Profile-based 3D-aided face recognition

Boris Efraty, Emil Bilgazyev, Shishir Shah, Ioannis A. Kakadiaris *
Computational Biomedicine Lab, Department of Computer Science, University of Houston, Houston, TX 77204, USA

A R T I C L E   I N F O
Article history:
Received 16 September 2010
Received in revised form 25 March 2011
Accepted 7 July 2011
Available online 19 July 2011

Keywords:
Profile-based face recognition
Facial data
3D
Multi-frame recognition
Multi-modal vision

A B S T R A C T
This paper presents a framework for automatic face recognition based on a silhouetted face profile (URxD-PV). Previous research has demonstrated the high discriminative potential of this biometric. 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 profile is extracted from the side-view image and its metadata is matched with the gallery metadata. We validate the accuracy of URxD-PV using data from publicly available databases.

1. 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 a complementary shape structure that is not visible in the frontal view, but can often be obtained 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 faces are available. Two different types of profiles are employed in our approach: (i) standard profiles – those generated synthetically from the 3D face model, and (ii) 2D profiles – those extracted from 2D images of side-view faces.

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 [4–6], 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 the camera and the object is not sufficiently large, then perspective projection has to be considered (based on imaging sensor parameters). In this paper, the term “profile” refers to the silhouette of nearly side-view head images.

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 used with 2D images to overcome challenges caused by variations in lighting, pose, and facial expressions [8,9]. An accurate 3D model provides shape information (which is not necessarily on the symmetry axis) that is highly valuable for modeling the profile shape. Therefore, by employing the tools from 3D face analysis the performance of a profile-based recognition framework can be improved.

The central concept of our approach is the use of 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. Two different types of profiles are employed in our system: (i) 3D profiles – those generated synthetically from the 3D face models, and (ii) 2D profiles – those extracted from 2D images of side-view faces.

* Corresponding author.
E-mail address: ioannisk@uh.edu (I.A. Kakadiaris).

© 2011 Elsevier Ltd. All rights reserved.
The schematic illustration of the profile-based face recognition system (URxD-PV) is depicted in Fig. 2. We provide algorithmic solutions for the entire 3D-aided profile-based recognition framework including profile modeling, landmark detection, shape extraction, accurate 3D profile generation and classification.

During the enrollment mode, the precise 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. We compare two different approaches for profile identification. The first method is based on matching 3D and 2D profiles represented explicitly using modified Hausdorff distance and the second method is based on an SVM classifier applied on the rotation-, translation-, and scale-invariant features extracted from the profiles.

The main contribution of our work is the incorporation of 3D facial shape information into a profile-based face recognition framework, resulting in a robust method for the extraction of fiducial points and a robust classifier for a range of poses. This is in contrast to the existing methods (see Section 2), which are designed to be efficient only on standard profiles and use 2D images as gallery. The proposed system is fully automatic and hence, can be used in applications related to access control. Parts of this work have appeared in [10–12]. In this paper we provide a more detailed description of the technique with the inclusion of additional experiments and demonstrate the system’s performance on several new datasets, which include infrared images and video sequences.

As part of the identification mode, we developed an algorithm for 2D profile extraction from a side-view image. However, the details of this module have not been presented, as shape extraction is not the main focus of this paper. The details about this algorithm and its experimental validation in natural scenes may be found in [12,13].

The rest of this paper is organized as follows: Section 2 reviews previous work, Section 3 describes methods used for profile representation and modeling using 3D data, while some identification approaches are discussed in Section 4. In Section 5, we present the components of a fully automatic face recognition system, which is based on the proposed methodology. Section 6 presents a performance evaluation using publicly available databases. Section 7 provides a summary of our main contributions.

2. 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 with the exploration of 3D-aided face recognition. The methods for recognition using the profile curve can be classified into one of two categories: landmark-based methods [15–18] or global methods [10,19,20]. Landmark-based methods rely on the attributes associated with a set of fiducial points and the similarity metrics based on those attributes. Global methods consider each profile as a geometric object and introduce a similarity metric between objects.

Harmon et al. [15] defined 17 fiducial points. After aligning two profiles based on selected landmarks, 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.

