Rotated Profile Signatures for Robust 3D Feature Detection

Timothy C. Faltemier, Kevin W. Bowyer, and Patrick J. Flynn

Abstract—While recent years have seen progress in face recognition from 3D images, non-frontal head pose is still a challenge to existing techniques. We introduce a new system for 3D face recognition that is robust to facial pose variation. Large degrees of facial pose variation may lead to a significant fraction of the features visible in frontal images being occluded. High accuracy automatic feature and pose detection is performed by a new technique called Rotated Profile Signatures (RPS). Experiments are performed on the largest available database of 3D faces acquired under varying pose. This database contains over 7,300 total images of 406 unique subjects gathered at the University of Notre Dame. Experimental results show that the RPS detection algorithm is capable of performing nose detection with greater than 96.5% accuracy across the pose variation represented in the data set used.

I. INTRODUCTION

Scenarios containing differences in expression, pose, and lighting must be addressed for 3D face recognition to become successful in less constrained environments. Many current 3D face recognition algorithms are able to automatically locate fiducial points based on the assumption that they will be provided a frontal image with a neutral expression. For example, one approach for performing nose detection is to assume that it is the point closest to the camera [4], [2]. Such a heuristic enables fast detection of the hypothesized nose tip. Under non-frontal pose, however, the heuristic can fail. Figure 1 shows examples where this assumption is valid (a,b) and when it is not (c,d). Another approach to automatic nose detection is to use the curvature information at each point on the face [5], [6], [7], [8], [9]. The use of curvature information solves many problems associated with changes in pitch (up/down). However, noise, holes, and changes in yaw (left/right) can be problematic. The Iterative Closest Point (ICP) matching algorithm used in many 3D face recognition systems has been shown to perform poorly when an accurate initial alignment is not provided. Many times, the initial alignment is provided by automatically detected feature points in the image. This suggests that an application’s recognition performance is often limited by the accuracy of the feature detection module.

In this paper, we describe a system for 3D face recognition that is more robust to non-frontal face poses. It includes an algorithm that is able to locate the nose tip automatically in the presence of pose or expression variation and occlusion with a high level of accuracy using a technique we call Rotated Profile Signatures (RPS). This method rotates the 3D face over a 180° interval in 5° increments, extracting the rightmost “profile” points on the image at each step. These profiles are then matched to a variety of nose models, as described in Section IV, resulting in a similarity score. As the nose is rotated into view, the similarity score reaches a minimum, indicating the correct nose location.

The remainder of the paper is organized as follows. Section II gives an overview of related work in the area of 3D face recognition. Section III discusses experimental materials and methods. The RPS method for efficient and accurate nose detection is introduced and results attained on the NDOff2007 data set are discussed in Section IV. Finally, Section V provides conclusions and discussion.

II. RELATED WORK

A recent survey of research on face recognition using 3D data is given in [12]. This section focuses on selected prior work that is most closely related to 3D feature detection and recognition.

Lu et al. [13] propose a method of feature extraction based on the directional maximum in a 3D image. A nose profile is represented by different subspaces and a nearest neighbor approach is used to select the best candidates for the nose tip. Of the nose candidates, the point that best fits the statistical feature location model (i.e. the nose should be below the
eyes and above the mouth) is selected as the final nose tip. The authors claim recognition results similar to those achieved from manually marked features on the FRGC v1.0 face image data set (953 frontal scans from 277 subjects) and the MSU data set (300 scans from 100 subjects with varying changes in yaw). The authors state that this method is only robust to changes in yaw, and that changes in pitch would result in “an expensive brute force search.”

Mian et al. [14] propose an expression invariant approach to 3D face recognition and report results on the FRGC v2.0 data set [1]. They perform automatic nose detection by slicing the 3D image horizontally, smoothing uniformly, and filling holes using linear interpolation. A circle centered at the maximum value of the slice is used to find the triangle with the largest area. The three triangle points consist of the circle center and the locations where the circle intersects the slice. A line is fit to the candidate nose tips, which should follow the nose bridge. The point with the maximum confidence level, based on the triangle altitude, is selected as the final nose tip. Because this method relies heavily on the triangle altitude for nose tip determination, it may perform poorly in the presence of large changes in the pitch or roll of an image.

