Automatic Face Segmentation and Facial Landmark Detection in Range Images

Maurício Pamplona Segundo, Luciano Silva, Member, IEEE, Olga Regina Pereira Bellon, Member, IEEE, and Chauã C. Queirolo

Abstract—We present a methodology for face segmentation and facial landmark detection in range images. Our goal was to develop an automatic process to be embedded in a face recognition system using only depth information as input. To this end, our segmentation approach combines edge detection, region clustering, and shape analysis to extract the face region, and our landmark detection approach combines surface curvature information and depth relief curves to find the nose and eye landmarks. The experiments were performed using the two available versions of the Face Recognition Grand Challenge database and the BU-3DFE database, in order to validate our proposed methodology and its advantages for 3-D face recognition purposes. We present an analysis regarding the accuracy of our segmentation and landmark detection approaches. Our results were better compared to state-of-the-art works published in the literature. We also performed an evaluation regarding the influence of the segmentation process in our 3-D face recognition system and analyzed the improvements obtained when applying landmark-based techniques to deal with facial expressions.

Index Terms—Face segmentation, facial landmark detection, range images, 3-D face recognition.

I. INTRODUCTION

With the development of 3-D face recognition techniques, face segmentation and landmark detection on depth information have become very important preprocessing steps for designing fully automatic recognition systems. The need for face segmentation was observed in some of the first methods aiming for 3-D face recognition, which were based on different strategies (e.g., image registration [1],[2], appearance recognition [3]–[5], and profile curves [6],[7]). In all these works, the face was considered as a rigid shape, and the same drawback was observed: recognizing faces with shape variations. It was found that even small changes between two neutral expression faces of the same subject could cause a decrease in the verification rate. Subsequent works in the literature successfully employed landmark detection to better deal with this problem.

Chang et al. [8] proposed the use of three different regions around the nose during the matching stage. Lu and Jain [9] proposed a modeling approach for expression deformation.

Bronstein et al. [10] proposed an expression-invariant face representation for recognition. Other works used similar ideas to support face recognition [11]–[13]. All the aforementioned approaches used facial landmarks to achieve their objectives.

Surface curvature classification [14] has been successfully applied for landmark detection [8],[15],[16] showing high localization rates. Colbry et al. [17] employed curvature information together with some heuristics to find facial landmarks in face images with arbitrary pose. However, this last approach required a large displacement tolerance to obtain good localization rates. Recently, Faltiemier et al. [18] used profile curves from different views of the input image to detect the nose tip in nonfrontal face images. This approach obtained high localization rates considering a small displacement tolerance. However, this approach is restricted to a single landmark.

For frontal face images, Lu and Jain [19],[20] improved precision and detection rates by combining surface curvature classification with relief curves of depth images, statistical model of landmark location, and point cornerness obtained from intensity images. In this method, a training stage was required to create a statistical model, and the use of intensity images may present some problems regarding illumination, pose, and alignment between intensity and range data [21].

To perform landmark detection, each range image must be first segmented seeking the isolation of the face region from the other parts (i.e., regions that do not belong to the face area) such as hair, neck, ears, and clothes. For images containing only one subject, many techniques have been adopted for face segmentation, like data clustering [5], skin color models [8], and histogram of depth coordinates [10], and even manual segmentation has been reported [16]. Some of these methods [5],[8] use the color information to perform segmentation and are subject to the same problems that arise when using intensity images [21].

Other works can detect more than one face per image either by creating 2-D projections of the 3-D data and applying appearance training methods [15], or performing techniques based on a boosted cascade classifier that uses color information [19] to establish the face location from the range data. However, these methods do not extract the entire face region. In [15], only the area containing the nose and the eyes is extracted, and in [19], the face segmentation technique excluded forehead and chin parts.

We present a methodology for face segmentation and facial landmark detection in range images. Our goal was to develop an automatic process to be embedded in a face recognition system using only depth information as input. Our first contribution is a
segmentation approach to find and extract the entire face region in frontal images without requiring any training stage. To this end, we combine edge detection, region clustering, and shape analysis. This approach was more accurate when compared to two other segmentation approaches [19], [21], considering the verification results of a baseline 3-D face recognition technique [22] and the verification results of our own 3-D face recognition system. Our second contribution is a method for facial landmark detection that combines surface curvature classification and depth relief curves. This approach presented an equivalent or better performance when compared to three state-of-the-art works presented in the literature [19], [20], [23], [24]. Results of the landmark detection were successfully applied to improve the performance of our automatic 3-D face recognition system under expression variations [13].

In our experiments, we used the Face Recognition Grand Challenge (FRGC) v1.0 and v2.0 databases\(^1\) and the BU-3DFE database [25], totaling 7450 images. These databases were chosen to support our experiments because they have been extensively used for research regarding 3-D face analysis [8], [12], [13], [21], [26]–[31].

The remainder of this paper is organized as follows. In Section II, we introduce our segmentation approach to extract the entire face region from an input range image. The methodology for facial landmark detection is presented in Section III. Details regarding experimental results are discussed in Section IV, followed by our final remarks in Section V.

