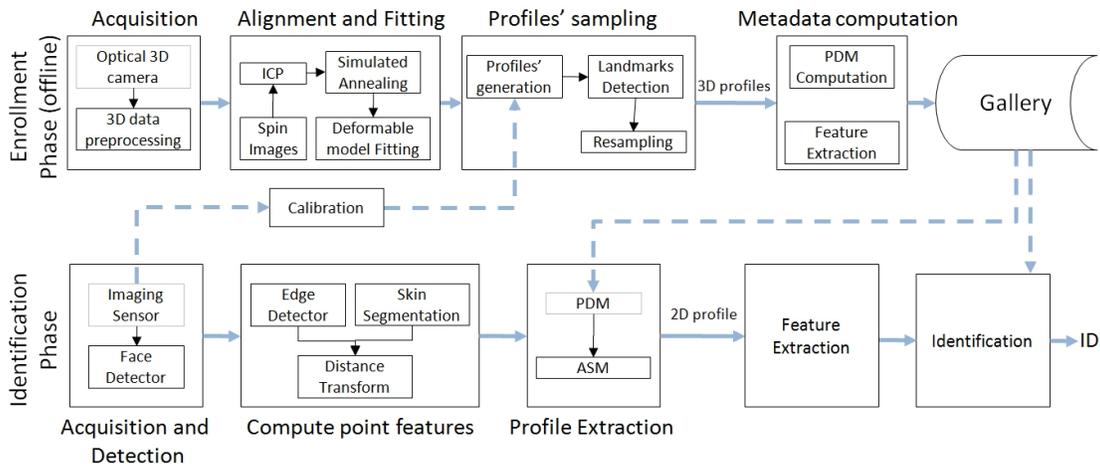


Fig. 9: The components of an integrated 3D-aided profile recognition system.

C. Multiple-frame Identification

A single face profile is a weak biometric, primarily because of pose uncertainty and inaccuracies in the acquisition and extraction stages. If the sequence of frames is available, we can compensate for these uncertainties by fusing the results of recognition from multiple frames.

An alternative approach for multi-frame profile-based recognition is based on a super-resolution approach, used to combine multiple extracted profiles into one with higher resolution, as described by Zhou and Bhanu [20]. However, they impose an assumption of negligible pose changes across the sequence. Our assumption, on the contrary, is that, using video frames acquired at a low frame rate, we will be able to accumulate evidence from more poses.

We consider 5 standard approaches for score fusion: (i) mean rank – rank of every subject is averaged across time, (ii) voting – count rank-1 occurrences for every subject across time, (iii) mean score – the matching score of every subject is averaged across time, (iv) best score – only the best score per subject is preserved, and (v) mean of best – only 10% of best scores are used to compute the average across time.

VI. INTEGRATION OF THE METHODS IN THE AUTOMATIC SYSTEM URxD-PV

Implementation of the automatic profile-based profile recognition system requires incorporation of previously described algorithms in the framework illustrated in Fig. 9. In this section we provide details only about the steps that are not yet covered in the previous sections.

A. Enrollment phase

During the enrollment phase (E) the raw data of each subject are converted to metadata and stored in the database as follows:

- E1. Acquire a facial shape with a 3D scanner and convert it to a polygonal mesh representation.
- E2. Align and fit the 3D data to a common reference model.
- E3. Generate multiple synthetic profiles by sampling a predefined range of rotation angles and locate a set of anatomical landmarks on them (Sections III-B and III-C).
- E4. Derive a set of features based on the profile geometry and landmark locations from profiles and store them as metadata to be used in the identification phase.

Step E1: An essential requirement of our system is the high accuracy in regions along the central line of face symmetry. In fact, the accuracy of the surface in some regions (e.g., near the ears) has no importance because they never appear on silhouetted profiles. Previous research [33] suggests that the recognition algorithm might benefit from 3D raw data with RMS error below 1 mm. Today's 3D sensor technology makes such data available and the accuracy increases as technology improves [8].

We apply a number of preprocessing steps are described in [9] for the purpose of mitigating sensor-specific problems and making all succeeding steps sensor-independent. These steps include hole filling, smoothing, and

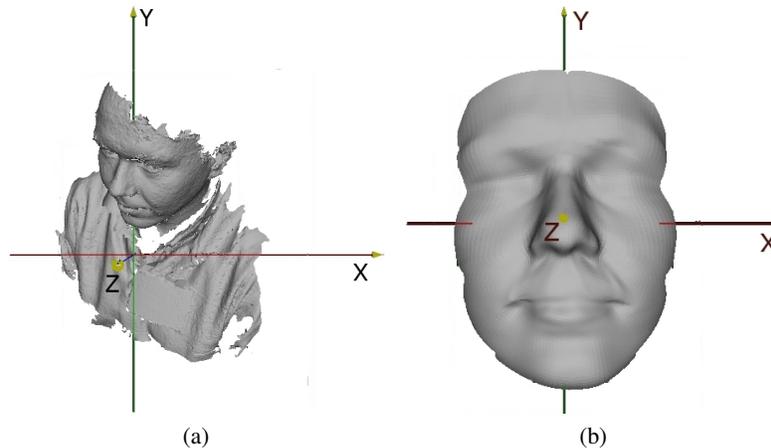


Fig. 10: The process of fitting a generic face model to the raw 3D scan. (a) Raw data and (b) fitted AFM.

resampling.

Step E2: This process aims to fit a generic face model to the raw 3D scan of the subject. This step is necessary to obtain 3D face models suitable to form the gallery database. We use an *Annotated Face Model* (AFM) [9] that defines the control points of a subdivision surface and is annotated into different areas (e.g., mouth, nose, eyes). Specifically, the model is first globally aligned to the subject and then local elastic deformations are applied to fit the model to the data.