Bhanu and Zhou [19] proposed a curvature-based matching approach using a dynamic warping algorithm. They reported a recognition rate of almost 90% on the University of Bern Database (UBD) of 30 subjects. Gao and Leung [21] 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 UBD database. Gao et al. [20,22] proposed new formulations of the Hausdorff distance. Initially, it was extended to match two sets of lines, and later, was computed by weighting points based on their significance.

As mentioned earlier, all these methods can only be applied to standard profiles and use 2D images as gallery. Kakadiaris et al. [10] introduced the use of a 3D face model for the generation of profiles under different poses for the gallery. A modified directional Hausdorff distance between the probe profile and the
gallery profile was used for identification. In addition, four different profiles under various rotation angles were used to introduce robustness to pose.

3. Profile representation and modeling

In this section, we first provide a description of the proposed profile shape representation and then describe the procedure for profile sampling in the pose space using a 3D model.

3.1. Profile representation

In our approach, we treat the profile as an open curve. The profile curve \( c \) may be described by a pair of arc-length parameterized 1D functions \( Y_\ell(t) \) and \( X_\ell(t) \), where \( 1 \leq \ell \leq [0,1] \). A set of \( k \) landmarks is defined by the landmark’s coordinates on a parametric curve: \( \{0 = v^1 < \cdots < v^k = 1\} \). The set contains both, anatomical (e.g., “chin”) and pseudo landmarks (e.g., “middle of the nose”). The complete set of landmarks is illustrated in Fig. 3(a). We approximate functions \( Y_\ell(t) \) and \( X_\ell(t) \) using a finite set of points and obtain an equivalent \( n \)-points shape model

\[
\mathbf{v} = [x_1, y_1, x_2, y_2, \ldots, x_0, y_0]^T \in \mathbb{R}^{2n},
\]

The positions of the points are obtained through uniform sampling of functions \( Y_\ell(t) \) and \( X_\ell(t) \) between a predefined subset of the landmarks.

3.2. Generating 3D profiles

Using a 3D model that does not contain any artifacts and has been aligned to a common reference model [9], profiles corresponding to a range of rotation angles are automatically generated to model the variations in the geometry of profiles caused by head rotations. 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 and has linear complexity. Algorithm 1 describes the proposed method and uses the following notation:

- \( [\mathbf{n}_i] \) is the set of normal vectors corresponding to each face of the triangular polygon mesh,
- \( e(i,j) \) is the edge shared by \( i \)-th and \( j \)-th faces and can also be defined by its vertices \( e \equiv (v_i, v_j) \) (each vertex is associated with coordinates in the 2D projection plane), and
- \( \mathbf{o} \) is the vector that defines the position of a 3D object with respect to the observer.

Generating a 3D profile has two main stages: (i) build a graph based on critical subset of edges, which may be on the profile (Steps 1–5); and (ii) find the path (a chain of line segments) that bounds the rest of the graph from the left side (Steps 6–15). We use spline interpolation to introduce smoothness and perform uniform resampling of the spline curve by arc-length.

Algorithm 1. Generating a 3D profile.

Input: \([\mathbf{n}_i], [v_i], [e_i] \) – sets of normals, vertices and edges, respectively, \( \mathbf{o} \) – observer vector

Output: Profile \( c \)

1: Set \( c = 0 \). 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 the line segments defined by \( \mathcal{E} \) in the projection plane using the plane sweep algorithm [23].
3: for all \( (v_1, v_2, v_3, v_4) \) with intersection \( p \)
4: Split \( (v_1, v_2, v_3, v_4) \) into four edges \((v_1, p_1, v_2, v_3), (v_3, p_2, v_4, v_2)\). Denote the modified set of edges by \( \tilde{\mathcal{E}} \) and the 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 the lowest and highest \( Y \) coordinate, respectively, in \( \tilde{\mathcal{V}} \).
10: Find edge \((u, v) \in \mathcal{L}_v \) inducing the smallest angle with vector in the direction \([-1,0]\).
11: repeat
12: \( c \leftarrow c \cup (u, v) \),
13: Let \((v, w)\) to succeed \((u, v)\) in \( \mathcal{L}_v \),
14: \( u \leftarrow v, v \leftarrow w \).
15: until \( v \neq a \)