II. FACE SEGMENTATION IN RANGE IMAGES

We developed a segmentation approach to extract the entire face region from an input range image that obeys the following constraints: 1) The image cannot contain more than one face; 2) the face must be among the closest regions to the acquisition device; 3) the face must be frontally posed; and 4) the face cannot be occluded by but except hair. To segment a face, first of all, we had to isolate the face region from background and other body parts. Although there are some constraints regarding the input image, many situations may interfere in the segmentation process, such as hairstyle, head accessories, and body parts.

To solve this problem, we designed an algorithm combining edge detection [32], region clustering [33], and shape analysis. The segmentation algorithm has two main stages: 1) the localization of the homogeneous regions in the input image and 2) the identification of the homogeneous regions that belong to the face. A visual diagram of the proposed approach is shown in Fig. 1. Details regarding each stage of the segmentation process are provided in the following sections.

A. Homogeneous Region Localization

To localize the homogeneous parts, we performed region clustering and edge detection in the input image’s depth information. Considering the databases used in our experiments, we observed that the images hardly present depth data in the background region and that the face is one of the closest regions to the acquisition system.

First, a $5 \times 5$ median filter is applied twice on the depth information to reduce noise. Then, the K-means clustering algorithm [34] is applied in the depth image to divide the depth data in the two main regions (i.e., $K = 2$): the region of interest (ROI), composed of foreground regions that are close to the acquisition device, and the rest of the data, including foreground regions that are far from the same device and all background regions that may appear in the depth data. Assuming that the face is always inside the region with the smallest depths (i.e., the ROI), we can reduce the search area, as shown in Fig. 1.

However, the region detection process alone is not enough to correctly isolate the face region from the input image regions, such as hair, neck, clothes, and head accessories. Therefore, edge detection is employed to separate these parts from the face region in the resulting K-means segmentation.

A thresholding operation was employed for edge detection in the gradient image, which is obtained by applying the Sobel operator [35] to the input depth information. A global threshold $T_G$ was defined based on the inflection point in the decreasing stretch of the gradient histogram computed for all images of the FRGC v2.0 database (i.e., the inflection point represents a significant move in the gradient histogram curve, and we considered this move as the change from homogeneous to nonhomogeneous gradient values). Fig. 2(a) shows the gradient histogram for the FRGC v2.0 database. The inflection point can be found as a negative peak in the first derivative of the histogram curve [see Fig. 2(a)], according to the following equation:

$$T_G = u \quad \text{if} \quad h'(u) < h'(v) \quad \forall v \neq u \quad (1)$$

where $h'(x)$ is the first derivative function of the histogram curve (i.e., in this paper, $T_G = 7$).

However, images with high resolution present smaller gradient values than images with lower resolution, and $T_G$ needs to

1Available at http://www.bee-biometrics.org.
be normalized according to the current image resolution. The resolution $r_F$ for an image $F$ is computed as the mean distance in the $y$-axis between neighbor points in the same column, i.e., only the $y$-axis is used because both the $y$- and $x$-axes present similar values for the same image [see Fig. 2(b) and (c)]. A global resolution $r_G$ is computed as the mean value of the resolution of all images from the FRGC v2.0 database (see Fig. 2(b); in this paper, $r_G = 0.6$), and the automatic threshold $T_F$ for an image $F$ is obtained through the following equation:

$$T_F = \frac{r_F}{r_G} T_G.$$  

A closing process [36] (i.e., two $3 \times 3$ dilation followed by one $3 \times 3$ erosion) is performed to link some broken lines created in the thresholding operation to obtain the final edge map. After that, the edge map was used to split the region resulting from K-means by combining both images through a logical AND operation, as shown in Fig. 1. The resulting image contains the homogeneous regions of the input image.

### B. Face Region Identification

In this stage, we need to identify which homogeneous regions belong to the face. To this end, these regions are submitted to a labeling process where all regions with a size below 0.5% of the image size are discarded. This percentage threshold, which was empirically defined, must be large enough to eliminate undesired regions (e.g., hair remains) and small enough to never discard a relevant region. Fig. 3(a) shows an example of homogeneous regions after labeling.

Face regions are found in the labeled image by searching for an elliptical shape, which is considered to be the most similar geometric shape to the face boundary. That is achieved by performing the distance transform [37] on the border line of each labeled region, as shown in Fig. 3, and looking for the combination of homogeneous regions presenting a shape similar to an ellipse. To evaluate the similarity between a set of regions and an ellipse, we used Algorithm 1, where the border distance, ellipse filling, and ellipse leaking are measured, with ellipses of higher size being prioritized. Fig. 4 shows all the combinations for the regions shown in Fig. 3(a). The best ellipse, according to Algorithm 1, is shown in Fig. 4(g).
Fig. 5. [(a)–(c)] Regions of the detected ellipse used to localize the face. (d) Example of a face split by mustache edges.