Xu et al. [15] propose an approach for locating the nose tip in 3D facial data. Their method uses a hierarchical filtering scheme combining two rules to extract the points that distinguish the nose from other salient points. The first rule states that the nose tip will be the highest point in a certain direction that is determined by finding the normals on the face. This rule eliminates many points, leaving a limited number of candidate points (the chin, the forehead, the cheeks, hair, etc.). The next rule attempts to model the cap-like shape on the nose tip itself. Each candidate point is characterized by a feature vector containing the mean and variance of its neighboring points. The vectors are projected into mean-variance space and a Support Vector Machine (SVM) is used to determine the boundary between nose tips and non-nose tips. The authors note that this rule also is challenged by wrinkles, clothing, or other cap-like areas on the face. The authors use three databases to test their algorithm. The largest database, the 3D Pose and Expression Face Models (3DPEF), contains 300 images of 30 subjects with small amounts of changes in pitch, yaw, and roll and a 99.3% nose detection rate is reported.

Rajwade et al. [16] demonstrate a method for automatic pose detection and correction using support vector regression on wavelet sub-bands. This technique is able to classify the 3D pose of subjects in an identity-invariant manner with an accuracy of ±9° in the x and y directions. The authors report correct classification results of up to 99% using data from two frontal data sources: the Freiburg database [17], and FRGC v1.0 data set. The authors synthetically create different 3D poses by rotating the complete 3D models from 0° to 90° around the Y axis and -30° to 30° around the X axis. This method is not representative of realistic biometric acquisition and may be challenged when only partial face data may be available due to variation in facial pose.

III. Experimental Materials and Methods

A. The NDOff2007 3D Face Data set

The images for the experiments performed in this paper are taken from the largest known 3D pose variation face database available at the time of writing: NDOff2007. Standard 3D face databases such as the Face Recognition Grand Challenge (FRGC) v2.0 [1] do not contain a significant component of non-frontal pose images. The NDOff2007 data set contains a total of 6,911 non-frontal images containing neutral expressions and a single frontal neutral image for each of 406 distinct subjects. Examples of different poses in the NDOff2007 data set can be seen in Figures 2 and 3 and the number of images per pose are found in Table I. The (0°, 0°) pitch and yaw values correspond to frontal pose. The pose space is not uniformly sampled; however, due to facial symmetry, results from the left should be similar to those from the right.

The images were acquired with a Minolta Vivid 910 range scanner [18]. The Minolta 910 uses a laser stripe and triangulation to construct a range scan. The image is initially in a range image format of 640x480 with a flag field indicating 0 if a 3D point is not present and 1 if it is. Both color texture (r, g, b) and 3D location (x, y, z) are produced, but not simultaneously, as the laser stripe requires a few seconds to cross the face and the color image is taken after the shape. This can result in texture misregistration or mesh distortion from subject movement. The number of 3D points on a frontal face image is typically around 110,000, and depends on the lens used as well as sensor-to-subject distance.
TABLE I

<table>
<thead>
<tr>
<th>Pitch</th>
<th>Yaw</th>
<th>90</th>
<th>60</th>
<th>45</th>
<th>30</th>
<th>15</th>
<th>0</th>
<th>-15</th>
<th>-30</th>
<th>-45</th>
<th>-60</th>
<th>-90</th>
</tr>
</thead>
<tbody>
<tr>
<td>60</td>
<td>140</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>45</td>
<td></td>
<td>163</td>
<td>133</td>
<td>165</td>
<td>88</td>
<td>160</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>30</td>
<td></td>
<td>204</td>
<td>193</td>
<td>186</td>
<td>291</td>
<td>203</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15</td>
<td></td>
<td>205</td>
<td>183</td>
<td>187</td>
<td>189</td>
<td>192</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>175</td>
<td>158</td>
<td>155</td>
<td>173</td>
<td>186</td>
<td>406</td>
<td>180</td>
<td>177</td>
<td>160</td>
<td>173</td>
<td>170</td>
<td></td>
</tr>
<tr>
<td>-15</td>
<td></td>
<td>163</td>
<td>171</td>
<td>123</td>
<td>140</td>
<td>152</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-30</td>
<td></td>
<td>154</td>
<td>165</td>
<td>173</td>
<td>172</td>
<td>175</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-45</td>
<td></td>
<td>183</td>
<td>161</td>
<td>141</td>
<td>160</td>
<td>157</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Fig. 3. Example ranges images corresponding to Figure 2 (Subject 05078). The images show the changes in yaw (90° to -90°), and changes in pitch (45° to -45°).