3.3. Landmark detection

The main principles of landmark extraction are well established for standard profiles [17,24]. These methods are usually driven by the curvature extrema and are subject to anthropometric standards of the face. For example, for an upward facing profile, the tip-of-nose is the extremum point of \( Y_\ell(t) \) with \( \ell \in [0,4,0.7] \). However, the accuracy of such methods is very low when applied to general profiles. In this paper, we use both the landmark information from standard profiles and a 3D face surface to accomplish the localization of landmarks. Specifically, we propose to use the 3D surface to extrapolate information about landmarks on a standard profile to a profile obtained with known rotation angles. The method is described in Algorithm 2 and the individual steps are illustrated in Fig. 3.

Fig. 3. Profile landmarks’ localization. Depiction of (a) profile landmarks and their names, (b) landmarks on the standard profile, (c) landmarks projected on a 3D model (red markers) and extensions of landmarks to 3D regions (black markers), (d) projection of landmarks on a rotated (not standard) profile, and (e) result of the refinement procedure. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article).
Algorithm 2. Landmark detection on a rotated profile.

Input: $\mathcal{G}$ – 3D surface, $\mathcal{C}_0$, $\mathcal{C}$ – standard and rotated profiles, respectively, $\{v^1, \ldots, v^n\}$ – landmarks on $\mathcal{C}_0$, rotation and projection parameters.

Output: $\{v^1, \ldots, v^n\}$ – landmarks on $\mathcal{C}$
1: Back-project landmarks from $\mathcal{C}_0$ onto a center line of $\mathcal{G}$ (Fig. 3(b)).
2: for all landmarks $i$ with coordinate $(x_i, y_i, z_i) \in \mathcal{G}$ do
3: Define a subset of points in $e_i$-neighborhood of $(x_i, y_i, z_i)$: $N_i = \{(x, y, z) \in \mathcal{G} \mid \|(x, y, z) - (x_i, y_i, z_i)\| \leq e_i\}$ (Fig. 3(c)).
4: Project $N_i$ onto the profile plane using rotation and projection parameters, thus obtaining a set of 2D points $\hat{N}_i = \{(\hat{x}, \hat{y})\}$ located on one side of $\mathcal{C}$.
5: Find coordinate $v^i$ on profile $\mathcal{C}$ with minimal distance to $\hat{N}_i$ (Fig. 3(d)).
6: end for
7: Refine all landmark coordinates based on neighbor landmarks depicted in Fig. 4(b) and Eq. (2) (Fig. 3(e)).

The last step in Algorithm 2 is a refinement step and is based on the elaboration of a general definition of the landmark. The position of the landmarks (detected by an automatic procedure or by a human operator) has a certain amount of uncertainty. For example, the position of a landmark that is defined by the curvature value is equivocal in the region with a 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 the second derivative), whereas point $E$ corresponds to the zero of the first derivative and is sensitive to rotations. In order to reduce such an ambiguity, we define a refinement operator. This operator detects the most protruding (or intruding) point in the region.

Let $ij$ be the indices of any two landmarks on a profile’s curve and $v^i, v^j$ denote their corresponding arc-length coordinates. For any coordinate $v$, the function $H(v^i, v^j, v)$ defines the signed distance between the vector connecting the points of $v^i, v^j$ and the point corresponding to $v$. The sign of $H(v^i, v^j, v)$ denotes whether the point lies on the left or the right side of the vector. The refined coordinate $v^* = \arg\max_{\{v^i \mid i < r < l \leq n\}} (\gamma \cdot H(l, 1, v))$, where the parameter $\gamma$ is defined to be $1$ (or $-1$) according to the convexity (or concavity) of the profile curve in the neighborhood of the landmark. Fig. 4(a) illustrates the effect of this operator, which detects point $F$. We apply this refinement operator to all the 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 (Fig. 4(b)). This simple and fast operator may be efficiently applied for landmark refinement on any profile, even after a rough initialization (Fig. 4(c)).