The alignment is a multistage process involving (i) establishing initial correspondence by using a spin-images approach [34], (ii) applying a variant of the Iterative Closest Point (ICP) algorithm, and (iii) refining registration results by applying Simulated Annealing to minimize the Z-buffers difference. The AFM is fitted to the mesh data using the elastically adapted deformable model framework of Metaxas and Kakadiaris [35]. The result of this method applied on raw 3D data is illustrated in Fig. 10 (more details may be found in [9]).

Step E4: The metadata, which is being stored in the gallery database, will be used in two different stages of the identification phase: (i) *Data used for matching* reflects subject-specific information and depends on the specific classifier or distance measure used for the identification, which is discussed in Section V. (ii) *Data used for profile extraction* reflects properties common for all the profiles in the current gallery – a statistical shape model. In our case, this data includes a hierarchical variant of a linear deformable model, which is reviewed in Section IV-B.

B. Identification phase

During the identification phase (I), the profile is extracted from a 2D image, and its metadata is matched with gallery metadata as follows:

11. Acquire an image and compute a tight region of interest (*ROI*) that contains the face (Face Detection).
12. Compute a set of features for each pixel in the *ROI*, which will be used to guide the shape extraction procedure (Section IV-A).
13. Extract the profile shape using the modified Active Shape Model (*ASM*) approach (Section IV-C).
14. Extract features (varies depending on classifier) from the profile shape (Section V).
15. Match/Classify the features (Section V).

Step II: At this stage may be employed any existing state-of-art detector specially trained on side-view exemplars, such as the well-known Viola-Jones detector [36] or the profile-specific multi-biometric approach proposed by Gentile *et al.* [37]. In our experiments, the region of interest is extracted by software developed by Pittsburgh Pattern Recognition Inc. [38].

TABLE I: Cohorts used as gallery.

Cohort	No. of datasets	Mesh quality	Collection
G1	19	good	UHDB
G2	50	moderate	UHDB
G3	50	low	UHDB
G4	30	good	UHDB
G5	50	very low	XM2VTS

TABLE II: Cohorts of images used as probe.

Cohort	Datasets per subject	Face size (pixels)	Quality (contrast)	Controlled environment	Pose
P2a,P3a	1	300	moderate	no	standard
P2b,P3b	1	300	moderate	no	non-standard
P5	2	250	good	yes	standard

VII. PERFORMANCE EVALUATION

A. Experiment Data

Collections: In our experiments we employ data from two publicly available collections. The face collection from the University of Houston [39] contains 3D data that was acquired with a 2-pod 3dMD™ system [40]. The quality of data varies through different sessions. Additionally, this collection includes side-view 2D images and video sequences acquired in visual and infrared spectrum. During the acquisition of these images individuals were requested to rotate and tilt their heads arbitrary in the predefined range or to assume standard side-view pose. The acquisition environment includes both controlled (indoor, stable background) and uncontrolled (driver) scenarios.

The XM2VTS face collection [41] from the University of Surrey, includes 3D data and side view images acquired from two different sides of the subject. The 3D data was acquired with a stereo-based 3D camera developed by the Turing Institute. The side view images are all acquired in controlled environment, however some of them have corrupted profiles due to pose or hair occlusions. To the best of our knowledge, these are the only two publicly available collections, including both 3D models and side-view images.

Cohorts: The gallery and probe datasets were partitioned into separate cohorts according to their attributes. Resulting cohorts are enumerated from 1 to 5. Cohorts 1 to 4 are part of a face collection from the University of Houston and cohort 5 is part of the XM2VTS collection. The contents of probe cohorts P1, P4a and P4b are video sequences of 100 frames each, then P4a and P4b contain images from the same scene in visual and infrared spectrum respectively. The contents of the remaining probe cohorts are single side-view images of both standard and arbitrary non-standard poses. The partition of gallery and probe data into cohorts according to the quality and contents is summarized in Tables I, II and III. Typical 3D gallery models and 2D probe images are depicted in Figs. 11 and 18.

B. Parameters

In our experiments, the profile is represented by a 116-point shape ($n = 116$); we also construct a 58-point shape for the coarse search stage. The employed sector-based model is constructed using the scheme in Fig. 5(b). The parametric representation with linear deformable total number retains a minimal number of principal variation modes in order to explain 98% of the variance.

For the gallery profiles sampling, we consider angles in the range $[-110^\circ, -70^\circ]$ for yaw and $[-25^\circ, 25^\circ]$ for roll. The angle offset from the standard side-view is limited in such a way that no landmark disappears from the profile due to full self occlusion. We do not create profiles for different pitch angles because they correspond to the in-plane rotation only and do not influence the geometry of the profile. The resolution of sampling is 5° . In total, we obtain 99 profiles per subject.

For our experiments, we use the libSVM [42] implementation of SVM classifier with linear kernel, which implemented the multi-class classification with one versus one strategy.

TABLE III: Cohorts of videos used as probe. Includes one video per subject, 100 frames per video.

Cohort	face size (pixels)	quality (contrast)	controlled environment	Pose
P1	140 – 270	varying	no	random
P4a	500	good	yes	random
P4b	140	good	yes (infrared)	same as in P4a

