Landmark Detection on 3D Face Scans by Facial Model Registration

Tristan Whitmarsh[1], Remco C. Veltkamp[2], Michela Spagnuolo[1]
Simone Marini[1], Frank ter Haar[2]

[1]IMATI-CNR, Genoa, Italy
[2]Dept. Computer Science, Utrecht University, The Netherlands

Abstract. This paper describes a way to locate landmarks on a 3D face scan, by registering a facial model to the scan data, and then mapping the landmark location from the model to the data.

Keywords: Landmark detection, face recognition

1 Introduction

In this paper we describe a first attempt to detect landmarks in 3D face scans using a facial template model. Obtaining such landmarks is a first step towards a face data processing at a semantic level. They can for instance be used for animation, face recognition or face comparison. An advantage of using 3D scans as opposed to photographs, is that more information is present in 3D scans in the form of a surface description. A first step in this process is to register the facial template model to the scanned data. When we fit both the shape and the expression of the facial model, we are able to achieve a more accurate registration of the facial model to the face scan. When we assign landmarks to this facial model and obtain an accurate registration of this model to a face scan, then we can obtain the landmarks in the face scan in a rather trivial manner.

Contribution Firstly, we have refined the existing CANDIDE-3 face model so as to provide more degrees of freedom in the registration process, thereby obtaining a better fit. Secondly, we have adapted the well know ICP registration algorithm so that it is capable of the specific task of facial model registration. Thirdly, by fitting a facial model to the scans, we have developed a new landmark detection method.

2 Related work

Several techniques have been developed to find landmarks on face scans [1] or even whole body scans [9], depending on heuristics with different measures of accuracy. The use of color markers is not sufficiently reliable, due to misalignment between scan and texture. The method in [1] is based on the recognition of regions with similar curvature, but leads to many false positive positions. The method in [9] is based on a semi-automatic template matching. In contrast, we propose a method that detects landmarks by registering a facial model to a face scan. Apart from landmark detection, other applications of facial model fitting are face recognition and categorisation. Also the registration of facial models onto images or video is a current field of research [4]. In this work the registration is used to acquire the orientation or shape of the face in the image or video. With the advancements of 3D range scanning we can now apply similar techniques to 3D face scans. While we use model fitting to locate landmarks, other use located landmarks to fit the model [10], a considerably easier task.

3 Scan Data

The scans for this research were made by the Minolta VI-910X and the Vitronic Viro-3D full human body scanner. We made multiple scans of the same subject with different facial expressions. This will test the robustness of the method to different expressions. The method described in this paper is robust to missing data and usually the scans acquired by only one frontal scan of the face contains enough facial details for the registration. Therefore the acquired data will not be post-processed. Other methods may require additional hole filling or the merging of various scan from different angles to fill in missing data. For some methods investigated in this research a point set for the scan data will suffice. However, for more advanced methods explained later, a triangle mesh has to be available. The scanners we used yield a triangle mesh directly, which suffices for this research.

4 Facial model

Our aim is to automatically register a facial model onto a 3D face scan in such a way that the features will match. By this is meant that for instance the tip of the nose of the facial model is at the same position as the tip of the nose on the face scan. For this, a distance measure is minimized between the facial model and face scan. If possible, the facial model should be transformed so that it resembles the shape of the scanned face. For this purpose several facial models have been examined, but most suffer from the lack of deformability: only width, height and depth can be adjusted, which do not cover the wide variety of facial shapes. The CANDIDE facial models distinguish themselves from the other models by the possibility of changing the shape with variables called action units. The CANDIDE-1 model [2] has 79 vertices, 108 triangles, and 11 action units that define only the expression of the facial model. The action units have a value between 1 and –1. Every action unit has a list of vertices that are moved by the action units and a vector for every vertex in the list which indicates the vector which multiplied by the action unit value is added to the vertex. When adjusting these parameters to better fit the model, the registration will improve. An extension of the CANDIDE-1 named CANDIDE-2 has been developed which includes the tongue, a larger part of the head and the shoulders. It is more detailed with 160 vertices and 238 triangles. Unfortunately, only 6 action units have been included and the large number of tongue and shoulder vertices makes it less useful for our purpose. The CANDIDE-3 facial model is an extension of the CANDIDE-1 model with the main purpose of simplifying animation by incorporating the MPEG-4 Facial Animation Parameters. It has 79 action units of which 14 are shape units and 65 are animation units. The shape units define the shape of the CANDIDE-3 model and the animation units the expression. With the more variation in shape and expression, the CANDIDE-3 model should be able to shape itself to the face scan to acquire an even better registration.