B. Data Preprocessing

Our algorithm operates automatically using the 3D shape from a frontal view of the face and its corresponding 2D color image for skin detection. First, hole-filling is performed in the range image by finding the “missing” points (with a zero flag value) that are surrounded by 4 or more “valid” points. Linear interpolation is used to find the x, y, and z coordinates based on the surrounding points. Next, a 3x3 median filter is used to smooth the data and remove spikes in the z-coordinate.

Finally, we use a skin detection algorithm described by Boehnen et al. [19] to minimize the possibility of the RPS algorithm incorrectly finding the nose due to clothing or hair. This algorithm first segments the foreground information from the background using the z-coordinate of the range image data. Then it performs skin classification on each pixel in 2D in the YCbCr color space and masks out the other regions as non-valid. To fill holes, they find the maximum and minimum x-coordinates of the skin pixels on each row.

Using these boundary values, they classify all pixels between these points as valid skin pixels.

IV. Rotated Profile Signatures

In this section, we describe a new technique for location of the nose tip in 3D face images containing a wide range of poses. The following steps are performed for θ ∈ {0, 5, 10, ..., 180} and a total of 37 profile signatures are collected.

• Assume set Pθ is a set of points sorted by y-coordinate p.y forming ordered pairs of (p.x, p.y, p.z).
• Let Pθ,i be the set p = (p.x, p.y, p.z) such that p ∈ Pθ and ⌊p.y⌋ = i. Pθ,i is the subset of Pθ all of whose members have a y-coordinate that quantizes to an integer i, ymin,i ≤ i ≤ ymax,i (recall measurement units are in mm).
• Let j∗ = arg max Pθ,i,j.x be the index of the point in Pθ,i with the largest x-coordinate. Each point located at j∗ is stored in Sθ,i forming the profile signature for the given rotation.

The 3D image is then rotated ∆θ degrees about the vertical axis and the process is repeated until each profile signature is created. In our experiments, we chose ∆θ = 5 as the rotation increment. We have experimented with larger and smaller increments and found that if the model is rotated more than 15° at a time, small noses may be mislabeled, and if it is less than 5°, the execution process takes longer with the same result. Abbreviated examples of the rotation process are depicted in Figures 4, 5, and 6. The top images show the 3D shape images and the bottom images show the corresponding profile images at each rotation increment.

Once the profile signatures Sθ have been created, each profile can then be matched to two different model signatures, as seen in Figure 7. These signatures were manually extracted from a single subject image rotated to a known profile position. Model m1 seen in Figure 7(a) is taken from the nose of an image with a 0° pitch and Model m2 seen in Figure 7(b) is the taken from the nose of an image with a 45° pitch. The matching process and nose tip determination is performed as follows:

• For each Sθ, determine the point Bθ,j best matching model m2 such that...
Fig. 4. An example of the rotation process given a subject (05078) with 15° yaw and 45° pitch. The top images show the 3D models with the extracted profile outlined in red and the bottom images show the associated extracted profiles.

Fig. 5. An example of the rotation process given a subject (05078) with 15° yaw and 0° pitch. The top images show the 3D models with the extracted profile outlined in red and the bottom images show the associated extracted profiles.

Fig. 6. An example of the rotation process given a subject (05078) with 15° yaw and -45° pitch. The top images show the 3D models with the extracted profile outlined in red and the bottom images show the associated extracted profiles.
Fig. 7. Model \( m_1 \) is extracted from an image with a 0° pitch, and model \( m_2 \) is extracted from an image with a 45° pitch.

\[
B_{\theta,z} = \arg \min_i \sum_{j=1}^{m_z} (S_{\theta,i+j} - m_{z,j}).
\]

(1)

• For each \( \theta \), if

\[
|B_{\theta,m_1} - B_{\theta,m_2}| < \Delta,
\]

where \( \Delta \) was experimentally determined to be 5mm, then the found nose tip is

\[
f_{\theta} = \frac{B_{\theta,m_1} + B_{\theta,m_2}}{2}.
\]

(3)

• Otherwise, \( f_{\theta} \) is either \( B_{\theta,m_1} \) or \( B_{\theta,m_2} \). We find \( \beta_{\theta,m_1} \) and \( \beta_{\theta,m_2} \), the next local minima in \( S_{\theta} \) below \( B_{\theta,m_1} \) and \( B_{\theta,m_2} \) respectively. \( f_{\theta} \) selects the point for which the distance between the point and the corresponding local minimum,

\[
\sqrt{(B_{\theta,m_2}.x - \beta_{\theta,m_2}.x)^2 + (B_{\theta,m_2}.y - \beta_{\theta,m_2}.y)^2}
\]

is minimized. We perform matching in this manner rather than using correlation because of the structure and simplicity of the problem. We assume that the \( y \)-coordinates for both the model and the probe are scaled appropriately during the initialization phase and therefore do not wish to incur the additional processing time required for calculating the correlation values.

