Improved Detection of Landmarks on 3D Human Face Data

Shu Liang[1], **Jia Wu**[2], **Seth M. Weinberg**[3], and **Linda G. Shapiro**[2]

[1]Department of Electrical Engineering, Tsinghua University, P.R. China

[2]Department of Computer Science and Engineering, University of Washington, Seattle, Washington

[3]Department of Oral Biology, Center for Craniofacial and Dental Genetics, University of Pittsburgh, Pittsburgh, Pennsylvania

Abstract

Craniofacial researchers make heavy use of established facial landmarks in their morphometric analyses. For studies on very large facial image datasets, the standard approach of manual landmarking is very labor intensive. With the goal of producing 20 established landmarks, we have developed a geometric methodology that can automatically locate 10 established landmark points and 7 other supporting points on human 3D facial scans. Then, to improve accuracy and produce all 20 landmarks, a deformable matching procedure establishes a dense correspondence from a template 3D mesh with a full set of 20 landmarks to each individual 3D mesh. The 17 geometrically-determined points on the individual 3D mesh are used for the initial correspondence required by the deformable matching. The method is evaluated on 115 3D facial meshes of normal adults, and results are compared to landmarks manually identified by medical experts. Our results show a marked improvement to prior results in the recent literature.

I. INTRODUCTION

Using 3D human face data to measure facial features is of great practical importance in craniofacial research and practice. Traditionally, direct anthropometry using calipers has been the standard technique for quantifying craniofacial dysmorphology, as well as for surgical planning and outcome assessment [1]. Some of the major downsides to direct anthropometry include the excessive time and invasiveness of the method, the amount of training required, the extent of measurement error, and limitations in the kinds of data that can be collected. Following the introduction of cost-effective 3D surface imaging solutions, computerized anthropometry has largely replaced more traditional direct methods for collecting quantitative information on human faces [2]. These systems are capable of capturing the full 3D geometry of the human face and head in just a fraction of a second.

While computerized 3D anthropometry represents a major advance, to obtain measurements, points on the face and head corresponding to traditional anthropometric landmarks must still be captured manually through the use of software. This can be a time consuming process, requiring a fair amount of training to master. Efficiency is particularly crucial when dealing with very large 3D facial datasets, such as the repositories currently being collected as part of the FaceBase Consortium [3]. Recognizing the need to move beyond manual data collection, our goal is to provide an automatic method to detect landmarks from 3D facial surfaces. Such a method requires the resulting automatically-generated landmarks to be located in the correct anatomical positions and the process to be extendable to as many landmarks as needed.

In the computer vision community, facial landmark detection methods can be classified as those that are solely dependent on geometric information and those that are supported by trained statistical feature models [4]. Of the methods that are dependent on geometric information, surface curvatures, shape index and geometric relationships are heavily used. For example, Lu and Jain [5] and Colbry *et al.* [6] find eye/mouth corners and nose/chin tips based on a fusion scheme of shape index on range maps and the "cornerness response," making use of distance relationships to find the points. Lin *et al.* [7] used curvature analysis to determine the eye sockets and detected the nose tip as the extreme vertex along the normal direction of the eye sockets.

Using not only geometric information but also trained models, Yu and Moon [8] located the nose tip and inner eye corners in 3D range maps with a genetic-algorithm-trained detector. Xu *et al.* [9] used the concept of "effective energy" to describe the relationships between neighboring points and an SVM classifier to select the correct nose tip points. Nair and Cavallaro [10] used a point distribution model to locate landmarks and register faces. Romero-Huertas and Pears [11] developed a graph-matching approach to locate the positions of the nose tip and inner eye corners. All of the above methods are limited to a fixed number of predefined points. If any new landmarks are needed, a new type of model must be created.

In this paper, we describe an improved method to automatically locate an arbitrary number of landmarks on a 3D facial mesh. First, a partial set of landmarks is located on each individual mesh by geometric techniques. Then, in order to improve positional accuracy and add additional landmarks, the geometrically-detected landmarks are used to initialize a deformable transformation that is used to create a dense correspondence between a template mesh and the individual (target) mesh, so the set of established landmarks on the template mesh can be transferred to the target. The distance between manually annotated landmarks labeled by experts in anthropometry and the automatically-detected final landmarks is computed to evaluate the accuracy of the method.

