Multiscale Integral Invariants For Facial Landmark Detection in 2.5D Data

Adam Slater, Yu Hen Hu*
Department of Electrical & Computer Engineering
University of Wisconsin - Madison
Madison, WI, USA
{ajslater, yhhu}@wisc.edu

Nigel Boston*
Department of Mathematics & ECE
University of Wisconsin - Madison
Madison, WI, USA
boston@math.wisc.edu

Abstract— **In this paper, we introduce a novel 3D surface landmark detection method using a 3D integral invariant feature extended from that proposed by Manay et al. for 2D contours. We apply this new feature to detect the nose tips of 2.5D range images of human faces. Using the Face Recognition Grand Challenge 2.0 dataset, our method compares favorably with a recently proposed competing method.** *

I. Introduction

The area of face recognition is a well-researched field, with many thoroughly studied approaches and algorithms. However, the vast majority of this research has been in the area of two-dimensional images. With the relatively recent availability of three-dimensional (3D) face databases and scanning equipment, 3D face recognition has begun to receive much more attention. While there have been a number of proposals in recent years [19] for various systems for face recognition using this newly-available three-dimensional data, very few papers have focused on the initial registration and landmark detection stage intrinsic to the operation of most of these algorithms. This paper seeks to explicitly address this particular aspect of 3D face recognition and presents a novel algorithm for the detection of 3D facial landmark points and the registration of 3D facial range images.

This area of facial feature detection and location has been receiving increased attention lately, with several differing approaches proposed. One of the earliest and most common methods has been the use of local mean and Gaussian curvature information [2,3,4,5,6]. These approaches either use the curvature to segment the face into regions for recognition or attempt to use local curvature to locate feature points for registration. Unfortunately, methods based on local curvature information can be unstable due to noisy data and are fairly prone to giving false positives.

A more recent method used by Medioni et al. [7] used the ICP (Iterative Closest Point) algorithm for registration. This is a general purpose algorithm for the iterative minimization of mean square error when aligning arbitrary point clouds. This method, while guaranteed to converge, may converge to local minima and is dependent on the assumption that one of the datasets to be aligned is a subset of the other and both are free from outliers, in the sense that that each point in at least one of the datasets has a valid corresponding point in the other [17].

The Point Signature[1,8,9] method proposed by Chua and Jarvis proposed a method of extracting local features by projecting local points onto a plane orthogonal to a local approximation of the surface normal vector. This technique was extended by the Spin Image representation [10,11], which modified the original technique to include invariance to rotations and translations. A further modification of this technique was proposed by Wang et al. [15,16], whose Local Shape Map scheme mitigated problems with data loss and ambiguousness in the Spin Map calculation. All of these methods, however, rely on the accurate approximation of surface normals, which are sensitive to expression variation and noise, and all suffer from a relatively poor tradeoff between storage required and performance.

A somewhat similar method, proposed by Xu et al. [12] is used as a basis for comparison with the algorithm proposed in this paper. This work relies on calculations of the angle of intersection between point normals and vectors to points in a local area. The method reduces the storage requirements of the previously discussed algorithms by only retaining the mean and variance of each set of intersection angles and by using an early filtration method to reduce its number of candidate points.

In this paper, we propose a 3D integral invariant feature for 3D surface and apply it to detect the nose tip landmark for a given range image of human face. Our method is inspired by the work of integral invariant signatures by Manay et al. [13], which approximate contour features based on integration, rather than differentiation, of local areas. Our approach is also influenced by Mortara et al.'s technique of tracing the contours of intersection of 3D spheres, similar to a Gaussian curvature calculation [14].

Most prior methods of 3D landmark detection or local feature extraction have been very sensitive to small-scale variations in the data which can be caused by expression variation or noisy data acquisition and suffer from a poor storage to performance tradeoff. Our method aims to rectify these difficulties while retaining the discriminative power of local feature methods like the point signature or spin image.

Specific contributions of this work include:

- Formulation of an integral invariant feature over 3D surfaces
- Development of an efficient, incremental feature extraction method of the proposed 3D integral invariant.