Another model taken into consideration is the highly detailed H-Anim template model of the head with Facial Animation Parameters, incorporated in the MPEG-4 standard. A disadvantage of this model is the more complicated method of shape and expression fitting.

5 Registration

5.1 Initial registration

Our registration method starts finding a reasonable initial registration. We assume that each face scan is oriented upwards faced towards the z-axis, that the tip of the nose has the smallest z-value, and that the scan contains the face only. Our initial registration aligns the model's nose vertex with the minimal z-value (nose) of the scan and resizes the model such that it fits the x, y, or z dimension of the scan. In many practical application our assumptions hold, and in the case they don't, we can apply Principal Component Analysis to normalize the scan's orientation and employ existing methods to find the head, and in particular the nose. Alternatively the method of [11] could be used to find the tip of the nose.

5.2 Iterative Closest Point algorithm

Many variants of the Iterative Closest Point (ICP) algorithm [5] have been developed to register two 3D range scans [6]. We have developed an ICP variant for the specific purpose of facial model registration. The ICP algorithm iteratively finds a set of N pairs of corresponding points on two 3D models P and Q. In our method, P represents the facial model and Q the face scan. After finding a point set $p_i \in P$ and $q_i \in Q$, $i = 1 \ldots N$, we have to find the transformation that minimizes the sum of the squared distances. For N we use the number of vertices in our model P. For these vertices we search for correspondences on Q. The easiest approach is to find the closest point in Q for every control point in P. Other methods have been investigated that incorporate the surface structure of the face scan. For instance the normal of the face scan can be used to find the tangent plane at a vertex. The closest point on the tangent plane with the control point p_i can be selected as the corresponding point q_i. Furthermore, we used a distance threshold to reject certain point pairs. The distance used for this decision should not be the distance between the control point pair, but between the point and it's closest point on the scan. It might for instance be possible that the control point is very far away from the scan whereas the closest point on the tangent plane in very close by. With this restriction in distance, a registration can be

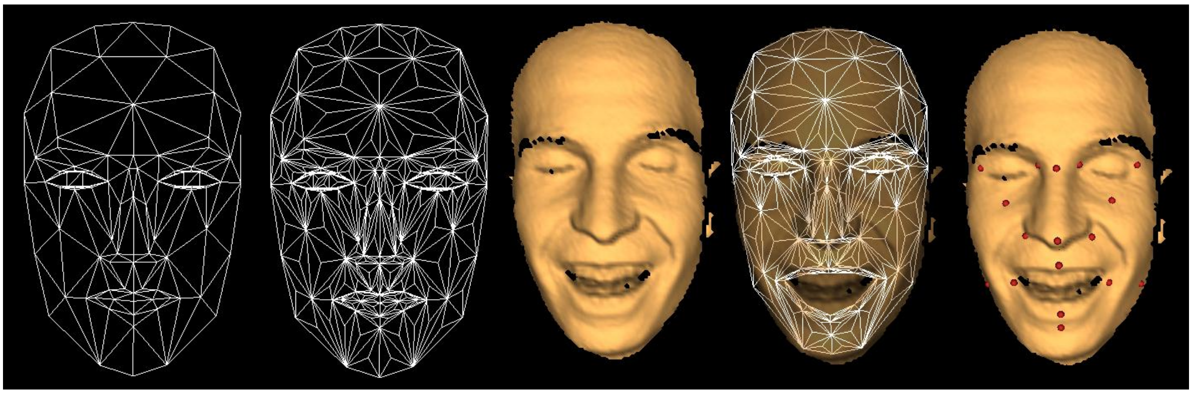

Fig. 1: CANDIDE-3 model, refined model, face scan, model fit to scan, and mapped landmarks.

done on face scans that are only partially complete. Many other correspondence point selection and rejection strategies have been developed [6].