The main contribution of this work is the fully working system that can produce all 20 established landmarks and outperforms all competing methods in the literature. A second contribution is the geometric methodology for finding the initial set of 17 landmark points that are used to intialize the deformable matching procedure and are critical to its success. The use of deformable matching for this purpose is not a new idea, but the specific matching algorithm used [12], which has not been used for this purpose before, is particularly good for this task, because it executes an order of magnitude faster than other algorithms we tested (e.g. about 10 minutes for the process in [12] compared to more than 4 hours for the method in [13]).

II. DATA FORMAT

Our data set consists of 115 3D facial meshes comprised of 30-40,000 points obtained from a 3dMD® digital stereophotogrammetry imaging system (Atlanta, GA). These systems are outfitted with multiple CCD cameras mounted at fixed angles and distances, to capture overlapping views of the face and head. Prior to 3D image capture, scalp hair obscuring the subject's face was cleared away through application of hairnets, hairbands, and various types of pins and clips. Subjects were positioned so that their heads were centered between the imaging pods and positioned slightly upward in order to ensure adequate coverage of the subnasal region. Extraneous data, including most of the subject's neck and shoulders, were removed. Surface meshes were pose-normalized for our automatic detection system using a method described in [14].

There are 27 landmarks mentioned in this paper, including the nasion (n), sellion (se), pronasale (prn), subnasale (sn), right and left alare (al), right and left alar curvature (ac), labiale superius (ls), stomion (sto), labiale inferius (li), sublabiale (sl), right and left subalare (sbal), right and left crista philtri (cph), right and left chelion (ch), ganthion (gn), right and left endocanthion (en), right and left exocanthion (ex), right and left superaurale (sa), right and left postaurale (pa), as shown in Fig. 1. The 20 established facial landmarks that we sought are listed in Table 1 and also marked with a triangle in Fig. 1. For ground truth, these 20 landmarks were located manually on each surface by a single trained personnel. All landmarking was performed on 3D models with color and texture mapping active, which made them more accurate than if placed on the mesh alone. However, our programs only had access to 3D meshes with no color/texture.

III. METHOD

Our method to find landmarks using deformable registration can be described in the following steps:

1. Compute 17 initial landmarks on the nose, mouth, eyes, and ears of both the template mesh T and the target mesh D using fully automatic geometric methods.

2. Use deformable registration with the geometric landmarks from step 1 for initialization to find a dense correspondence from T to D.

3. Transfer the 20 manually-marked established landmarks from T to the corresponding points of D to obtain final landmarks.

A. Generating Initial Landmarks

The geometric method for automatic detection of initial landmarks requires that the head is normalized to face forward, the origin point lies in the center of the head, and the data are in (x, y, z) coordinates as shown in Fig. 1. The face width is defined along the x-axis, the height along the y-axis, and the depth along the z-axis. Our geometric methodology can find 17 landmarks automatically, including seven nose points, four eye points, two mouth points and four ear points, marked with an asterisk in Fig. 1. Of the 17, ten are established landmark points and seven more (the sellion, the right and left alar curvature, right and left superaurale, right and left postaurale) are support points that help the deformable matching procedure to find a better correspondence from the template to the target.

1) Nose Landmarks—For each 3D head, there is a set of points Z_{max} at the maximum z-value. The geometric center of these points (x_{prn}, y_{prn}) is the pronasale (prn). The sellion (se) and subnasale (sn) can be found as the local minima on either side of the pronasale on the line with the same x-value as the pronasale, as shown in Fig. 2(a). (Note that the nasion, an established landmark, is easy to find on skull CT data, but impossible to localize on 3D mesh data, so we find the sellion instead and leave it to the deformable matching to estimate the nasion.) To find the left and right alare and alar curvature, the region is restricted to $y_{sn} < y < y_{prn}$. The left and right alare (al) are located by calculating the surface normal angles of the points. Given the surface normal vector $n(n_x, n_y, n_z)$ of a point in the region, the point with the largest n_x value has a surface normal pointing toward the left side of the face. This point is selected as al_l, and the point with the smallest n_x value is selected as al_r (Fig. 2(c)).