Algorithm 1 Pseudocode for similarity measure between homogeneous regions and ellipses.

**Input:** A set of homogeneous regions \( S = \{r_1, r_2, \ldots, r_N\} \)

**Output:** A similarity value \( V \)

1. Create a distance map \( D \) combining the distance map of each region using the following equation for each image point \( i \): \( D_i = \min \{D_i^1, D_i^2, \ldots, D_i^N\} \)
2. Obtain the bounding box of \( S \), and define an ellipse with the center, height, and width values of this bounding box
3. Considering \( P \) as the set of points in the border of the ellipse, compute the distance \( d = \sum_{p \in P} D_p / |P| \)
4. Compute the fraction \( \alpha \) of nonregion points inside the ellipse to measure the ellipse filling
5. Compute the fraction \( \beta \) of region points inside the bounding box and outside of the ellipse to measure the ellipse leaking
6. Compute the bounding box area \( A \)
7. \( V = \alpha \beta d / A \)

After the ellipse detection follows the selection of the labeled regions that belong to the face, according to Algorithm 2. In the step 2 of Algorithm 2, regions close to the central part of the face are selected, preventing the selection of neck and hair parts, which usually appear near the ellipse border. However, sometimes, the detected edges split the face into small parts, and some of these parts may not be selected by step 2. Steps 4–7 were designed to select the missing parts in these cases. Steps 4–6 are employed to select mouth or cheek parts that may be split when the face presents a high level of expression or mustache, as shown in Fig. 5(d). Step 7 is applied to select the forehead region and prevent the selection of hair or head accessories.

Algorithm 2 Pseudocode for face region selection.

**Input:** A set of labeled regions \( S = \{r_1, r_2, \ldots, r_N\} \) and an ellipse \( E \)

**Output:** A set of face regions \( F \)

1. Define a circle \( C \) in the ellipse center with a diameter equal to two thirds of the ellipse width [see Fig. 5(a)]
2. Add to \( F \) all labeled regions in \( S \) with at least one pixel inside \( C \)
3. Divide the ellipse in four slices: the top slice \( E_T \), the bottom slice \( E_B \), the left slice \( E_L \), and the right slice \( E_R \) [see Fig. 5(b)]
4. Add to \( F \) the region in \( S \) with the highest number of pixels inside \( E_B \)
5. Add to \( F \) the region in \( S \) with the highest number of pixels inside \( E_L \)
6. Add to \( F \) the region in \( S \) with the highest number of pixels inside \( E_R \)
7. Considering \( r_i \in S \) as the region with the highest number of pixels inside \( E_T \), \( r_i \) will be selected if it is also the region with the highest number of pixels inside the inferior part of \( E_T \) [see Fig. 5(c)] or if it has twice the number of pixels of the most common region in the inferior part of \( E_T \)

The regions selected as face are represented as a binary image that indicates the face location in the input image. However, this binary image may lack information that can be fulfilled by linking disconnected points in the same column or in the same row. After that, we performed a logical AND between the resulting binary image and the input range image to obtain the final segmentation of the face region.

III. FACIAL LANDMARK DETECTION

Facial landmark detection is an important step to optimize some tasks regarding recognition, for instance, face registration, facial pose estimation, and location of the face rigid regions [8].

We are interested in a small group of facial landmarks: nose tip, eye corners, and nose corners. These landmarks were selected because they allow precise location of the rigid face regions such as the nose and forehead. To localize them, we combined relief curves [18], [38] obtained from the depth data and from the resulting image of the surface curvature analysis [8], [39]. A visual diagram of the landmark detection approach is shown in Fig. 6.

A. Curvature-Based Surface Classification

According to [40], by analyzing the surface curvature information of a face, we can observe that some face regions (e.g., the nose tip and inner eye corners) present the same curvature.
Fig. 7. Examples of surface classification applied on images of the same person displaying different expressions: (a), (b), and (c) are face segmentation results; (d), (e), and (f) show classification of surfaces from (a), (b), and (c), respectively. Regions in light gray are peaks, and the white ones are pits.

To compute the curvature type for each point of the face image, we defined a local surface in a small neighborhood \( N \times N \), as in [41] (i.e., \( n = 11 \)). Then, we used a least square fitting technique to calculate the coefficients of a biquadratic approximation of the surface, and we estimated the Gaussian curvature \( (K) \) and the mean curvature \( (H) \) by computing the partial derivatives, according to [39]. Once we had both \( K \) and \( H \) values, we could classify the local surface according to Table I [39].

After classifying all points of the input image, two surface types were used in our approach to find the desired facial landmarks: peak and pit. Eye corners present a pitlike surface, and the nose tip presents a peaklike surface, as can be seen in Fig. 7. Nose corners may present different classification over different facial expressions, so another face characteristic was employed to find them, as explained hereinafter.

### B. Landmark Detection

Our landmark detection has two stages: First, we find the \( y \)-coordinates for the nose tip and eyes, and then, we find the \( x \)-coordinates for the nose tip, nose corners, and inner eye corners.
case, the nose tip y-coordinate. To detect the nose corners, we calculated the gradient information of this curve in order to identify one peak on each side of the nose tip, as shown in Fig. 11.