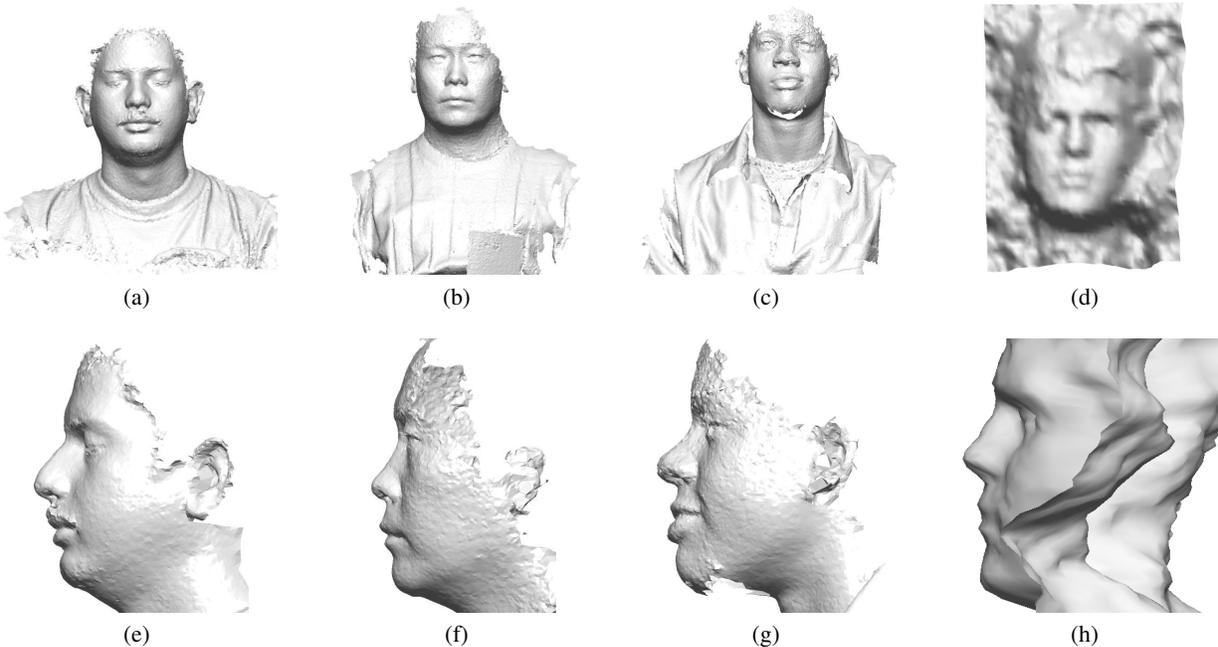


Fig. 11: Samples of raw 3D models from cohorts: (a) G1 (quality of G4 are comparable), (b) G2, (c) G6, and (e-h) side views of the models in the top row.

For the unconstrained scenario, probe (P1-P3) profiles are extracted by both a human operator and the proposed algorithm, and both sets of profiles are employed for further experiments. We employ Cumulative Match Characteristic (CMC) curves for demonstration of recognition accuracy.

C. Experiments

For our experiments we employ different configurations of URxD-PV system that summarized in the Table IV.

Experiment 1: Performance of classification algorithms for uncontrolled environment: In this experiment, we validate recognition performance of the system on the drivers single-frame cohort. It employs the gallery cohort G2 and probe cohort P2. We present overall performance of the considered identification algorithms on the probe images with both standard and non-standard poses in Fig. 12(a). CMC curves for each type of pose are depicted in Figs. 12(a,b).

The performance of Hausdorff distance-based matching (rank-1 recognition rate is 86% for P2 probe cohort) is clearly superior to the classification with SVM (rank-1 recognition rate is 60% for P2 probe cohort). This can be attributed to problematic properties of our training and testing sets: (i) training sets are not independent from each other as soon as they related to the same 3D mesh and specific poses, and (ii) training and testing sets (3D profiles and 2D profiles) are not originating from the same distributions. The first problem causes overfitting and may be solved by acquiring multiple independent 3D scans per person; whereas, solution of the second requires learning robust features that explore similar behavior for 3D and 2D profiles.

TABLE IV: URxD-PV system configurations.

Configuration	Classifier	Fusion	Experiments
URxD-PV1	HD	–	1,3,4,5
URxD-PV2	SVM	–	1
URxD-PV3	HD	mean rank	2
URxD-PV4	HD	voting	2
URxD-PV5	HD	mean	2
URxD-PV6	HD	max	2
URxD-PV7	HD	mean of best	2,4
URxD-PV8	SVM	mean rank	2
URxD-PV9	SVM	voting	2
URxD-PV10	SVM	mean	2
URxD-PV11	SVM	max	2
URxD-PV12	SVM	mean of best	2

We observe that recognition is higher for the nearly standard profiles (rank-1 recognition rate is 98% for HD and 76% for SVM classifier), than for non-standard profiles (rank-1 recognition rate is 67% for HD and 54% for SVM classifier). This fact may be attributed to the fact that standard profiles contain more discriminative information.

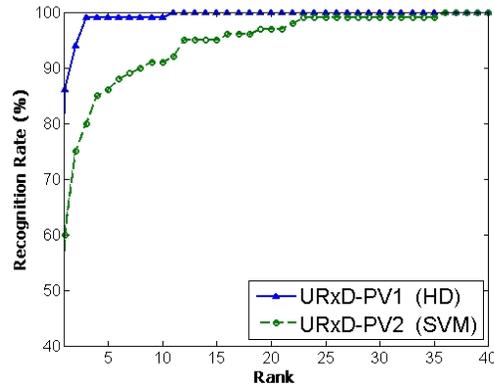
Experiment 2: Analyzing performance on multi-frame input: The purpose of this experiment is assessing the influence of various fusion schemes (see Section V-C) for multi-frame identification. The first and the second columns in Fig. 13 describe the performance of the algorithms on data from a visual spectrum and on infrared sequences in the controlled environment. The third column depicts recognition results on unconstrained scenario videos. The profiles of the first two sequences are extracted with the aid of a simple thresholding approach and, for the third sequence, manually extracted profiles are employed.

Comparison of various fusion schemes shows that the “mean of best” and the “best” are the preferable schemes for the majority of probe sequences. For instance, when we consider HD algorithm, rank-1 recognition rate for “mean of best” is 97% for the sequences of P4a, 89% for the sequences of P4b and 100% for P1. Corresponding recognition rates for the SVM classifier are 93%, 82% and 84%. As in the previous experiment, the recognition based on Hausdorff distance (top row of Fig. 13) is better than SVM-based recognition (bottom row of Fig. 13), however the superiority of “mean of best” fusion scheme is preserved. The remaining fusion rules are too sensitive for the outliers’ scores.