Given the surface normal vectors of two neighboring triangles, the angle between these two vectors is a dihedral angle. All the points on the edges with dihedral angles larger than 30° make up the set of sharp edge points S shown in Fig. 2(b).

Two more nose points, the left and right alar curvature (ac), are defined by equations (1) and (2). The ac_l point is selected as the leftmost point (with the largest x-value) of the region, and the ac_r is the rightmost point (with the smallest x-value) as shown in Fig. 2(c).

$$ac\_l = \left\{ (x,y,z) \,\middle|\, \arg\max_{x,y,z} x, \, y_{sn} < y < y_{prn}, (x,y,z) \in S \right\} \quad (1)$$

$$ac\_r = \left\{ (x,y,z) \,\middle|\, \arg\min_{x,y,z} x, \, y_{sn} < y < y_{prn}, (x,y,z) \in S \right\} \quad (2)$$

Both points are selected from sharp edge point set S, lower than the pronasale and higher than the subnasale.

2) Mouth, Ear and Eye Points—As shown in Fig. 3(a), to locate the left and right chelion (ch) points, the mouth is extracted by first restricting to the region where $y < y_{sn}$. Among the points on sharp edges in this region, we select the point with the largest x-value as ch_l and the point with the smallest x-value as ch_r, as defined in (3) and (4).

$$ch\_l = \left\{ (x,y,z) \,\middle|\, \arg\max_{x,y,z} x, \, y < y_{sn}, (x,y,z) \in S \right\} \quad (3)$$

$$ch\_r = \left\{ (x,y,z) \,\middle|\, \arg\min_{x,y,z} x, \, y < y_{sn}, (x,y,z) \in S \right\} \quad (4)$$

Two points on each ear are used to help define the eye position and registration: the point with the largest y value as superaurale (sa) and the point with the smallest z value as the postaurale (pa) on the outline of the ear (Fig. 3(b)).

To locate the endocanthion (en) and exocanthion (ex) points on the left eye, the region E_l is restricted to $y_{pa} < y < y_{se}$ along the y-axis and $x_{se} < x < 0.8x_{pa}$ along the x-axis. Then ex_l and en_l can be found as those points with the least Euclidean distance to the corner with largest x-value and smallest y-value (the right-back-most point) and smallest x- and z-values (the left-back-most point), respectively, among all the candidates on the sharp edges. The ex_r and en_r points can be detected similarly as shown in Fig. 3(c). All the geometric restrictions are based on the relationship of points to already-detected landmarks. Thus the pose normalization step [14] is critical to the success of the geometric methods. For most individuals, including adults (Fig. 4(a)) and children (Fig. 4(b)), the geometric method produces good initial results. In some cases, such as in Fig. 4(c), where the subnasale is mistakenly located because of the mustache in the upper lip area, initial results can be improved by the following deformation step.

B. Deformable Registration

The purpose of this step is to improve the accuracy of the initial landmarks and to add additional landmarks that cannot be easily detected on meshes. Given a template 3D mesh T and a target 3D mesh D with sets of geometrically-generated initial landmark points I_T and I_D, respectively (Fig 5(a) and (b)), deformable registration is a process that can be applied to find a dense correspondence from mesh T to mesh D. Due to its speed, we used the deformable registration method of Allen *et al.* [12], initialized by the correspondences between the points of I_T and the corresponding points of I_D. After completion, all landmarks

marked by an expert on the template mesh T are transferred to the corresponding points on the target mesh D, and these become our final landmarks. Note that the number of final landmarks is determined by the number of the landmarks on the template mesh, and the final landmarks need not have any relation to the initial landmarks.

The deformable registration algorithm uses an optimization framework to find a set of transformations that move all the points in T to a deformed surface that matches well with D, minimizing an energy function

$$E = \alpha E_d + \beta E_s + \gamma E_m \quad (5)$$

whose terms include the data error E_d (how well the deformed template matches the target), the smoothness error E_s (of the deformation) and the marker error E_m (how well the initial landmarks match). The process is iterative. In early iterations, the marker error contributes more to the global optimization. As the process moves towards the end of the registration, the data error dominates the transformation. See [12] for details.