To find the x-coordinates of the eye corners, we computed the x-projection of the curvature image by calculating the percentage of pit curvature points for every column in a set of neighbor rows centered in the eye y-coordinate. The left eye corner x-coordinate is the beginning of the first peak in this x-projection, and the right eye corner x-coordinate is the end of the second peak in the same projection, as shown in Fig. 12.

IV. Experimental Results

The experiments presented in this section were designed to evaluate the accuracy of our approaches for face segmentation and landmark detection and to analyze their influence in a 3-D face recognition system. Three databases were used for these experiments. The FRGC database versions 1.0 and 2.0, totaling 4950 images of 557 subjects with variations of facial expression, resolution, pose, and other characteristics, such as different hairstyles. Images have the size of 640 × 480 and were acquired by a Minolta Vivid 910 laser scanner and the BU-3DFE database [25], totaling 2500 images of 100 subjects with high facial expression variations. Only surface meshes are available in the original data of the BU-3DFE database, so we converted these models to frontally posed range images through point sampling considering a 0.6-mm resolution.

All recognition experiments were performed using a standard baseline approach based on the iterative closest point (ICP) [22] for 3-D face matching and our automatic 3-D face recognition system [13] based on simulated annealing (SA) [42] and the surface interpenetration measure (SIM) [43]. The ICP was employed in our experiments because many recognition works in the literature are based on it. In our implementation, an initial solution was obtained by aligning the centers of mass of the two face images, and then, a coarse alignment was performed by applying the ICP using 5% of the points of the test image. In the last step, a fine alignment was obtained by increasing the percentage of sample points of the test image to 25%. The final similarity value was obtained by using the M-estimator Sample Consensus (MSAC) [44] over the error between corresponding points in the overlapping area of the registered images.
PAMPLONA SEGUNDO et al.: FACE SEGMENTATION AND FACIAL LANDMARK DETECTION IN RANGE IMAGES

Fig. 13. Incorrect segmentation results caused by [(a)–(e)] irrelevant region selection and (f) missing parts of the face.

The MSAC threshold was set to 3.0 mm. Similarly, in our recognition system, an initial solution is obtained by aligning the centers of mass of the two face images. Then, a coarse alignment is performed using an SA-based searching procedure to minimize the MSAC, and a fine alignment is obtained by an SA-based searching procedure to maximize the SIM [13].

A. Segmentation Results

Our segmentation approach extracted the entire face region correctly in 99.3% of the FRGC v1.0 database (936 of 943), 99.7% of the FRGC v2.0 database (3993 of 4007), and 96.1% of the BU-3DFE database (2403 of 2500), as confirmed by a visual inspection of the results (i.e., resulting faces containing irrelevant data or missing face parts were considered as incorrect segmentation). Errors were mainly caused when regions of neck, hair, or head accessories were not separated from the face by edge detection. Some occurrences are shown in Fig. 13. In Fig. 13(a)–(e), irrelevant regions were not correctly separated from the face, and in Fig. 13(f), the forehead region was not selected as face. The average time to segment one face was around 1 s in a Pentium-D 3.4-GHz processor. Fig. 14 shows some examples of final face segmentation by using the proposed approach.

To evaluate the influence of segmentation results in 3-D face recognition systems, we compared our approach to two other segmentation techniques [19], [21]. The first one is a baseline segmentation method based on a boosted cascade classifier developed by Viola and Jones [45], available in the OpenCV Library\(^2\), that uses intensity image for face extraction. The background was removed by using the range map, as

\(^2\)Available at http://www.intel.com/technology/computing/opencv.
proposed in [19], to reduce the search area and the number of false detections. The face area corresponds to a set of range points inside the detected square, as shown in Fig. 15. Some examples of the final segmentation for this technique are shown in Fig. 16. The second segmentation technique was proposed by Mian et al. [21] and consists of extracting a spherical area with a radius of 80 mm around the nose tip, as shown in Fig. 17(a).

Face images after segmentation using this technique are shown in Fig. 17(b)–(d).

For this comparison, we used only 933 noiseless images with neutral expression from FRGC v2.0, eliminating effects caused by expression variations and noise. With this configuration, the entire face region is relevant to the recognition task, and the segmentation technique that maximizes the extracted face region and minimizes the amount of irrelevant data will present the best verification results. Thus, all these noiseless images with neutral expression were segmented by all the three techniques. Nonface regions detected as face by the baseline approach were manually removed, and a ground truth nose tip location for the images of the database was used for Mian’s segmentation technique.