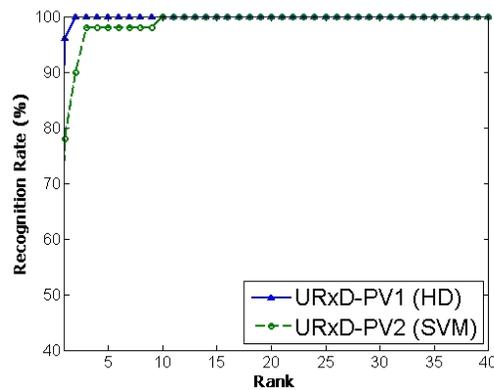
The drop in performance for the infrared sequence (middle column of Fig. 13) is attributed to the fact that it corresponds to smaller face size (in pixels). Note that the number of subjects in the gallery used for recognition of sequences in P1 (right column of the Fig. 13) is smaller than for two other sequences, but the quality and resolution and contrast of some sequences is very low, which particularly influences the performance of the SVM classifier.

Experiment 3: Influence of the quality of the 3D gallery data: In this experiment, we analyze the influence of a 3D quality mesh on the recognition results. For this purpose, we present results of moderate quality based on cohort G2 of moderate quality and compare them to results based on the G3 and G5 gallery cohorts with lower qualities. Only configuration URxD-PV1 (HD matching) is used for this experiment. As one would expect, the quality of the 3D mesh is critical for recognition performance, which is shown in the results presented in Fig. 14. Rank-1 recognition rate is 86% for the cohort of moderate quality and it is only 68% for the cohort G3 of low quality, characterized by missing data and noisy regions. The most drastic drop of the performance (only 46% rank-1 recognition) is observed for the cohort G5, which has low quality even for full 3D face recognition.

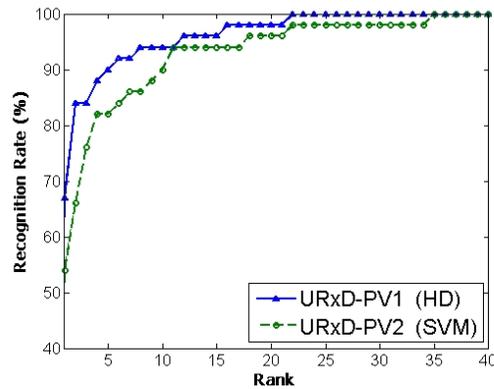
Experiment 4: Quality of extractor and its influence on identification: The accurate extraction of the profile from unconstrained scenes is a complicated task, especially for images with low contrast. Even considering manually extracted profiles, the shape may vary from operator to operator or even for the same operator in two different sessions. To account for this variation, we performed an experiment where two independent operators acquired two sets of profiles that were compared to profiles extracted by the proposed automatic algorithm. The modified Hausdorff distance is used as a distance metric. The cumulative distribution functions of these distances across the three considered sets of extracted profiles are presented in Fig. 15. For instance, examining these distributions, we may observe that 90% of automatically extracted profiles have less than 0.015 distance from manually extracted profiles and at the same time less than 98% of manually extracted profiles have such a distance from each other.



(a)



(b)



(c)

Fig. 12: Recognition results on drivers side-view images using two identification approaches: (a) Performance on both datasets P2a and P2b together, (b) performance on nearly standard side-view images (P2a only), and (c) performance on non-standard side-view images (P2b) only.

Note that the distances are normalized such that the unit distance is corresponding to the distance from [UN] to [NB] landmarks.

The distances between manually extracted profiles is lower than the distance between automatically and manually selected profiles, especially for regions below the chin and above the eyebrows. This is natural because human perception of the profile does not change much from operator to operator and it is less influenced by cast shadows and edges of the background than our algorithm.

In order to demonstrate the influence of the automatic profile extractor on recognition, we use the same system configuration (URxD-PV1) once on profiles extracted from probe images automatically in the unconstrained scenarios and then on profiles extracted manually by a human operator. Figs. 16 (a,b) depict comparisons for single- and multi-frame inputs, the fusion scheme ‘mean-of-best’ is considered for the multi-frame cohort (URxD-PV7).

As one may expect, we observe a decline in performance for both datasets. For instance, rank-1 recognition rate is 86% for single frame cohort (P2) using manually extracted profiles and 78% using automatically extracted profiles. For the multi-frame cohort (P1) we observe a drop in performance from 100% to 84%. The majority of misclassifications are related to sequences of low resolution and poor contrast. The negative influence on SVM classifier is much higher, as it is much more sensitive to inaccuracies introduced by an automatic extractor. The Hausdorff distance metric is known for its robustness to noise and, thus, the drop in performance is significantly smaller.

Experiment 5: Influence of sampling range and density: In order to demonstrate the sensitivity of the algorithm to the predefined range of gallery sampling angles, we compare recognition results based on the original gallery to the results based on wider or narrower ranges, where each range is reduced by 5° from each side. Results of such comparison are depicted in Fig. 17 (a,b) separately for standard and non-standard poses. We also present the CMC curve based on matching of the standard pose only. In a similar manner, Fig. 17 (c,d) depicts the influences of angular sampling density on recognition by comparison of the current sampling density of 5° to the alternative sparser sampling densities of 10° and 20° . These experiments were applied on P3a and P3b cohorts with manually extracted profiles and HD identification algorithm (URxD-PV1).

The results show a clear tendency for the widely sampled pose domain to be more robust on non-standard poses. For instance, rank-1 recognition is 78% for wide region (current settings), 76% for slightly narrower region and only 56% for the sampling region with 10° reduced from each side. On the other hand narrow sampled pose domain regions will slightly outperform if we consider only nearly standard poses. For instance, sampling in the narrow region results in 98% rank-1 recognition as compared to 96% recognition for other settings (wide and moderate). However, even in this case, sampling only one point corresponding to standard pose (ultra-narrow) is worse than other options and results in 92% rank-1 recognition for nearly standard poses and only 56% for non-standard poses. Unlike the area of sampling region, the frequency of sampling has less influence on the performance. Even very sparse sampling of 20° results in a performance very similar to denser frequencies.