Only one template mesh was used in the experiments reported here. We did try several different template meshes with similar results. In prior work, we have developed systems that selected the *best matching* template mesh using a fast heuristic matching algorithm for this purpose [13].

IV. RESULTS

In order to validate our automatic landmark results, we compared the landmark positions with the landmarks manually placed by the experts. The average distances from our automatic landmarks to the ground truth given by medical experts are shown in Table I with their standard deviations. Recall that the automatic landmarks are computed based on the shape information of the heads without textures, while the experts used the color/texture data.

The average distance of the 10 points generated by the geometric method alone to the expert points is 3.12 mm. After the deformable registration with one template to all targets, the average distance of these 10 landmarks reduces to 2.74mm, with a smaller standard deviation for eight out of the 10 points (though a few points increase in average error). The final results with all 20 points have an average distance of 2.64 mm to the ground truth. Results in the literature on similar data are much worse with average distances ranging from 5 to 13 mm, as shown in Table II. Note that none of the other methods attempted to find the full set of 20 landmarks.

V. CONCLUSIONS

In this paper, an improved method was introduced to automatically detect landmarks from 3D human facial data. First, geometric information was used to locate 17 prominent points. Then a deformable transformation between target mesh and data mesh determined 20 established landmarks and located them more accurately than with the geometric method alone. Our method has an average error of 2.64 mm over a sample of 115 heads and is superior to prior published methods. Further studies will include testing our method on a larger database, using these landmarks in morphometric studies and comparing landmark-based analysis against shape-based analysis of human faces.

Acknowledgments

REFERENCES

1. Wong JY, Oh AK, Ohta E, Hunt AT, Rogers GF, Mulliken JB, Deutsch CK. Validity and reliability of craniofacial anthropo-metric measurement of 3d digital photogrammetric images. The Cleft Palate-Craniofacial Journal. 2008; 45(3):232–239. [PubMed: 18452351]

2. Heike CL, Cunningham ML, Hing AV, Stuhaug E, Starr JR. Picture perfect? reliability of craniofacial anthropometry using three-dimensional digital stereophotogrammetry. Plastic and reconstructive surgery. 2009; 124(4):1261. [PubMed: 19935311]

3. Hochheiser H, Aronow BJ, Artinger K, Beaty TH, Brinkley JF, Chai Y, Clouthier D, Cunningham ML, et al. The facebase consortium: a comprehensive program to facilitate craniofacial research. Developmental biology. 2011; 355(2):175–182. [PubMed: 21458441]

4. Perakis, P.; Passalis, G.; Theoharis, T.; Kakadiaris, IA. Tech. Rep. University of Athens; Greece: 2010. 3d facial landark detection and face registration.

5. Lu, X.; Jain, AK. IEEE International Conference on Automatic Face and Gesture Recognition. IEEE; 2006. Automatic feature extraction for multiview 3d face recognition; p. 585-590.

6. Colbry D, Stockman G, Jain A. Detection of anchor points for 3d face verification. IEEE Workshop on Advanced 3D Imaging for Safety and Security. 2005; 3:118.

7. Lin, TH.; Shih, WP.; Chen, WC.; Ho, WY. Proceedings of the 44th annual Southeast regional conference. ACM; 2006. 3d face authentication by mutual coupled 3d and 2d feature extraction; p. 423-427.

8. Yu, TH.; Moon, YS. IEEE International Conference on Biometrics: Theory, Applications and Systems. IEEE; 2008. A novel genetic algorithm for 3d facial landmark localization; p. 1-6.

9. Xu C, Tan T, Wang Y, Quan L. Combining local features for robust nose location in 3d facial data. Pattern Recognition Letters. 2006; 27(13):1487–1494.

10. Nair P, Cavallaro A. 3-d face detection, landmark localization, and registration using a point distribution model. IEEE Transactions on Multimedia. 2009; 11(4):611–623.

11. Romero-Huertas, M.; Pears, N. 2nd IEEE International Conference on Biometrics: Theory, Applications and Systems. IEEE; 2008. 3d facial landmark localisation by matching simple descriptors; p. 1-6.

12. Allen, B.; Curless, B.; Popovi, Z. ACM Transactions on Graphics (TOG). Vol. 22. ACM; 2003. The space of human body shapes: reconstruction and parameterization from range scans; p. 587-594.