Then, we computed the similarity value for each combination of two images segmented by the same approach (864,578 combinations between images from different subjects and 4978 between images from the same subject), and Fig. 18 shows the verification results for the three approaches. As observed, our

### Table II

<table>
<thead>
<tr>
<th>Landmarks</th>
<th>NT</th>
<th>LE</th>
<th>RE</th>
<th>LN</th>
<th>RN</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean (mm)</td>
<td>2.69</td>
<td>4.26</td>
<td>4.15</td>
<td>6.16</td>
<td>7.22</td>
</tr>
<tr>
<td>[19]-Mean (mm)</td>
<td>5.00</td>
<td>5.70</td>
<td>6.00</td>
<td>N.A.</td>
<td>N.A.</td>
</tr>
<tr>
<td>[24]-Mean (mm)</td>
<td>2.18</td>
<td>5.59</td>
<td>4.74</td>
<td>N.A.</td>
<td>N.A.</td>
</tr>
<tr>
<td>Std (mm)</td>
<td>2.14</td>
<td>3.51</td>
<td>3.13</td>
<td>3.09</td>
<td>2.76</td>
</tr>
<tr>
<td>[19]-Std (mm)</td>
<td>2.40</td>
<td>3.00</td>
<td>3.30</td>
<td>N.A.</td>
<td>N.A.</td>
</tr>
<tr>
<td>[24]-Std (mm)</td>
<td>6.83</td>
<td>16.08</td>
<td>9.76</td>
<td>N.A.</td>
<td>N.A.</td>
</tr>
<tr>
<td>T10</td>
<td>99.89</td>
<td>97.14</td>
<td>98.52</td>
<td>90.56</td>
<td>90.03</td>
</tr>
<tr>
<td>[24]-T10</td>
<td>99.50</td>
<td>98.50</td>
<td>98.50</td>
<td>N.A.</td>
<td>N.A.</td>
</tr>
<tr>
<td>T15</td>
<td>99.89</td>
<td>99.47</td>
<td>99.68</td>
<td>99.79</td>
<td>99.47</td>
</tr>
</tbody>
</table>

### Table III

<table>
<thead>
<tr>
<th>Landmarks</th>
<th>NT</th>
<th>LE</th>
<th>RE</th>
<th>LN</th>
<th>RN</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean (mm)</td>
<td>2.73</td>
<td>3.69</td>
<td>3.35</td>
<td>4.83</td>
<td>5.84</td>
</tr>
<tr>
<td>Std (mm)</td>
<td>1.39</td>
<td>2.26</td>
<td>2.33</td>
<td>2.03</td>
<td>1.73</td>
</tr>
<tr>
<td>T10</td>
<td>99.95</td>
<td>98.88</td>
<td>99.15</td>
<td>99.55</td>
<td>99.15</td>
</tr>
<tr>
<td>T15</td>
<td>99.95</td>
<td>99.65</td>
<td>99.63</td>
<td>99.95</td>
<td>99.95</td>
</tr>
<tr>
<td>[23]-T10</td>
<td>90.00</td>
<td>90.00</td>
<td>90.00</td>
<td>N.A.</td>
<td>N.A.</td>
</tr>
</tbody>
</table>

### Table IV

<table>
<thead>
<tr>
<th>Landmarks</th>
<th>NT</th>
<th>LE</th>
<th>RE</th>
<th>LN</th>
<th>RN</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean (mm)</td>
<td>1.87</td>
<td>6.33</td>
<td>6.33</td>
<td>6.66</td>
<td>6.49</td>
</tr>
<tr>
<td>Std (mm)</td>
<td>1.12</td>
<td>4.82</td>
<td>5.04</td>
<td>3.36</td>
<td>3.40</td>
</tr>
<tr>
<td>T10</td>
<td>99.88</td>
<td>84.32</td>
<td>82.40</td>
<td>83.56</td>
<td>83.56</td>
</tr>
<tr>
<td>T15</td>
<td>99.88</td>
<td>94.60</td>
<td>95.20</td>
<td>99.08</td>
<td>99.52</td>
</tr>
</tbody>
</table>
B. Landmark Detection Results

After performing segmentation, the faces were submitted to our landmark detection process without any adjustment (i.e., incorrect results were not rectified). The facial landmarks were manually labeled twice by two different people in order to evaluate their localization displacement, and the ground truth position to each landmark was the average of these two values.

The localization displacement was estimated as the 3-D Euclidean distance between the ground truth position and the position obtained through our detection approach. To show our improvement in precision, some statistics of the localization displacement obtained by our approach are presented in Tables II–IV, together with some results obtained by Lu and Jain [19], [20], Romero-Huertas and Pears [23], and Yu and Moon [24]. The FRGC v2.0 database was employed in the experiments presented in [23], and the FRGC v1.0 database was used in [19] and [24]. The mean error and the standard deviation were computed considering the localization displacement between the ground truth and our automatic detection approach. Detection rates were obtained by counting the percentage of detected landmarks whose localization displacement was below the tolerance threshold. As can be seen, our approach presented smaller mean displacement and standard deviation values, and consequently, nose tip and inner eye corners were correctly detected in more images than in other approaches when using small tolerance values.

We computed the localization displacement histogram (LDH) for each landmark using the images of the FRGC v2.0 database. The obtained LDH graphics for our detection of nose tip, inner eye corners, and nose corners are shown in Fig. 19.