VIII. CONCLUSIONS

We presented a fully automatic system for profile-based 3D aided side-view face recognition. It is intended to be used in the absence of frontal view face images. The system uses a 3D scan of the face to model the geometry of the face profile contour under various poses. In this paper, we provided solutions for all stages of the integrated system for accurate landmarking, statistical shape modeling, 3D profile generation, and extraction of the profile shapes from side-view images.

We introduced a modified ASM-based algorithm for profile curve extraction from images that employs simple image features and does not require a set of 2D images in the training stage.

The above experiments explore the basic factors that may influence the performance of automatic 3D-aided profile-based recognition such as quality of 3D data, accuracy of the extracted profile, fusion schemes for multiframe recognition, and configuration of the sampling domain. All experiments were performed using publicly available databases.

REFERENCES

- [1] W. Zhao, R. Chellappa, P. Phillips, and A. Rosenfeld, “Face recognition: A literature survey,” *ACM Computing Surveys*, vol. 35, no. 4, pp. 399–458, 2003.
- [2] X. Tan, S. Chen, Z. Zhou, and F. Zhang, “Face recognition from a single image per person: A survey,” *Pattern Recognition*, vol. 39, no. 9, pp. 1725–1745, 2006.
- [3] N. Davidenko, “Silhouetted face profiles: A new methodology for face perception research,” *Journal of Vision*, vol. 7, pp. 1–17, 2007.
- [4] T. Papatheodorou and D. Rueckert, *Face recognition*. I-Tech Education and Publishing, July 2007, ch. 3D Face Recognition, pp. 417–446.
- [5] C. Li and O. Barreto, *Profile-based 3D face registration and recognition*. Springer-Verlag, 2005, vol. 3506, ch. 10, pp. 478–488.
- [6] L. Yin and M. Yourst, “3D face recognition based on high-resolution 3D face modeling from frontal and profile views,” in *Proc. ACM SIGMM Workshop on Biometrics Methods and Applications*, New York, NY, Nov. 8 2003, pp. 1–8.