13. Teng, CC.; Austin-Seymour, MM.; Barker, J.; Kalet, IJ.; Shapiro, LG.; Whipple, M. Proceedings of the AMIA Symposium. American Medical Informatics Association; 2002. Head and neck lymph node region delineation with 3-d ct image registration.; p. 767

14. Wilamowska, K.; Shapiro, L.; Heike, C. IEEE International Symposium on Biomedical Imaging: From Nano to Macro. IEEE; 2009. Classification of 3D face shape in 22q11. 2 deletion syndrome; p. 534-537.
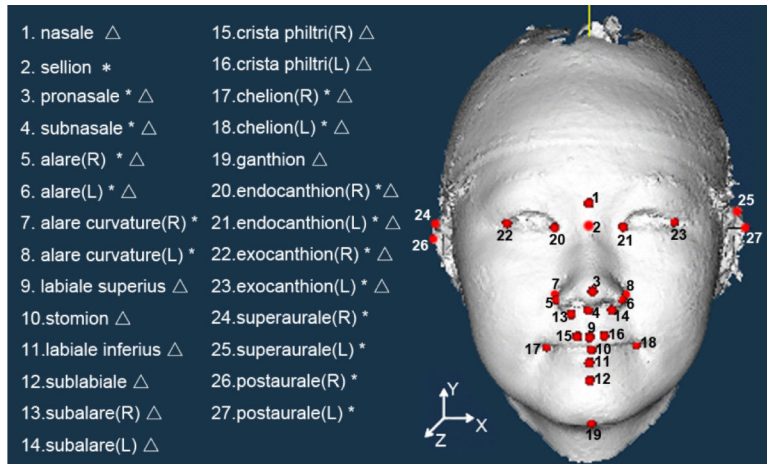
Fig. 1.
Normalized head with landmarks. The geometric method finds the points marked with *.
The points marked with   are found using the deformation method and compared with the
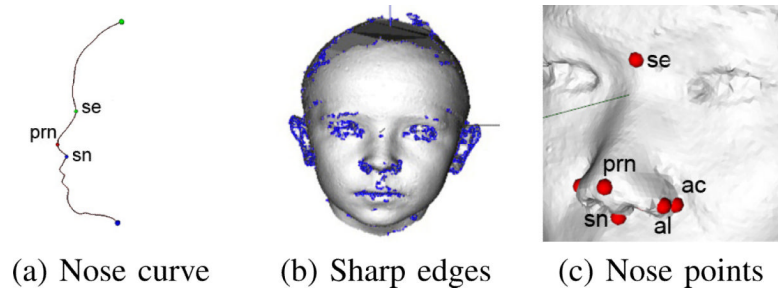ground truth.

(a) Nose curve   (b) Sharp edges   (c) Nose points

Fig. 2.
Initial landmarks on nose
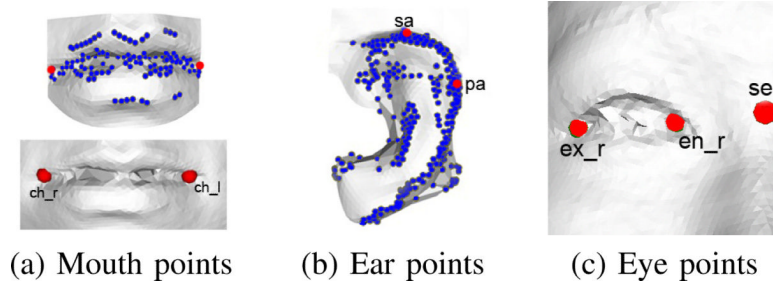
(a) Mouth points (b) Ear points (c) Eye points

Fig. 3.
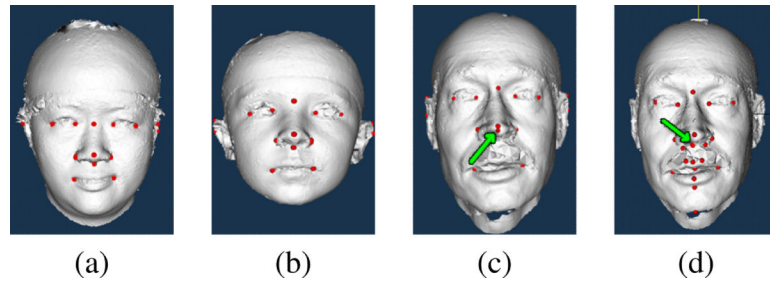Initial landmarks on mouth, ears, and eyes