The average time spent to locate all desired facial landmarks in a segmented face image was around 0.3 s in a Pentium-D 3.4-GHz processor. Some results of the landmark detection are shown in Fig. 20.

Then, these landmarks were employed to make a 3-D recognition system able to recognize face images under expression variations. To this end, we performed the matching of all neutral images of the FRGC v2.0 database against each other, as well as images with nonneutral expression against neutral images. This experiment reproduced a genuine face recognition system, where we have uncontrolled acquired images matched against controlled images (i.e., noiseless and neutral expression) previously obtained and stored in a database.

To improve the matching, two different landmark-based techniques were considered: The first one extracts different regions of the face that are less affected by facial expression variations in 3-D face recognition [i.e., the nose and upper head regions, as shown in Fig. 21(a)–(c)], and the second one uses the entire face region, but it is divided in nine sectors, and the sectors containing invariant regions, highlighted in Fig. 21(d), are prioritized during the matching process [13].

Face images were matched against each other using different regions: the entire face, the nose region, the upper head region,
cause changes in forehead and nose areas (e.g., frown, disgust, and sad expressions). The combination of rigid solutions presented better results than using a single region for recognition in more than a half of the expressions analyzed, and a higher stability for different expressions.

Table VI shows the verification rates for our face recognition system at $10^{-3}$ FAR using the entire face region, the circular nose area, the elliptical nose area, the upper head region, the sector approach [Fig. 21(d)] [13], and the combination between landmark-based techniques by the sum rule [47]. Our recognition system presents a better performance in comparison to the ICP-based recognition system, and the combination of landmark-based techniques obtained the highest verification rates in almost all expressions analyzed.

Our experimental results show an expressive increase of verification rates when landmark-based techniques were applied for face recognition. By comparing the use of the entire face region for recognition against the combination of the rigid regions automatically extracted when using ICP, we obtained rates up to 46% higher for high expression level (i.e., SUR 2), 36% higher for low expression level (i.e., SUR 1), and 5% higher for neutral expression. By using our recognition system, we obtained rates up to 18% higher for high expression level (i.e., HAP 2) and 10% higher for low expression level (i.e., PUF).

### V. Conclusion

We have presented automatic techniques for face segmentation and landmark detection in frontally posed range images using only depth information as input. Our segmentation approach extracts the entire face region by combining edge detection, region clustering, and shape analysis. Our landmark detection approach uses surface curvature classification and relief curves of face images to detect the nose tip, nose corners, and inner eye corners.

The entire face region was correctly extracted by our approach in 99.3%, 99.7%, and 96.1% of the images of the FRGC v1.0, FRGC v2.0, and BU-3DFE databases, respectively. There was an increase in the verification rate of an ICP-based recognition system of 2% at $10^{-3}$ FAR in comparison to the two other segmentation approaches [19], [23], [24] and also an increase in the verification rate of 0.4% at $10^{-3}$ FAR when using our face recognition approach.

The desired landmarks were correctly found in 99% of the FRGC databases and also 99% for nose landmarks and 93% for eye corners in the BU-3DFE database. The nose tip was the most precise landmark, presenting the lowest mean error and the highest detection rates for all databases. Our landmark detection approach obtained equivalent or better performance when compared to three other techniques [19], [23], [24] and is also robust to facial expressions, which can be seen in our experimental results for FRGC v2.0 and BU-3DFE databases. The detected landmarks were also employed to extract different rigid regions of the face. These rigid regions were employed in a recognition experiment in order to evaluate the improvement obtained in verification rates when face images containing expression variations were used as probe images. Verification rates up to 46% higher were obtained for images with high