5.3 Transformation step

To find the translation, rotation, and scaling to transform the facial model to minimize the sum of the squared distances, Procrustes analysis [8] can be used. Procrustes analysis is a matrix calculation that can find the rotation, translation and scaling of one set of points to minimise the sum of the squared distance with an other corresponding point set. Although Procrustes analysis can include scaling we encountered a shrinkage effect. When the optimisation of action units and the ICP is applied a lot of times, a shrinkage of the facial model has been observed. To counteract this shrinkage we not only use the sum of squared distance as a distance measurement, but also include the size of the model. As a size for the model we take the average edge length. We now define the distance to be the sum of the squared distance divided by the average edge length. Instead of using a uniform scaling factor in the extended Procrustes analysis, we separately perform a scaling step for the width, height and depth of the model. We calculate the distance with the new sum of the squared distance divided by the average edge length measurement for a number of values in a certain range, and take the value with the smallest distance. Increasing the number of values, or narrowing the range, increases the accuracy of the fitting.

5.4 Optimising action units

For optimising the CANDIDE action units we opted for the simplest solution. An action unit has a value between –1 and 1. By simply calculating the distance of *n* evenly divided values in between you will get a best solution with an error depending on *n*. To increase the accuracy you could simply increase *n*. Besides, the scans also have an error depending on the means of acquisition. Unfortunately the threshold of some of the animation units is taken too large and optimising them can result in unrealistic situations. Also, optimising for instance the eyelids is practically impossible because of the limited accuracy of the scans at the eye regions. Therefore a subset of the animation units is chosen.

A problem encountered when testing the facial models and in particular the CANDIDE-3 model is the limited accuracy of the model. For the initial registration the CANDIDE-3 model is detailed enough. However, when for instance the action units are optimised, some inaccuracy occurs. An example is optimising the animation unit that drops the jaw when the model is registered on a face scan with an open mouth. Because the chin has very limited vertices, the distance is most often less for a registration with the model with an closed mouth, than with the model with an open mouth. This can be resolved by using a facial model with more vertices or as also examined, by using more control points than only the vertices of the facial model. We have split each triangle into three, see Figure 1. This gives a significant improvement of the fitting.

6 Acquiring landmarks after registration

After the registration is complete, the landmarks can be found on the face scan. The landmark locations on the facial model are either actual vertices of the model, or locations

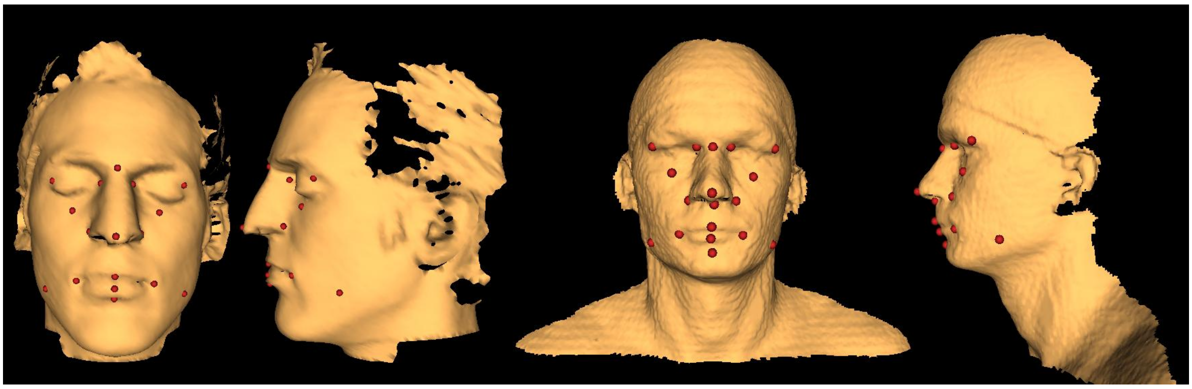

Fig. 2: Landmarks detected on a Minolta scan (left), and a Vitronic scan (right).

on the triangles. Finding the corresponding position on the face scan can be done in the same manner as how the pairs of corresponding points are found in the ICP process. Either simply the closest point is found in the face scan, or a more sophisticated method is used using the normals or surface description of the face scan.