Fig. 4.
Initial landmarks generated by the geometric method. (a)(b) show good results while (c) is a poor result which is improved by deformable registration. (d) is the result after deformable registration for (c), the subnasale (on nose tip in (c)) is largely improved (below nose in (d)).
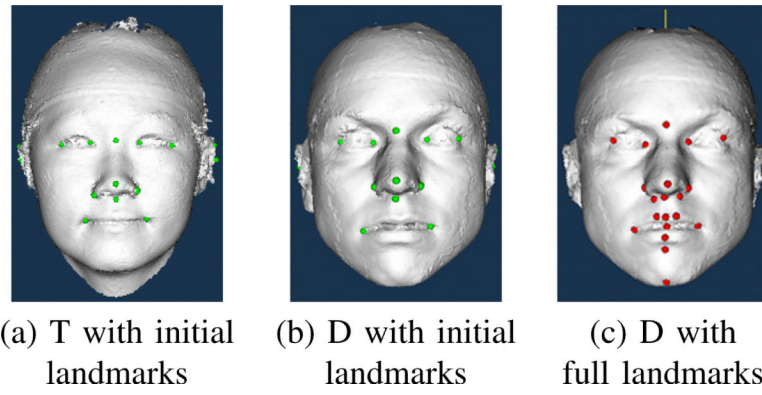
(a) T with initial
landmarks

(b) D with initial
landmarks

(c) D with
full landmarks

Fig. 5.
The deformable registration process: (a) and (b) show T (target) and D (destination) with initial landmarks I_T and I_D respectively. (c) shows D with transferred landmarks, which is the final result of our method.

TABLE I

Average distances and the standard deviations of automatically generated landmarks compared with the ground truth given by medical experts

Point Name	Geometric Method Average Distance(mm)	Deformable Method Average Distance(mm)
Nasion	—	2.92±1.62
Pronasale	1.29±0.68	1.59±0.81
Subnasale	2.35±2.16	2.45±0.80
Alare(R)	3.24±2.61	1.78±1.15
Alare(L)	3.14±2.41	3.07±1.15
Labiale Superius	—	2.27±1.15
stomion	—	1.49±0.90
Labiale Inferius	—	2.27±1.41
Sublabiale	—	3.17±1.87
Subalare(R)	—	2.36±1.06
Subalare(L)	—	1.59±0.93
Crista Philtri(R)	—	2.31±1.27
Crista Philtri(L)	—	1.99±1.03
Chelion(R)	3.14±2.41	3.08±2.14
Chelion(L)	2.80±2.38	3.08±1.64
Gnathion	—	5.31±3.54
Endocanthion(R)	4.78±1.45	2.39±1.09
Endocanthion(L)	4.58±1.70	2.78±1.50
Exocanthion(R)	3.15±2.21	3.34±1.63
Exocanthion(L)	2.72±1.86	3.68±1.91

TABLE II

Average distances(mm) and the standard deviations of our method and methods in the literature.

Point Name	Our method	Yu [8]	Nair [10]	Lu [5]	Colbry [6]	Perakis [4]
prn	1.6±0.8	2.2±6.8	8.8	8.3±19.4	4.1±5.1	4.9±2.4
ch(R)	3.1±2.1	—	—	6.0±16.9	6.9±8.6	5.6±4.3
ch(L)	3.1±1.6	—	—	6.2±17.9	6.7±9.3	6.4±4.2
gn	5.3±3.5	—	—	—	11.0±7.6	6.0±4.3
en(R)	2.4±1.1	4.7±9.8	12.1	8.3±17.2	5.5±4.9	5.1±2.5
en(L)	2.8±1.5	5.6±16.1	11.9	8.2±17.2	6.3±5.0	5.5±2.6
ex(R)	3.3±1.6	—	20.5	9.5±17.1	—	5.8±3.4
ex(L)	3.7±1.9	—	19.4	10.3±18.1	—	5.7±3.5