---

### TABLE V

<table>
<thead>
<tr>
<th>Method</th>
<th>F</th>
<th>C</th>
<th>E</th>
<th>U</th>
<th>C+E+U</th>
</tr>
</thead>
<tbody>
<tr>
<td>NEU</td>
<td>93.33</td>
<td>97.95</td>
<td>98.35</td>
<td>93.27</td>
<td>98.19</td>
</tr>
<tr>
<td>HAP 1</td>
<td>66.94</td>
<td>73.31</td>
<td>83.28</td>
<td>70.89</td>
<td>82.40</td>
</tr>
<tr>
<td>HAP 2</td>
<td>23.44</td>
<td>37.30</td>
<td>48.85</td>
<td>32.87</td>
<td>48.28</td>
</tr>
<tr>
<td>OPM 1</td>
<td>72.96</td>
<td>91.48</td>
<td>92.59</td>
<td>86.30</td>
<td>95.19</td>
</tr>
<tr>
<td>OPM 2</td>
<td>66.58</td>
<td>81.46</td>
<td>91.33</td>
<td>81.46</td>
<td>92.01</td>
</tr>
<tr>
<td>FRO 1</td>
<td>74.22</td>
<td>85.16</td>
<td>88.28</td>
<td>84.38</td>
<td>90.63</td>
</tr>
<tr>
<td>FRO 2</td>
<td>59.04</td>
<td>78.19</td>
<td>74.47</td>
<td>64.36</td>
<td>84.04</td>
</tr>
<tr>
<td>DIS 1</td>
<td>87.21</td>
<td>97.67</td>
<td>97.67</td>
<td>90.70</td>
<td>98.84</td>
</tr>
<tr>
<td>DIS 2</td>
<td>52.07</td>
<td>23.17</td>
<td>43.90</td>
<td>37.80</td>
<td>37.20</td>
</tr>
<tr>
<td>SUR 1</td>
<td>57.43</td>
<td>89.11</td>
<td>93.32</td>
<td>73.76</td>
<td>93.81</td>
</tr>
<tr>
<td>SUR 2</td>
<td>42.53</td>
<td>80.84</td>
<td>80.40</td>
<td>77.68</td>
<td>88.84</td>
</tr>
<tr>
<td>SAD 1</td>
<td>75.11</td>
<td>94.05</td>
<td>87.67</td>
<td>82.16</td>
<td>93.61</td>
</tr>
<tr>
<td>SAD 2</td>
<td>57.95</td>
<td>79.55</td>
<td>70.83</td>
<td>63.64</td>
<td>78.41</td>
</tr>
<tr>
<td>PUF</td>
<td>33.94</td>
<td>42.87</td>
<td>32.58</td>
<td>32.49</td>
<td>53.39</td>
</tr>
<tr>
<td>OTH</td>
<td>63.92</td>
<td>72.15</td>
<td>83.86</td>
<td>72.15</td>
<td>83.84</td>
</tr>
</tbody>
</table>

### TABLE VI

<table>
<thead>
<tr>
<th>Method</th>
<th>F</th>
<th>C</th>
<th>E</th>
<th>U</th>
<th>C+E+E+S</th>
</tr>
</thead>
<tbody>
<tr>
<td>NEU</td>
<td>99.40</td>
<td>99.36</td>
<td>99.52</td>
<td>99.32</td>
<td>99.96</td>
</tr>
<tr>
<td>HAP 1</td>
<td>92.52</td>
<td>94.43</td>
<td>95.31</td>
<td>92.08</td>
<td>99.90</td>
</tr>
<tr>
<td>HAP 2</td>
<td>76.89</td>
<td>85.25</td>
<td>89.51</td>
<td>75.41</td>
<td>94.10</td>
</tr>
<tr>
<td>OPM 1</td>
<td>99.26</td>
<td>96.30</td>
<td>96.30</td>
<td>99.26</td>
<td>97.04</td>
</tr>
<tr>
<td>OPM 2</td>
<td>94.73</td>
<td>96.94</td>
<td>99.15</td>
<td>94.05</td>
<td>99.66</td>
</tr>
<tr>
<td>FRO 1</td>
<td>95.31</td>
<td>93.75</td>
<td>92.19</td>
<td>92.19</td>
<td>87.50</td>
</tr>
<tr>
<td>FRO 2</td>
<td>92.02</td>
<td>86.70</td>
<td>87.23</td>
<td>90.96</td>
<td>90.96</td>
</tr>
<tr>
<td>DIS 1</td>
<td>100.00</td>
<td>100.00</td>
<td>97.67</td>
<td>100.00</td>
<td>97.67</td>
</tr>
<tr>
<td>DIS 2</td>
<td>76.83</td>
<td>43.90</td>
<td>57.32</td>
<td>73.17</td>
<td>79.27</td>
</tr>
<tr>
<td>SUR 1</td>
<td>93.07</td>
<td>97.52</td>
<td>99.51</td>
<td>94.55</td>
<td>95.90</td>
</tr>
<tr>
<td>SUR 2</td>
<td>88.23</td>
<td>96.67</td>
<td>97.72</td>
<td>91.21</td>
<td>87.17</td>
</tr>
<tr>
<td>SAD 1</td>
<td>93.39</td>
<td>99.12</td>
<td>100.00</td>
<td>95.15</td>
<td>92.07</td>
</tr>
<tr>
<td>SAD 2</td>
<td>86.36</td>
<td>93.94</td>
<td>94.70</td>
<td>84.10</td>
<td>96.97</td>
</tr>
<tr>
<td>PUF</td>
<td>87.33</td>
<td>88.69</td>
<td>91.40</td>
<td>91.63</td>
<td>97.51</td>
</tr>
<tr>
<td>OTH</td>
<td>86.61</td>
<td>94.94</td>
<td>93.04</td>
<td>90.51</td>
<td>89.87</td>
</tr>
</tbody>
</table>

and the entire face with prioritized sectors. The matching results for the ICP-based and our recognition approaches are shown separately in Tables V and VI.