4. Identification

To, or match an extracted profile to a subject in the gallery, we consider two general approaches: matching or classification. Both these approaches, and the methods employed for multiple-frame identification, are discussed in the following sections.

4.1. Hausdorff distance-based matching

In this approach, the matching score between the probe profile and every profile in the gallery is computed. The decision is made based on 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. We propose to employ a modified Hausdorff distance as the distance metric for the score computation.

For two finite point sets $M = \{m_1, \ldots, m_n\}$ and $T = \{t_1, \ldots, t_n\}$ with associated weights $\{w^M_1, \ldots, w^M_n\}$ and $\{w^T_1, \ldots, w^T_n\}$, the distance is defined as

$$\frac{1}{n} \max_{m_i \in M} \min_{t_j \in T} \frac{\sqrt{w^M_i \sqrt{w^M_i}}}{C_0} \left( \sum_{m_i \in M} \min_{t_j \in T} \frac{\sqrt{w^M_i \sqrt{w^M_i}}} {C_0} \right),$$

where $M$ and $T$ are the probe and gallery $n$-point shapes, respectively, and $h_M$ and $h_T$ are the normalization factors of the distance between the $UN$ and $NB$ landmarks, respectively (see Fig. 3(a)), which are used to eliminate any effects caused by scale. The set of weights for a probe profile reflects the accuracy of a shape extractor (its value equals $1$ for manually extracted profiles). The set of weights for a gallery profile reflects prior knowledge about the discriminative properties of the various regions and is found experimentally. Each gallery shape is preregistered to the probe shape using Procrustes analysis [25].

4.2. SVM-based classification

A typical classification approach employs 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 distribution.

We define five types of features based on the properties of the profile $p$ between two landmarks $i$ and $j$: (i) Euclidean distance between the landmarks, (ii) arc-length distance between the landmarks, (iii) mean curvature of the region between the landmarks, (iv) $L_2$-norm of the curvature along the contour between the landmarks (proportional to bending energy), and (v) $L_2$-norm of angular chain codes for the region between the landmarks under the assumption of uniform discretization. All of these features are translation- and rotation-invariant. Scale invariance is obtained by normalizing the distances with respect to the size of the profile. This approach, unlike the Hausdorff distance-based matching approach, does not require registration to a common reference shape. Also, it is impossible to use a direct distance metric in the feature space and hence, we need to employ the machine learning approach.

Fig. 4. Depiction of landmark refinement concepts: (a) geometric entities related to Eq. (2); (b) neighbor relations used for the refinement of landmarks; and (c) example of refinement output (discs) compared to input before refinement (circles).
The Support Vector Machine (SVM) classifier, as implemented in [26], is employed as the machine learning technique for classification. We use forward selection with 2-fold cross validation for attribute selection, in which variables are added to the model one at a time [27].

4.3. 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 alternate approach for multi-frame profile-based recognition is based on a super-resolution approach, where multiple extracted profiles are combined into a single profile with higher resolution, as described by Zhou and Bhanu [19]. However, their approach is based on an assumption that the pose changes across the sequence are negligible. On the contrary, our approach is based on the assumption that by using video frames acquired at a low frame rate, we will be able to accumulate evidence from more poses.

We examine five standard approaches for score fusion: (i) mean rank – rank of every subject is averaged across time, (ii) voting – rank-1 occurrences for every subject across time are counted, (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 the best scores are used to compute the average across time.

5. Integration of the methods in the automatic system URxD-PV

Implementation of the automatic profile-based profile recognition system requires incorporation of the previously described algorithms in the framework illustrated in Fig. 5. In this section, we present the details of only those steps that have so far not been discussed in the earlier sections.

5.1. Enrollment mode

During the enrollment mode (E) the raw data acquired from each subject are converted to metadata and stored in the database. The steps involved in this conversion are:

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 generated profiles (Sections 3.2 and 3.3).
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 high accuracy in regions along the central line of face symmetry. In fact, the accuracy of the surface in certain regions (e.g., near the ears) bears no importance because these regions never appear on silhouetted profiles. We apply a number of preprocessing steps such as hole filling, smoothing, and resampling [9], for the purpose of mitigating sensor-specific problems.