* Yu Hen Hu and Nigel Boston are supported by the National Science Foundation under Grant No. CCF-0434355

- An empirical feature dimension selection method using linear discriminant analysis.
- A hierarchical multi-modal 3D surface landmark detection method for locating nose tip using both 3D range image and corresponding 2D color image.

In the rest of this paper, the proposed 3D surface integral invariant will be presented in section II. An efficient incremental feature extraction method is discussed in section III. Multi-modal, multi-stage pattern classification for nose tip landmark detection is presented in section IV.

II. Integral Invariants

Our method is inspired by the work of Manay et al.[13], specifically. However, whereas this prior work focused on one-dimensional contours in $\mathbf{R}^2$, we extend this method to two-dimensional surfaces in $\mathbf{R}^3$.

The method approximates the value of the integral invariant signature

$$I_\gamma(p) = \int_{\bar{\gamma}} h(p,x)\,d\mu(x)$$

where p is a point of interest, x is a point in the local neighborhood of p, and $d\mu(x)$ represents an infinitesimal geodesic distance on the surface. γ represents the surface itself, and $\bar{\gamma}$ represents the volume enclosed by the surface. $h(p,x)$ is a kernel function which satisfies

$$\int_{\gamma} h(p,x)\,d\mu(x) = \int_{g\bar{\gamma}} h(gp,x)\,d\mu(x) \quad \forall g \in G$$

where G is a group and gp is the image of p under the group action of g in G. It has been shown that any function $I_\gamma(p)$ which satisfies this relationship is invariant with respect to the group G.

In this work, we begin with the three-dimensional analog to the special Euclidean invariant kernel proposed in [13],

$$h(p,x) = \chi(B_r(p) \cap \bar{\gamma})(x)$$

which represents the indicator function of the intersection of a sphere $B_r(p)$ of radius r centered at the point p with the volume enclosed by the surface γ. This kernel is invariant with respect to the special Euclidean group, which includes any rigid transformations of the data, such as translations or rotations. This gives rise to the corresponding integral invariant:

$$I_\gamma(p) = \int_{B_r(p)\cap\bar{\gamma}} dx$$

This integral represents the volume of intersection of a sphere of radius r, centered at point p, and volume enclosed by the surface γ. This new kernel also remains invariant under special Euclidean group.

III. Feature Extraction

The single parameter of this invariant is the radius of sphere of intersection, r. To obtain a complete representation of each point's local region, it's essential to perform this calculation for a variety of radii. However, performing this calculation over a scale-space will result in redundant information, since each larger scale also contains the information from each smaller scale. In order to minimize the amount of redundant information contained by the local surface representation, it's necessary to minimize the number of repeated calculations between scale levels. To this end, we redefine our kernel function $h(p,x)$ as:

$$h(p,x) = \chi(B_{r[k]}(p)) \cap \bar{\gamma})(x) - \chi(B_{r[k-1]}(p)) \cap \bar{\gamma})(x)$$

where $r[k-1] < r[k]$. This results in an integral invariant equivalent to the volume of intersection of the surface γ and a spherical shell with interior radius $r[k\text{-}1]$ and exterior radius $r[k]$:

$$I_\gamma(p) = \int_{B_{r[k]}(p)\cap\bar{\gamma}} dx - \int_{B_{r[k-1]}(p)\cap\bar{\gamma}} dx$$

This function is invariant under the special Euclidean group. As a discrete approximation to this integral, we used the middle Riemann sum of the intersection volume. For each point (x_p, y_p, z_p), we approximated this integral as:

$$\begin{aligned} I_\gamma(x_p, y_p, z_p, r[k]) \approx &\sum_{x,y,z} f(z - z_p, S(x - x_p, y - y_p, r[k])) - \\ &\sum_{x,y,z} f(z - z_p, S(x - x_p, y - y_p, r[k-1])) \end{aligned}$$

$$f(z_1, z_2) = \begin{cases} 2Az_2 & z_1 > z_2 \\ A(z_1 - z_2) & -z_2 < z_1 \le z_2 \\ 0 & z_1 \le -z_2 \end{cases}$$

$$S(x,y,r) = \begin{cases} r^2 - x^2 - y^2 & x^2 + y^2 \le r^2 \\ 0 & x^2 + y^2 > r^2 \end{cases}$$

where A is the area of each point's local region. In our experiments, the data was sampled uniformly over the x-y plane, so A was constant. Although better approximations are possible and an area for future research, this was sufficient for our implementation.