- [7] J. Durou, M. Falcone, and M. Sagona, "Numerical methods for shape-from-shading: A new survey with benchmarks," *Computer Vision and Image Understanding*, vol. 109, no. 1, pp. 22–43, 2008.
- [8] K. Bowyer, K. Chang, and P. Flynn, "A survey of approaches and challenges in 3D and multi-modal 3D+2D face recognition," *Computer Vision and Image Understanding*, vol. 101, no. 1, pp. 1–15, Jan. 2006.
- [9] I. Kakadiaris, G. Passalis, G. Toderici, M. Murtuza, Y. Lu, N. Karampatziakis, and T. Theoharis, "Three-dimensional face recognition in the presence of facial expressions: An annotated deformable model approach," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, no. 4, pp. 640–649, Apr. 2007.
- [10] T. Cootes, C. Taylor, D. Cooper, and J. Graham, "Active shape models - their training and application," *Computer Vision and Image Understanding*, vol. 61, no. 1, pp. 38–59, Jan. 1995.
- [11] I. Kakadiaris, H. Abdelmunim, W. Yang, and T. Theoharis, "Profile-based face recognition," in *Proc. 8th IEEE International Conference on Automatic Face and Gesture Recognition*, Amsterdam, The Netherlands, Sep. 17-19 2008, pp. 1–8.
- [12] B. Efraty, D. Chu, E. Ismailov, S. Shah, and I. Kakadiaris, "3D-aided profile-based face recognition," in *Proc. IEEE International Conference on Image Processing*, Cairo, Egypt, Nov. 7-11 2009, pp. 4125–4128.
- [13] B. Efraty, E. Ismailov, S. Shah, and I. Kakadiaris, "Towards 3D-aided profile-based face recognition," in *Proc. 3rd IEEE International Conference on Biometrics: Theory, Applications and Systems*, Arlington, VA, Sep. 28-30 2009.
- [14] F. Galton, "Numerical profiles for classification and recognition," *Nature*, vol. 83, pp. 127–130, March 1910.
- [15] L. D. Harmon, M. K. Khan, R. Lasch, and P. Ramig, "Machine identification of human faces," *Pattern Recognition*, vol. 2, no. 13, pp. 97–110, 1981.
- [16] C. Wu and J. Huang, "Human face profile recognition by computer," *Pattern Recognition*, vol. 23, pp. 255–259, 1990.
- [17] Z. Liposcak and S. Loncaric, "A scale-space approach to face recognition from profiles," in *Proc. 8th International Conference on Computer Analysis of Images and Patterns*, London, UK, Sep. 1999, pp. 243–250.
- [18] G. Kaufman and K. Breeding, "The automatic recognition of human faces from profile silhouettes," *IEEE Transactions on Systems, Man and Cybernetics*, vol. 6, pp. 113–121, Feb. 1976.
- [19] G. Pan, L. Zheng, and Z. Wu, "Robust metric and alignment for profile-based face recognition: An experimental comparison," in *Proc. 7th IEEE Workshop on Applications of Computer Vision*, vol. 1, Jan. 2005, pp. 117–122.
- [20] X. Zhou and B. Bhanu, "Human recognition based on face profiles in video," in *Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Washington, DC, Jun. 2005.
- [21] Y. Gao, "Efficiently comparing face images using a modified Hausdorff distance," in *Proc. IEEE Conference on Vision, Image and Signal Processing*, Dec. 2003, pp. 346–350.
- [22] Y. Gao and M. Leung, "Human face profile recognition using Attributed String," *Pattern Recognition*, vol. 35(2), pp. 353–360, 2002.
- [23] —, "Line segment Hausdorff distance on face matching," *Pattern Recognition*, vol. 35, no. 2, pp. 361–371, 2002.
- [24] A. Bottino and S. Cumani, "A fast and robust method for the identification of face landmarks in profile images," *WSEAS Transactions on Computers*, vol. 7, pp. 1250–1259, 2008.
- [25] L. Gu and T. Kanade, "3D alignment of face in a single image," in *Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, New York, NY, Jun. 17-22 2006, pp. 1305–1312.
- [26] M. Mahoor and M. Abdel-Mottaleb, "Facial features extraction in color images using enhanced active shape model," in *Proc. 7th International Conference on Automatic Face and Gesture Recognition*, Washington, DC, USA, Apr. 2-6 2006, pp. 144–148.
- [27] M. de Berg, M. van Krefeld, M. Overmars, and O. Schwarzkopf, *Computational geometry: Algorithms and applications*, 2nd ed. Springer, 2000.
- [28] T. Cour, F. Benezit, and J. Shi, "Spectral segmentation with multiscale graph decomposition," in *Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, San Diego, CA, Jun. 20-25 2005, pp. 1124–1131.
- [29] D. Comaniciu and P. Meer, "Mean shift: A robust approach toward feature space analysis," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 5, pp. 603–619, 2002.
- [30] B. Efraty, E. Ismailov, S. Shah, and I. Kakadiaris, "Profile-based 3D aided face recognition," University of Houston, Tech. Rep., 2010.
- [31] T. Cootes, C. Taylor, H. Kang, and V. Petrovic, "Modeling facial shape and appearance," in *Handbook of Face Recognition*. Springer, 2005, pp. 39 – 63.
- [32] T. Cootes and C. Taylor, "Statistical models of appearance for computer vision," University of Manchester, Tech. Rep., Oct 2001.
- [33] K. Chang, K. Bowyer, and P. Flynn, "Face recognition using 2D and 3D facial data," in *Proc. Workshop on Multimodal User Authentication*, Santa Barbara, CA, 2003, pp. 25–32.
- [34] A. Johnson and M. Hebert, "Using spin images for efficient object recognition in cluttered 3D scenes," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 21, no. 5, pp. 433–449, 1999.
- [35] D. Metaxas and I. Kakadiaris, "Elastically adaptive deformable models," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 10, pp. 1310–1321, 2002.
- [36] P. Viola and M. Jones, "Robust real-time face detection," *International Journal of Computer Vision*, vol. 57, no. 2, pp. 137–154, 2004.
- [37] J. Gentile, K. Bowyer, and P. Flynn, "Profile face detection: A subset multi-biometric approach," in *Proc. 2nd International Conference on Biometrics Theory, Applications and Systems*, Arlington, VA, Sep. 29 - Oct. 1 2008.
- [38] Pittsburgh Pattern Recognition, "PittPatt face tracking & recognition software development kit," 2009. [Online]. Available: http://www.pittpatt.com/products/ft_r_sdk/
- [39] URxD-PV, "UHDB22: CBL database for biometrics research," available at <http://cbl.uh.edu/URxD/datasets/>.
- [40] "3dMD," 2009. [Online]. Available: <http://www.3dmd.com/>
- [41] K. Messer, J. Matas, J. Kittler, J. Luetin, and G. Maitre, "XM2VTSDB: The extended M2VTS database," in *Proc. Audio- and Video-Based Biometric Person Authentication*, Washington, DC, Mar. 22-23 1999, pp. 72–77.
- [42] C.-C. Chang and C.-J. Lin, "LIBSVM: A library for support vector machines," 2001, software available at <http://www.csie.ntu.edu.tw/>.