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 that are suitable to constitute the gallery database. We use the 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 outcome of this method when applied on raw 3D data is illustrated in Fig. 6 (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) metadata used for matching include subject-specific information and depend on the specific classifier or distance measure, which is discussed in Section 4; (ii) metadata used for profile extraction reflect properties common for all the profiles in the current gallery – a statistical shape model. In this paper, these data include a hierarchical variant of a linear deformable model [13].
5.2. Identification mode

During the identification mode (I), the profile is extracted from a 2D image and its metadata are matched with gallery metadata. The steps involved are:

11. Acquire an image and compute a 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.
13. Extract the profile shape using a modified Active Shape Model (ASM) method.
14. Extract features (varies depending on classifier) from the profile shape (Section 4).
15. Match/classify the features (Section 4).

Step 11: Any of the existing state-of-art detectors that are specially trained on side-view exemplars, such as the well-known Viola-Jones detector [28] or the profile-specific multi-biometric approach proposed by Gentile et al. [29] can be used for face detection. In our experiments, the ROI is extracted by the software developed by Pittsburgh Pattern Recognition Inc. [30].

Steps 12–13: To extract the profile shape from side-view images, we employ an algorithm proposed by Efraty et al. in [12,13]. This algorithm is based on a ASM concept and as it uses a 3D model, it does not require a set of 2D images in the training stage.

6. Performance evaluation

6.1. Collections

In our experiments, we employ data from two publicly available collections. The face collection from the University of Houston [31] contains 3D data that were acquired with a 2-pod 3DM© system [32]. The quality of data varies with different sessions. Additionally, this collection includes side-view 2D images and video sequences acquired in the visible and infrared spectrum. During the acquisition of these images individuals were requested to either rotate and tilt their heads arbitrarily in the predefined range or assume standard side-view poses. The acquisition environment included both controlled (indoor, stable background) and uncontrolled (driver) scenarios.

The XM2VTS face collection [33] from the University of Surrey includes 3D data and side-view images acquired from two different sides of the subject. The side-view images were acquired in a controlled environment. However, some of these images have corrupted profiles due to pose or hair occlusions. To the best of our knowledge, these are the only two publicly available collections that include both 3D models and side-view images.

6.2. Cohorts

The gallery and probe datasets were partitioned into separate cohorts based on their attributes. The 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. Cohorts P4a and P4b contain images from the same scene in the visible 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 the gallery and probe data into cohorts is summarized in Tables 1, 2 and 3. Typical 3D gallery models and probe images are depicted in Figs. 7 and 8.

<table>
<thead>
<tr>
<th>Cohort ID</th>
<th>Number of datasets</th>
<th>Mesh quality</th>
<th>Collection</th>
</tr>
</thead>
<tbody>
<tr>
<td>G1</td>
<td>19</td>
<td>Good</td>
<td>UHDB</td>
</tr>
<tr>
<td>G2</td>
<td>50</td>
<td>Moderate</td>
<td>UHDB</td>
</tr>
<tr>
<td>G3</td>
<td>50</td>
<td>Low</td>
<td>UHDB</td>
</tr>
<tr>
<td>G4</td>
<td>30</td>
<td>Good</td>
<td>UHDB</td>
</tr>
<tr>
<td>G5</td>
<td>50</td>
<td>Very low</td>
<td>XM2VTS</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Cohort ID</th>
<th>Datasets per subject</th>
<th>Face size (pixels)</th>
<th>Quality (contrast)</th>
<th>Controlled environment</th>
<th>Pose</th>
</tr>
</thead>
<tbody>
<tr>
<td>P2a, P3a</td>
<td>1</td>
<td>300</td>
<td>Moderate</td>
<td>No</td>
<td>Standard</td>
</tr>
<tr>
<td>P2b, P3b</td>
<td>1</td>
<td>300</td>
<td>Moderate</td>
<td>No</td>
<td>Non-standard</td>
</tr>
<tr>
<td>P5</td>
<td>2</td>
<td>250</td>
<td>Good</td>
<td>Yes</td>
<td>Standard</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Cohort ID</th>
<th>Face size (pixels)</th>
<th>Quality (contrast)</th>
<th>Controlled environment</th>
<th>Pose</th>
</tr>
</thead>
<tbody>
<tr>
<td>P1</td>
<td>140–270</td>
<td>Varying</td>
<td>No</td>
<td>Random</td>
</tr>
<tr>
<td>P4a</td>
<td>500</td>
<td>Good</td>
<td>Yes</td>
<td>Random</td>
</tr>
<tr>
<td>P4b</td>
<td>140</td>
<td>Good</td>
<td>Yes (infrared)</td>
<td>Same as in P4a</td>
</tr>
</tbody>
</table>