The local shape of a 3D surface can be represented by a feature vector that consists of I_γ for increasing values of $r[k]$. The extent of k determines the dimension of the feature vector. Obviously, the feature dimension is dependent on particular type of 3D surfaces to be represented. Hence, it is appropriate to determine the feature dimension empirically based on the avaiable 3D surface data.

Specifically, a subset of 100 loosely registered face scans taken from face recognition grand challenge V.2.0 data set [18] are used as the training data. For each face scan, we choose a subsample of the facial surface points including the nose tip to compute the 3D features. At each point, we calculate I_γ for $r[k]$ varying from *4mm* to *100mm* at *2mm* increments. We assume the feature vectors corresponding to the nose tip and those not at the nose tip will form two K dimensional probability distributions where K is the feature dimension.

We then compute the Mahalanobis distance similar to that used in Fisher's linear discriminant analysis between these two distributions for each feature dimension K:

$$J = (m_1 - m_2)^T S^{-1} (m_1 - m_2)$$

where m_1 and m_2 respectively are the mean vectors of the nose and non-nose distribution, and S is the covariance matrix corresponding to the non-nose distribution. The larger this distance, the more likely the feature vector is able to discriminate nose tip from other regions on the facial surface. A plot of this distance as a function of the radii used for computing the 3D feature is depicted in Figure 1. In the current implementation, we chose to retain features of radii less than 60mm, because features greater than 60mm appear to give little benefit, as shown in Fig. 1.

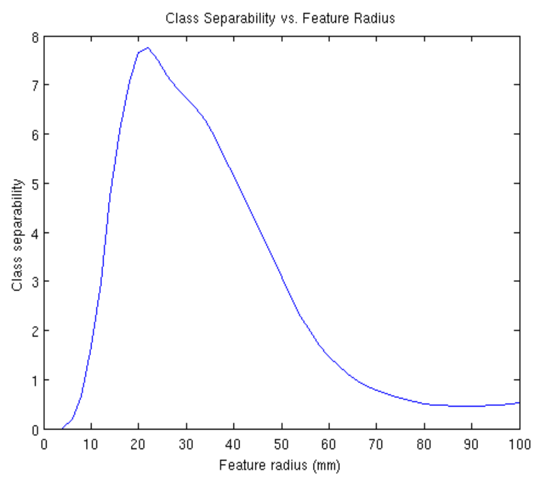

Figure 1: Class separability based on Mahalanobis distance for feature radii from 4mm to 100mm, averaged over a training set of 100 faces

IV. MULTI-MODAL, MULTI-STAGE NOSE TIP DETECTION

Once the feature dimension is determined, a simple quadratic discriminant analysis linear classifier is applied to classify the feature vectors into one of two classes: nose tip versus non-nose tip. While there are more sophisticated classifiers that potentially would yield better performance, due to space limitation, we present only linear classification results in this paper.

Since the FRGC database facial range images also include hair and clothes, initial tests of the nose-tip detection algorithm yielded excessive false positive classifications. Fortunately, for each range image, the database also provides an aligned color 2D image. Therefore, we employ skin color segmentation to restrict the search within the skin-colored facial region. This is accomplished with a simple Bayesian lookup-table based skin color classifier on the chrominance components of the YCbCr color-space. This classifier outputs a map of the probability that each point on our surface is human skin.

Secondly, we used a more complex classifier using other readily recognizable points on the face. In developing this second classifier, we first determined the average uniqueness of each point on several of our training samples using the Mahalanobis distance from the mean of the face as a criterion. This measurement is shown in Fig. 2. It became evident that the medial canthus (inside eye corner) is an easily classified point on the face using our 3-dimensional invariant, due largely to its strongly positive curvature and relative invariance to expression variation. Our second classifier uses the same local feature discussed earlier to detect the medial canthi, then computes the distance from each canthus to each nose candidate point. These distances are then classified using a quadratic discriminant to give another probability map.