Figure 1 shows, from left to right, the CANDIDE-3 model, the refined models, a face scan made with the Minolta scanner, consisting of 8481 vertices and 16425 triangles, a registration of the refined CANDIDE-3 model onto the face scan including optimisation of the action units, and the locations of the landmarks. Note that the registered model adequately fits the facial expression. Figure 2 shows some more results for the landmark locations on a Minolta scan with 39874 vertices and 78374 triangles (left) and a Vitronic scan with 12928 vertices and 23727 triangles (right). These examples demonstrate that the landmarks around the eyes, nose, and mouth are located well, but the gonion, the point of the angle of the jawbone, is not well located, because the CANDIDE-3 model does extend that far to the back. Even during in vivo landmark detection this point is difficult to find when covered with a lot of tissue, and requires palpation.

7 Concluding remarks

We are currently working on a quantitative evaluation of the landmark detection method. This will be done by computing the distance between the automatically located landmark and a ground-truth position, manually selected in the 3D scan using an interactive viewer. Secondly, we will evaluate our technique on scans of faces with different expressions. A third research line is face recognition on the basis of our fitting method. One way to do so is using the landmarks, e.g. as in [1]. Another way is by comparing the fitted models directly. A big advantage of this approach is that setting the action units of the models to neutral makes the face comparison expression invariant. Future extensions also include more landmarks, for instance those used in the CAESAR project [9].

Acknowledgement

We thank TNO Human Factors, the Netherlands, for the use of their Vitronic scanner, and Laurent Moccozet, MiraLab, for helpful discussions.

References

[1] A. Scheenstra, 3D Facial Image Comparison using Landmarks, MSc thesis Institute of Information and Computing Science, Utrecht University, 2005.

[2] M. Rydfalk, *CANDIDE, a parameterized face*, Report No. LiTH-ISY-I-866, Dept. of Electrical Engineering, Linköping University, Sweden, 1987.

[3] J. Ahlberg, *CANDIDE-3 -- an updated parameterized face*, Report No. LiTH-ISY-R-2326, Dept. of Electrical Engineering, Linköping University, Sweden, 2001.

[4] J. Ahlberg, Extraction and Coding of Face Model Parameters, Department of Electrical Engineering, Linköpings universitet, Sweden, Linköping 1999.

[5] Y. Chen and G. Medioni. Object modelling by registration of multiple range images. *Image and Vision Computing*, 10(3):145-155, 1992.

[6] S. Rusinkiewicz and M. Levoy. Efficient variants of the ICP algorithm. In *Proc. 3rd Int. Conf. on 3-D Digital Imaging and Modeling (3DIM)*, pages 145-152, 2001.

[7] Jason Luck, Charles Little, William Hoff. Registration of Range Data Using a Hybrid Simulated Annealing and Iterative Closest Point Algorithm. Colorado School of Mines, Sandia National Laboratori.
[8] Gower, J.C., Generalized Procrustes analysis. Psychometrika, 40(1), pp. 33-51, 1975.
[9] K. M. Robinette, H. Daanen, E. Paquet. The CAESAR project: a 3D surface anthropometry survey. Proc. Second International Conference on 3D Digital Imaging and Modeling, 1999.
[10] K. Kahler, J. Haber, H. Yamauchi, H-P. Seidel. Head stop: generating animated head models with anatomical structure. Proc. ACM SIGGRAPH Symp. on Computer Animation, pp55-64, 2002.
[11] X. Lu, A. K. Jain. Automatic Feature Extraction for Multiview 3D Face Recognition. In: Proc. Int. Conf. Automatic Face and Gesture Recognition, 2006.