An example of this process can be seen in Figure 8. The images show model \( m_1 \) repeatedly matched to selected configurations starting at the top of the face and moving down to the bottom. Figure 8(c) shows the configuration resulting in the minimum matching score and is reported as the nose tip. Each distinct value of \( \theta \) yields a nose tip matching score \( N_{\theta} \). The vector of \( N_{\theta} \) exhibits a global minimum at the correct rotation \( \theta^* \). However, noise can affect the smoothness of the entries in the vector. We select \( \theta^* \) by identifying a run of 4 \( \theta \) values whose sum of scores is minimal, and select \( \theta^* \) as the \( \theta \) yielding the minimum sum matching scores. The original yaw orientation of the image can be calculated by subtracting the known profile yaw (90°) from the resulting \( \theta^* \) value.

An example of the RPS algorithm applied to the image in Figure 5 (which contained a 15° rotation to the left) is as follows. Figure 9 shows that \( r = 105^\circ \). This suggests that in order to rotate the input image to the optimal profile configuration (90° to the right), a 105° rotation is performed on the input image. This is equivalent to the original image being rotated 105-90 = 15° to the left.

In order to determine how well the RPS algorithm locates the nose tip, each of the 7,317 images was first manually annotated for “ground truth” nose points. Upon completion of the RPS process, the automatically detected nose point is produced. The distance between the manually annotated nose tip and the RPS detected nose tip is calculated. If it is less than 10mm, it is considered a match.

The benefit of using the RPS method for nose detection is its indifference to any amount of model yaw, provided that the nose is visible. The results seen in Table II confirm this statement and show the detection rate at each pose in
the NDOff2007 data set where the numbers along the top represent changes in yaw and the numbers along the side represent changes in pitch. With a 0° pitch, each pose reports detection results of almost 100% regardless of 90° to -90° changes in yaw. The high level of detection is consistent for each pose variation except (60°, 60°). We believe that this irregularity is due to the limited amount of 3D point information available on the surfaces of the faces in that extreme pose. Examples of selected misidentified images for this pose set are seen in Figure 10. Due to the functionality of the laser scanner, facial points that are obscured due to pose do not produce valid 3D points, which leaves large portions of the face missing in the resulting data. The combination of large changes in pitch and yaw seen here greatly reduce the accuracy of the RPS algorithm.

V. CONCLUSIONS AND FUTURE WORK

In this paper we have demonstrated an accurate and efficient method for 3D nose detection and pose categorization. On a dual processor 2.4 Ghz Xeon machine after image preprocessing, nose detection can be performed in under one second per image. This processing time can be decreased even further if a multi-resolution approach is employed. The experiments in this paper are performed on the largest data set to date with various degrees of pose variation and consisting of over 7,300 3D face images.

VI. ACKNOWLEDGEMENTS

Biometrics research at the University of Notre Dame is supported by the National Science Foundation under grant CNS01-30839, by the Central Intelligence Agency, by the Department of Justice / National Institute for Justice under grant 2006-IJ-CX-K041, and by UNISYS Corp.

REFERENCES

<table>
<thead>
<tr>
<th>Pitch</th>
<th>90</th>
<th>60</th>
<th>45</th>
<th>30</th>
<th>15</th>
<th>0</th>
<th>-15</th>
<th>-30</th>
<th>-45</th>
<th>-60</th>
<th>-90</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>60</td>
<td>90</td>
<td>73.7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>45</td>
<td>100</td>
<td>99.3</td>
<td>100</td>
<td>98.8</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>30</td>
<td>99.3</td>
<td>99.4</td>
<td>98.4</td>
<td>100</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>99.4</td>
<td>98.3</td>
<td>98.4</td>
<td>100</td>
<td>99.4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>99.6</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>99.3</td>
<td>99.4</td>
<td>99.3</td>
</tr>
<tr>
<td>-15</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>99.4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>-30</td>
<td>98.7</td>
<td>100</td>
<td>98.2</td>
<td>98.8</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-45</td>
<td>97.8</td>
<td>98.2</td>
<td>95.7</td>
<td>96.9</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>