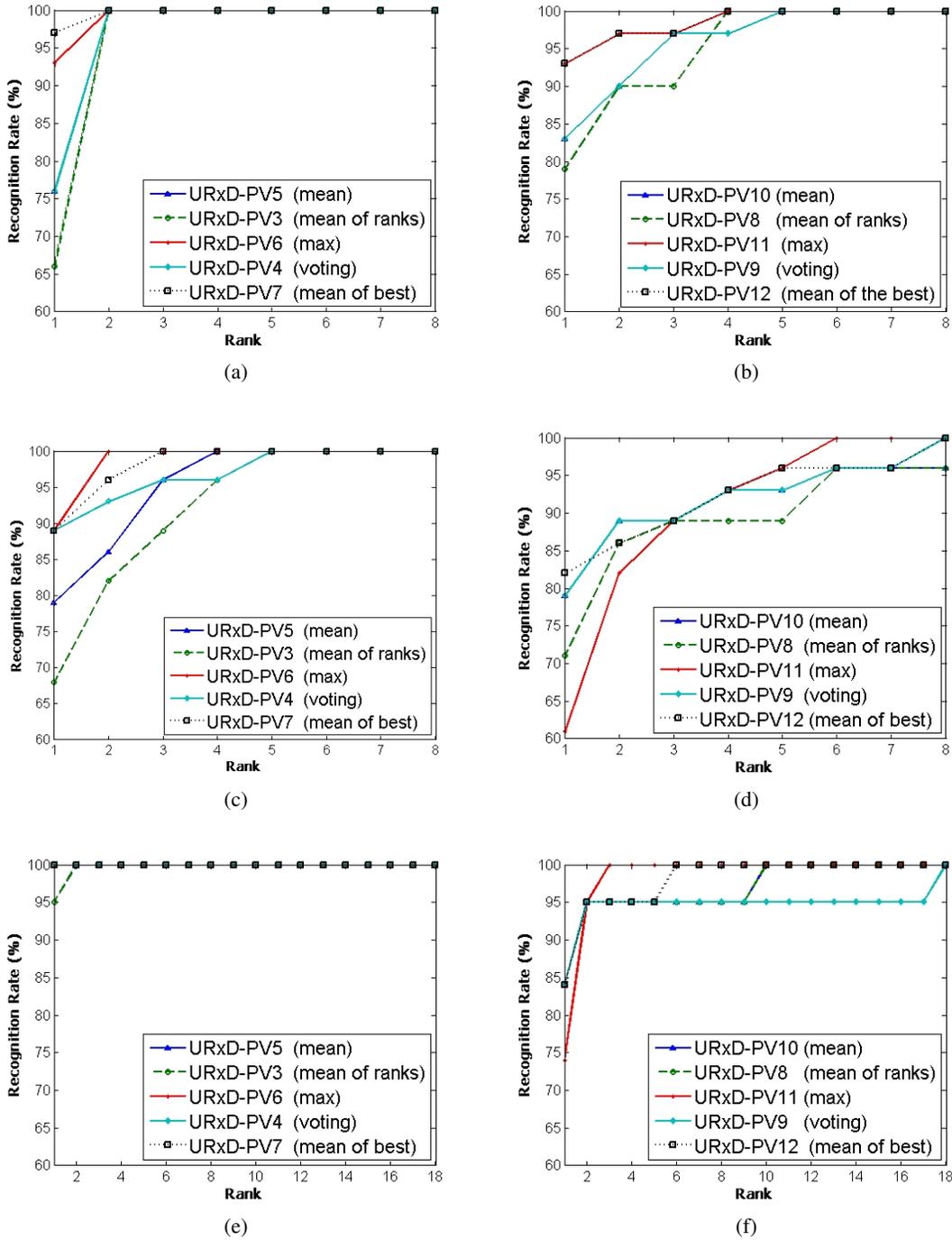


Fig. 13: Recognition results on three probe cohorts (rows): (a,b) P4a, (c,d) P4b, and (e,f) P1. The identification is performed using two algorithms (columns): (a,c,e) HD, and (b,d,f) SVM.

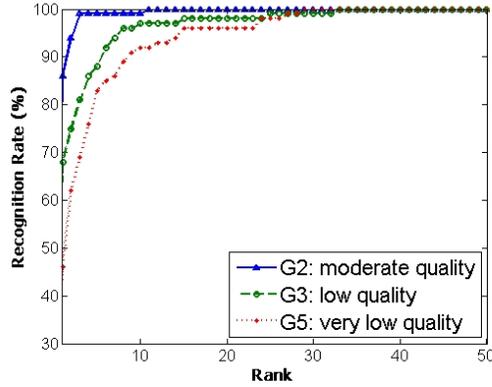


Fig. 14: Comparison of the identification performance based on 3D models with different qualities, with the same identification algorithm (HD) applied.

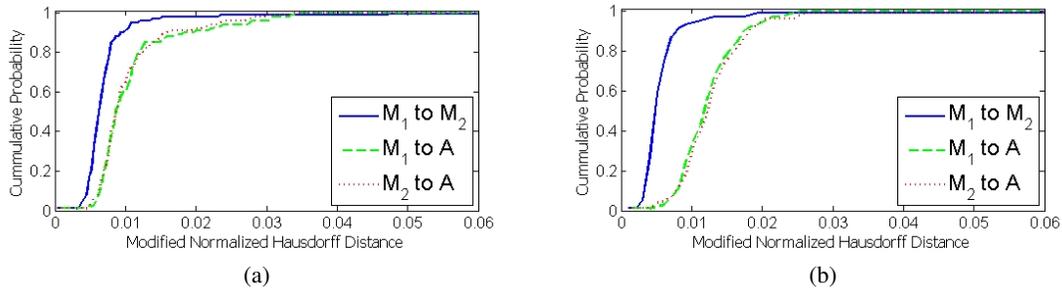


Fig. 15: Comparison of profiles extracted by two different human operators (M_1 , M_2) and our algorithm (A). (a) Full profile $[TO, BO]$, and (b) trimmed central section $[UN, NB]$.

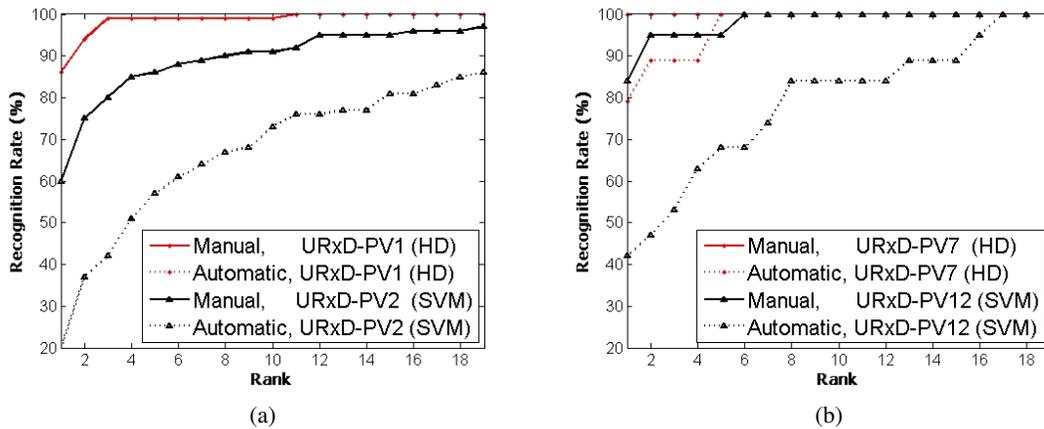


Fig. 16: Influence of profile extractor quality on recognition results in unconstrained scenarios: (a) single frame cohort (P2), and (b) multi-frame cohort (P1).