6.3. Parameters

In our experiments, the profile is represented by a 116-point shape. 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 only correspond to the in-plane rotation and do not influence the geometry of the profile. The sampling resolution is $5^\circ$ and therefore, the total number of profiles in the training set is 99 per subject. The gallery data are exactly the same for both identification methods.

For our experiments, we use an SVM classifier with linear kernel, combined with one-versus-one strategy for the multi-class classification [34]. The classes are well-balanced since we sample the same number of 3D profiles per subject. The number of classes for each experiment is equal to the number of subjects in the gallery. We use Cumulative Match Characteristic (CMC) curves to demonstrate recognition accuracy.

6.4. Experiments

For our experiments we use different configurations of the URxD-PV system that are summarized in Table 4. Note that previous methods discussed in Section 2 cannot be applied directly to our data as they neither assume a 3D gallery nor do they take into account the pose variations in the probe.

Experiment 1: Performance of the classification algorithms in an uncontrolled environment

In this experiment, we validate recognition performance of the system on the driver’s single-frame cohort. It employs the gallery cohort G2 and probe cohort P2. The overall performance of the
chosen identification algorithms is evaluated on the probe images with both standard and non-standard poses (Fig. 9(a)). CMC curves for each type of pose are depicted in Figs. 9(a,b).

The performance of Hausdorff distance (HD)-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 the challenging aspects of our training and testing sets: (i) training sets are not independent from each other when related to the same 3D mesh and to a specific pose, and (ii) training and testing sets (3D profiles and 2D profiles) do not originate from the same distribution. The first problem causes overfitting and may be solved by acquiring multiple independent 3D scans per person; whereas, the solution to the second requires the learning of robust features that explore similar behavior for 3D and 2D profiles.

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

Experiment 2: Analyzing performance of a multi-frame input

The purpose of this experiment is to assess the influence of various fusion schemes (described in Section 4.3) for multi-frame identification. The first and the second rows in Fig. 10 describe the performance of the algorithms on data from a visual spectrum and on infrared sequences, respectively, in the controlled environment. The third row 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 used.

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 the Hausdorff distance-based matching, the 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. The corresponding recognition rates for the SVM classifier are 93%, 82%, and 84%, respectively. As in the previous experiment, of the HD-based approach (first column of Fig. 10) the recognition rate of the SVM-based approach (second column of Fig. 10), the “mean of best” fusion scheme achieves superior results. Our experiments suggest that
the remaining fusion rules are too sensitive to the presence of the outliers’ scores.

The drop in the performance for the infrared sequence (second row of Fig. 10) is attributed to the fact that it corresponds to a smaller face size (in pixels). Note that the number of subjects in the gallery used for the recognition of sequences in P1 (third row of the Fig. 10) is smaller than for the other two sequences. Also the quality, resolution, and contrast of some sequences are very low, and these factors also affect 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 the quality of the 3D mesh on the recognition results. For this purpose, we present results based on cohort G2 and compare them to results based on the G3 and G5 gallery cohorts. Only the configuration URxD-PV1 (HD matching) is used in this experiment. As one would expect, the quality of the 3D mesh is critical for recognition performance, as shown in the results presented in Fig. 11. The rank-1 recognition rate is 86% for the cohort of moderate quality and it is only 68% for the cohort G3, which is of low quality as it contains missing data and noisy regions. The most drastic drop in the performance (only 46% rank-1 recognition) is observed for the cohort G5, which has low quality data even for full 3D face recognition.