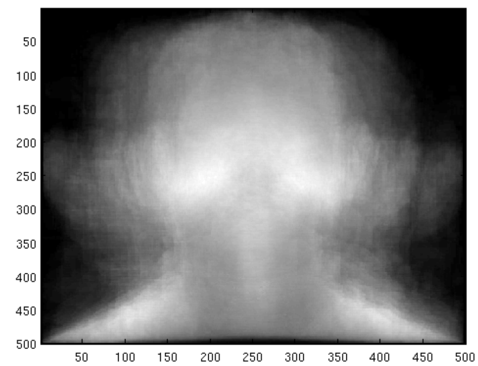

Figure 2: Average Mahalanobis distance of facial points from the mean, evaluated on 100 loosely-registered face scans with feature radii from 4mm to 100mm in increments of 2mm

After computing these three measurements, the integral invariant, the Bayesian color probability map, and the medial canthus distance probability map, the three are multiplied together to give the final conditional probability of each point being a nose tip.

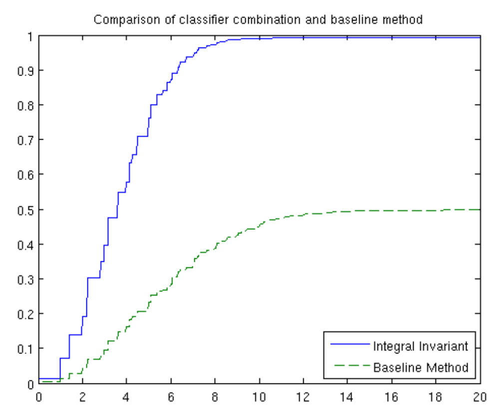

Figure 3: Comparison of our classifier and a baseline algorithm

V. EXPERIMENTAL RESULTS

For our experiments, we used the FRGC 2.0 database of 4007 2.5D range scans and associated 2D color images of human faces. These scans, although all of the head region, are relatively unconstrained, in that many include hair and clothing as well as pose and expression variation. Of these range scans, 101 were used as a training set, and the algorithm was evaluated on the remaining 3906. In this test, we used 10 radii for our algorithm, ranging from 4mm to 60mm. A small amount of preprocessing was performed on the data before the algorithm was applied, consisting of a median filter and removal of very large outlier points, as well as resampling on a 1mm square rectangular grid. For the sake of comparison, we also implemented an earlier algorithm proposed by Xu et al.[12], chosen for its thorough algorithm description. The parameters used for this algorithm were the same given in the paper.

Figure 3 summarizes our experimental results as a cumulative distribution of distances from the manually selected nose tip on each range image. As an illustration of the performance of each method, we examined the percentage of noses which were detected within 1cm of the manually selected position. Using only the integral invariant, 98.08% of these were detected to this tolerance. With skin color segmentation, this improved to 98.52%. Using the canthus distance further improved classification to 99.08%. Using all three classifiers slightly degraded performance to 99.03%, which we believe is due to overfitting. Our baseline method determined 45.3% of the noses to this degree of accuracy. This was due in part to a high number of false positives due to hair and clothing variations and early culling of important feature points, possibly due to the density of our range scans or a small amount of surface noise.

VI. Conclusion & Future Work

This paper presents a novel method for landmark detection in 2.5D facial range scans. This method is based on the Integral Invariants of Manay et al., but extends their algorithm to 3 spatial dimensions and performs classification over multiple scales. This strategy outperforms other common techniques in its resistance to false positives and its low space requirements.

The method presented in this work lends itself very well to some future algorithmic optimizations. First, a cascaded classifier seems a logical direction for this work, and would improve the efficiency of our feature extraction. Also, there are some implementation details that could benefit from future algorithm refinements; for example, at present, our feature extraction process iterates over each pixel in a local area multiple times for each point extracted. This should be possible in only a single pass, and is an area for future improvement.

Another area for improvement is the algorithm used for our discrete approximation to the integral invariant signature. Further extensions to this work could include the investigation of alternative methods for combining the classifier results and choosing feature radii. A final area of future research would be an unconstrained face detection problem on 2.5D data, although we know of no appropriate dataset yet in existence.