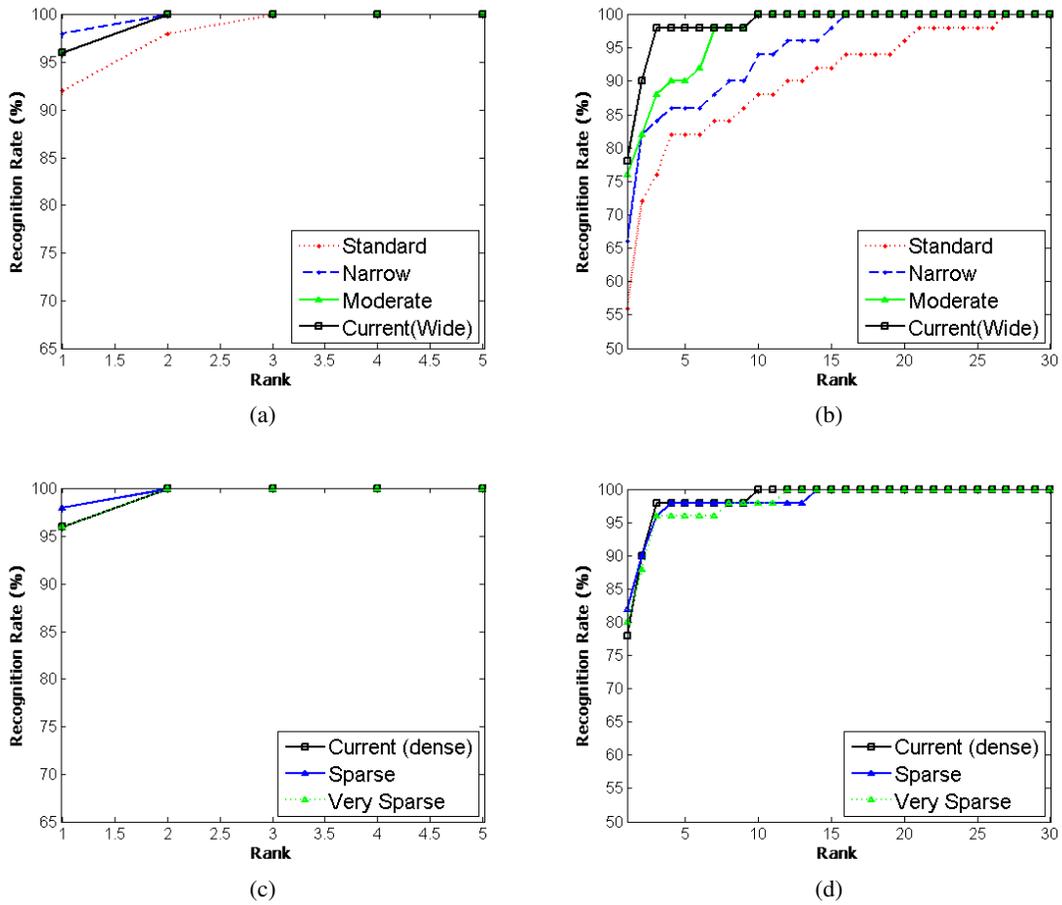


Fig. 17: Recognition results using various sampling domains: (a,c) probe cohort P2a, and (b,d) probe cohort P2b.

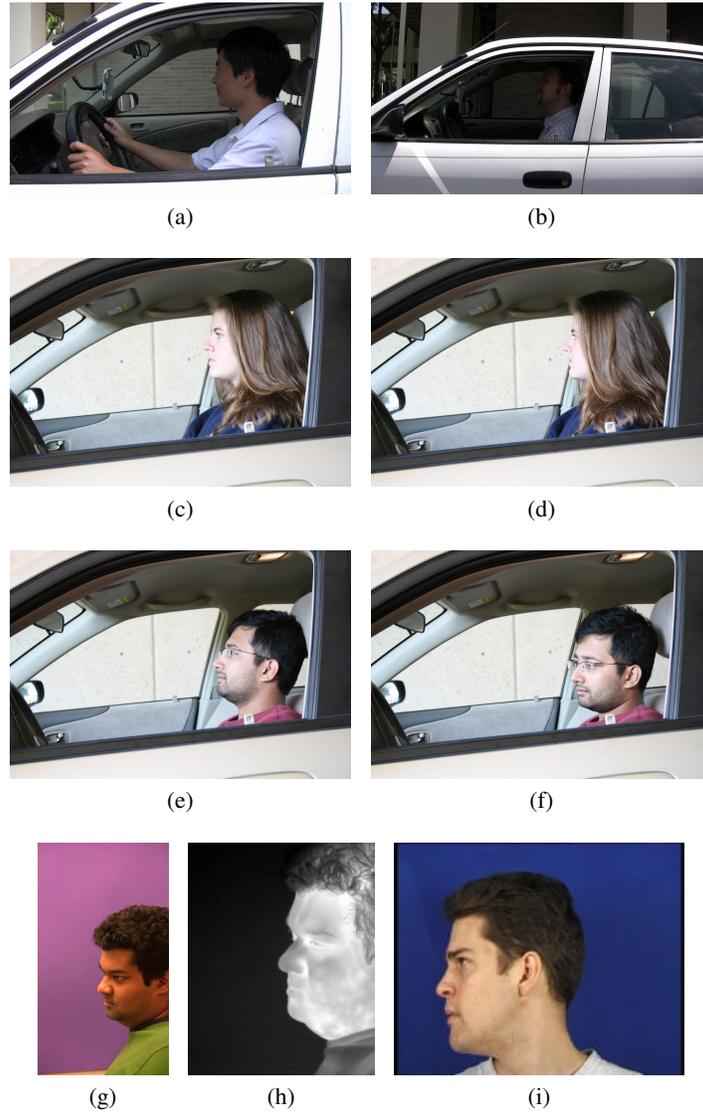


Fig. 18: Samples of images from probe cohorts: (a) sample frame from sequence in P1 (high resolution), (b) another sample frame from sequence in P1 (low resolution), (c) sample from P2a, (d) P2b, (e) P3a, (f) P3b, (g) sample frame from sequence in P4a, (h) P4b, and (i) P5.