Table V shows the verification rates for the ICP-based recognition system considering a false acceptance rate (FAR) at $10^{-3}$ using the entire face region, the circular nose area [see Fig. 21(a)] [8], the elliptical nose area [Fig. 21(b)] [8], the upper head region [Fig. 21(c)] [46], and the combination of the rigid regions by the sum rule [47]. For each one, different rates were computed using faces with different expressions: neutral (NEU), happy (HAP), open mouth (OPM), frown (FRO), disgust (DIS), surprised (SUR), sad (SAD), puffy cheeks (PUF), and other types (OTH). Some of these expressions present low and high levels of expression, labeled as 1 and 2, respectively. The recognition rate using the entire face region is still being affected by facial expressions, and landmark-based techniques presented a different behavior according to changes in the face shape. The circular and elliptical nose areas are affected by expressions that cause facial changes near the nose location (e.g., disgust and puffy cheek expressions), and the upper head technique presented some difficulties with expressions that...
levels of expressions, 36% higher for images with normal expressions, and 5% higher for images with neutral expression.

Finally, our results show the benefits of both face detection and landmark detection in a face recognition system and how much they improve its verification rates. Our presented approaches were encapsulated in our automatic 3-D face recognition system [13] based on SA + SIM, and we obtained a verification rate of 99.9% at zero FAR for neutral images and 96.5% at 10−3 FAR when all images of the FRGC v2.0 database were matched against each other (i.e., including matches between images with nonneutral expressions).

ACKNOWLEDGMENT

The authors would like to thank Dr. J. Phillips and Dr. P. Flynn for allowing them to use the FRGC images, Dr. L. Yin and The Research Foundation of State University of New York for allowing them to use the BU-3DFE images, and G. Vernize and P. Ferreira for the manual labeling of the landmarks in all images from the FRGC database.

REFERENCES


Maurício Pamplona Segundo received the B.S. degree in computer science from the Universidade Federal do Paraná (UFPR), Curitiba, Brazil, in 2007, where he has been working toward the M.Sc. degree in the IMAGO Research Group since 2008. Since 2005, he has already been with the IMAGO Research Group. His research interests include biometrics, automatic feature extraction, pattern recognition, image segmentation, image registration, and security. Mr. Pamplona Segundo’s undergraduate research work received awards from the Brazilian Computer Society.

Luciano Silva (M’99) received the B.S. and M.Sc. degrees in informatics from the Universidade Federal do Paraná (UFPR), Curitiba, Brazil, in 1998 and 2000, respectively, and the Ph.D. degree in electrical engineering and industrial informatics from the Universidade Tecnologica Federal do Paraná (UTFPR), Curitiba, Brazil, in 2003. He spent a year (2002–2003) in a doctoral stage at the Ohio State University, Columbus. In 1996, he joined the IMAGO Research Group, UFPR. In 2009, he spent a sabbatical year with the University of Southern California, Los Angeles. In 2004, he joined the faculty at UFPR, where he is currently an Associate Professor. His research interests include computer vision, biometrics, range images, and 2-D/3-D image analysis.

Dr. Silva is a member of the Brazilian Computer Society. His undergraduate work, M.Sc. dissertation, and Ph.D. thesis all received awards from the Brazilian Computer Society, as well as some research works he advised.

Olga Regina Pereira Bellon (M’99) received the B.S. degree in electrical engineering from the Universidade Federal do Espírito Santo (UFES), Vitória, Brazil, in 1987, and the M.Sc. and Ph.D. degrees in electrical engineering from the Universidade Estadual de Campinas (UNICAMP), Campinas, Brazil, in 1990 and 1997, respectively.

In 1995, she founded the IMAGO Research Group, Universidade Federal do Paraná (UFPR), Curitiba, Brazil, where she is currently a full Professor. From 1997 to 2000, she was the Head of the Graduate Program of Informatics. From 2000 to 2002, she was the General Research Coordinator of UFPR. She spent two sabbatical years in the U.S., one at the Ohio State University, Columbus, in 2003, and the other at the University of Southern California, Los Angeles, in 2009. In 2004, she was the General Chair of the XVII Brazilian Symposium on Computer Graphics and Image Processing (SIBGRAPI) joint with II Ibero-American Symposium on Computer Graphics (SIACG). Her main research interests include range images, computer vision and graphics, pattern recognition, biometrics, digital preservation of natural and cultural objects, and virtual museums.

Dr. Bellon is a member of the IEEE Computer Society, Brazilian Computer Society, and Brazilian Health Informatics Society. In 2009, she joined the Educational Committee of the International Association of Pattern Recognition. She was elected President of the Special Committee in Computer Graphics and Image Processing of the Brazilian Computer Society, in 2007 and was reelected in 2009. She is the Program Chair of the 23rd Conference on Graphics, Patterns and Images in 2009.

Chauã C. Queirolo received the B.S. and M.Sc. degrees in computer science from the Universidade Federal do Paraná (UFPR), Curitiba, Brazil, in 2007 and 2009, respectively.

Since 2004, he has been with the IMAGO Research Group, UFPR. His research interests include range imaging, computer vision, computer games, and dinosaurs.

Mr. Queirolo’s undergraduate research work and M.Sc. dissertation had both received awards from the Brazilian Computer Society.