VII. References

[1] C. S. Chua and R. Jarvis, "Point Signatures: A New Representation for 3D Object Recognition," *International Journal of Computer Vision*, vol. 25(1), pp. 63 – 85, October 1997

[2] G. Gordon, "Face recognition based on depth maps and surface curvature. Geometric Methods in Computer Vision," *SPIE* vol. 1570, pp. 1–12, July 1991.

[3] J. C. Lee and E. Milios, "Matching range images of human faces," In the Proceedings of Int'l Conf. on Comp. Vision, pp. 722–726, 1990.

[4] T. Nagamine, T. Uemura, and I. Masuda, "3D facial image analysis for human identification. International Conference on Pattern Recognition," *ICPR*, pp. 324–327, 1992.

[5] H. T. Tanaka, M. Ikeda, and H. Chiaki, "Curvature-based face surface recognition using spherical correlation-principal directions for curved object recognition," Third International Conference on Automated Face and Gesture Recognition, pp. 372–377, 1998.

[6] AB Moreno, Á Sánchez, JF Vélez, and FJ Díaz, "Face recognition using 3D surface-extracted descriptors," Irish Machine Vision and Image Processing Conference, September 2003.

[7] G. Medioni and R. Waupotitsch, "Face recognition and modeling in 3D," IEEE Int'l Workshop on Analysis and Modeling of Faces and Gestures, pp. 232-233, October 2003.

[8] Y. Wang, C. Chua, and Y. Ho, "Facial feature detection and face recognition from 2D and 3D images," *Pattern Recognition Letters*, vol 23, pp. 1191–1202, 2002.

[9] C.-S. Chua, F. Han, and Y.-K. Ho, "3-D human face recognition using point signature," in *Proc. 4th IEEE Int. Conf. Automatic Face Gesture Recognition*, pp. 233-238, 2000.

[10] Andrew E. Johnson , Martial Hebert, "Using Spin Images for Efficient Object Recognition in Cluttered 3D Scenes," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 21(5), pp. 433-449, May 1999.

[11] Y. Li and W.A.P. Smith, "Face Recognition using Patch-based Spin Images," *International Conference on Pattern Recognition*, vol. 1(20-24) , pp. 408- 411, Aug. 2006

[12] C. Xu, Y. Wang, T. Tan, and L. Quan, "Robust nose detection in 3D facial data using local characteristics," *International Conference on Image Processing*, vol. 3(24-27), pp. 1995- 1998, Oct. 2004.

[13] S. Manay and B.-W. Hong, "Integral Invariants For Shape Matching.," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28(10), pp. 1602-1618, Oct. 2006.

[14] M. Mortara, Gi. Patané, M. Spagnuolo, B. Falcidieno, J. Rossignac, "Blowing Bubbles for Multi-Scale Analysis and Decomposition of Triangle Meshes," *Algorithmica*, vol. 38(1), pp. 227-248, October 2003.

[15] Y. Wang, G. Pan, Z. Wu, and S. Han, "Sphere-spin-image: A viewpoint-invariant surface representation for 3d face recognition," in *International Conference on Computational Science*, pp. 427–434, 2004.

[16] Z. Wu, Y. Wang, and G. Pan, "3d face recognition using local shape map," *ICIP*, pp. 2003–2006, 2004.

[17] D. Chetverikov, D. Svirko, D. Stepanov, and P. Krsek, "The Trimmed Iterative Closest Point algorithm," in the proceedings of International Conference on Pattern Recognition. vol. 3(11-15), pp. 545-548, Aug. 2002.

[18] P.J. Phillips, P.J. Flynn, T. Scruggs, K.W. Bowyer, K. Jin Chang Hoffman, J. Marques, W. Jaesik Min Worek, "Overview of the face recognition grand challenge," in the proceedings of IEEE Computer Society Conference on Compuer Vision and Pattern Recognition, vol. 1, pp. 947-954, June 2005.

[19] K.W. Bowyer, K. Chang, P. Flynn, "A survey of approaches and challenges in 3D and multi-modal 3D+2D face recognition," *Computer Vision and Image Understanding*, vol. 101(1), pp. 1-15, 2006.