Experiment 4: Influence of sampling range and density

To demonstrate the sensitivity of the algorithm to the predefined range of gallery sampling angles we compare the 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. The results are depicted in Figs. 12(a,b), separately for standard and non-standard poses. We also present the CMC curve based on matching only the standard poses. In a similar manner, Figs. 12(c,d) depict the influences of angular sampling density on recognition by comparing the current sampling density of 5° to sparser sampling densities of 10° and 20°. These experiments were applied on P3a and P3b cohorts with manually extracted profiles and the HD identification algorithm (URxD-PV1).

The results indicate a clear tendency for the widely sampled pose domain to be more robust on non-standard poses. For instance, rank-1 recognition rate is 78% for a wide region (current settings), 76% for a slightly narrower region and only 56% for the sampling region with 10° reduced from each side. However, when only nearly standard poses are considered, the narrow sampled pose region will

Fig. 11. Comparison of the identification performance based on 3D models with different qualities, using the same identification algorithm (URxD-PV1).

Fig. 12. Recognition results using various sampling domains: (a,c) probe cohort P2a, and (b,d) probe cohort P2b.
result in better performance. 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 the standard pose (ultra-narrow) results in 92% rank-1 recognition for nearly 56% for non-standard poses. Unlike the area of the 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.

7. Conclusions

We presented a fully automatic system (URxD-PV) 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 work, we provided methods for all steps of the integrated system for accurate landmarking, statistical shape modeling, 3D profile generation, and extraction of the profile shapes from side-view images.

We performed numerous experiments to explore the basic factors (such as the quality of 3D data, fusion schemes for multiframe recognition, and configuration of the sampling domain) that may influence the performance of an automatic 3D-aided profile-based recognition system.

Acknowledgments

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. Special thanks to Fahd Zahed for manually annotating the profiles used in our experiments.

References


Boris Efraty received a B.A. degree in Mathematics in 1999, M.Sc. degree in Industrial Engineering in 2001 and Ph.D. in 2008 from Ben-Gurion University of the Negev, Israel. He is currently employed as a Post Doctoral Researcher within the department of Computer Science at the University of Houston. His research interests are in the field of computer vision, biometrics, machine learning, pattern recognition, image analysis, and applications of statistical analysis on images.

Emil Bilgazyev received a B.S. degree in Computer Engineering in 2005 from KTU ”Manas”, MEE from University of Houston in 2008. Currently, he is working towards a Ph.D. in Computer Science. His current research is in profile-based face recognition.

Shishir Shah is an Assistant Professor in the Department of Computer Science at the University of Houston. He joined the department in 2005 and founded the Quantitative Imaging Laboratory. He received his M.S. and Ph.D. degrees in Electrical and Computer Engineering from The University of Texas at Austin. He received his B.S. degree in Mechanical Engineering from The University of Texas at Austin. Dr. Shah’s research interests are in the fields of computer vision, video analytics, biomedical image analysis, and pattern recognition.
Ioannis A. Kakadiaris received the Ptychion (B.Sc.) degree in Physics from the University of Athens, Greece, in 1989, the M.Sc. degree in Computer Science from Northeastern University, Boston, Massachusetts, in 1991, and the Ph.D. degree in Computer Science from the University of Pennsylvania, Philadelphia, in 1997. Dr. Kakadiaris (a Hugh Roy and Lillie Cranz Cullen University Professor of Computer Science, Electrical & Computer Engineering, and Biomedical Engineering) joined the University of Houston (UH) in August 1997 after completing a postdoctoral fellowship at the University of Pennsylvania. He is the founder and director of the UH Computational Biomedicine Laboratory and director of the Division of Bio-Imaging and Bio-Computation at the UH Institute for Digital Informatics and Analysis. Professor Kakadiaris research interests include biomedical image analysis, biometrics, computer vision, and pattern recognition. He is the recipient of the year 2000 NSF Early Career Development Award, UH Computer Science Research Excellence Award, UH Enron Teaching Excellence Award, James Muller VP Young Investigator Prize, and the Schlumberger Technical Foundation Award.