An Intrinsic Coordinate System for 3D Face Registration

W.P. Koppen C.H. Chan W.J. Christmas J. Kittler
Centre for Vision, Speech and Signal Processing, University of Surrey, Guildford, UK
{p.koppen, chiho.chan, w.christmas, j.kittler}@surrey.ac.uk

Abstract

We present a method to estimate, based on the horizontal symmetry, an intrinsic coordinate system of faces scanned in 3D. We show that this coordinate system provides an excellent basis for subsequent landmark positioning and model-based refinement such as Active Shape Models, outperforming other –explicit– landmark localisation methods including the commonly-used ICP+ASM approach.

1. Introduction

The analysis of faces aims at discovering patterns that determine both variation and similarity of the shape and texture. This is important for example in surveillance and biometric identification [7, 9] where the goal is to learn a function that discriminates between individuals, but also in face caricaturing and facial compositing [13, 11] that focus on learning the axes of variation of specific facial features, as well as medical research studies and clinical purposes [1, 3] in which groups of people are classified into carrying a trait or not. In all cases, the analysis requires an established correspondence between faces. Finding this correspondence is called registration.

Registration may be carried out (or initialised) manually by placing a set of landmarks in each image, but for large datasets this will become a tedious if not prohibitive process, and automatic methods may provide an alternative. In 2D images faces scaled to the same size are often assumed to be roughly aligned. By modelling the depicted face, e.g. using Active Appearance Models (AAM), a more accurate registration can be achieved. In 3D images the Iterative Closest Point (ICP) algorithm computes the translation and rotation parameters to align one shape to another. Similar to the AAM method in 2D images, 3D shape models can be employed for more accurate registration, an example of which is the Active Shape Model (ASM).

In general the model-based approach is posed as a refinement of an initial rigid alignment, $i.e.$, when the feature points (landmarks) of a face have been roughly determined the model can position them more precisely. The extent to which it can do this is limited though as it is a local optimisation rather than global, and so gross localisation errors unavoidably lead to misregistration.

In this paper we introduce a new method to initialise face registration that is based on symmetry detection to find an intrinsic coordinate system of the face. An arbitrary reference shape, e.g. a fixed set of landmarks, can subsequently be positioned in this reference frame to provide initial registration. The method is fully automatic and copes well with cases where big gaps of missing data exist in the face surface. We compare it against ICP- and AAM-based registration procedures. We further investigate in each of the cases if and how well ASM can refine the resulting estimates.

It should be noted that although people have little difficulty locating the face and its features, manual annotation is not exact either. As has been shown before and as we confirm in this paper, there is still a considerable amount of variation in the position of manually placed landmarks. This variation provides a good reference to assess the performance of automatic methods.

1.1 Landmark-based registration methods

Whereas our method estimates a reference frame into which any set of landmarks can be positioned, other methods are based on the detection of specific points directly. This section describes some approaches.

1.1.1 Iterative closest point

A popular approach is the class of algorithms based on the iterative closest point technique. In this approach a sequence of estimates progressively reduce the error of alignment between two sets of points (here the set of landmarks against a shape). At each iteration the correspondence between the two point sets is computed.

Based on this, a transformation is computed that reduces the error. In our experiments we have used the publicly available LMICP algorithm [6].

1.1.2 Active appearance models

In principle active appearance models provide a method of 2D face recognition [5], but we can use it to obtain a registration of 3D faces. A 3D face scan consists of three parts; 1) a surface, 2) a texture, and 3) a mapping of the texture onto the surface. The texture mapping provides a conversion from the 2D image to the 3D shape, which can be used to map AAM detected points.

In our dataset the mapping defines identical triangles on the surface, $\mathbf{r}_1^{3D}, \mathbf{r}_2^{3D}, \mathbf{r}_3^{3D}$, and on the texture, $\mathbf{r}_1^{2D}, \mathbf{r}_2^{2D}, \mathbf{r}_3^{2D}$. A point $\mathbf{p}^{2D}$ in the texture image can be expressed in terms of the triangle points enclosing it using barycentric coordinates:

$$\mathbf{p}^{2D} = t_1\mathbf{r}_1^{2D} + t_2\mathbf{r}_2^{2D} + t_3\mathbf{r}_3^{2D} \tag{1}$$

where (t_1, t_2, t_3) are the barycentric coordinates. The corresponding 3D coordinates are found by

$$\mathbf{p}^{3D} = t_1\mathbf{r}_1^{3D} + t_2\mathbf{r}_2^{3D} + t_3\mathbf{r}_3^{3D} \tag{2}$$

1.2 Active shape models

Active shape models learn the patterns of variability (deformation) from a training set of annotated images. The learned model can subsequently be applied to iteratively refine a set of points in a new image. An advantage of this method is that any rigid estimation of landmarks can now be deformed to fit the specific shape and thus provides potentially a much better registration.

In [10] this method is modified to work on 3D shapes. Landmark locations are simply expressed in 3D, and landmark shapes are expressed using the shape index, which is calculated from the principal curvatures:

$$S = \frac{1}{2} - \frac{1}{\pi}\tan^{-1}\frac{\kappa_1 + \kappa_2}{\kappa_1 - \kappa_2} \tag{3}$$

The initial set of points can be estimated using any preferred method, e.g. ICP (see [10]), AAM, or the method we propose in the following section.

2. An intrinsic coordinate system

In this section we describe a method to accurately determine an intrinsic coordinate system for any 3D face shape primarily based on its horizontal symmetry. The use of symmetry to determine an intrinsic coordinate system for the face is actually not new. In their early work on 3D face recognition [2] the authors propose an iterative approach of guessing and validating a plane of symmetry based on Gaussian curvature. In fact, our approach shares some resemblance to this method as we too derive the symmetry from Gaussian curvature. The key difference however, is that the method proposed here computes the plane of symmetry directly and only from a very small set of points – the local extrema.

2.1 The horizontal axis

We determine the horizontal axis from the local extrema in the curvature of the face. Gaussian curvature can be defined as the product of the two principal curvatures

$$K = \kappa_1\kappa_2$$

with $\kappa_1 > \kappa_2$. Thus we can distinguish three types of shape with extremal curvature: saddles ($\kappa_1 > 0 > \kappa 2$), cups ($\kappa_1 > \kappa_2 > 0$), and caps ($0 > \kappa_1 > \kappa_2$). Let us denote the points at which K has an extremum

$$P = \{\mathbf{p}_1, \mathbf{p}_2, \mathbf{p}_3, \ldots\} \tag{4}$$

then two points, $\mathbf{p}_i$ and $\mathbf{p}_j$ can only be symmetric if they are of the same shape type.

Furthermore, if the two points are symmetric on the face, then the vector $\boldsymbol{n} = \mathbf{p}_i - \mathbf{p}_j$ is normal to the plane of symmetry, A, and the intersection of $\boldsymbol{n}$ with A should lie at $(\mathbf{p}_i + \mathbf{p}_j)/2$ midway between $\mathbf{p}_i$ and $\mathbf{p}_j$. Since the face is quasi-symmetric this does not hold exactly but is reasonably accurate.

Based on the above observations two collections of vectors can be defined:

$$V^+ = \{\mathbf{p}_i - \mathbf{p}_j | \mathbf{p}_i, \mathbf{p}_j \text{ of the same type}, i \neq j\} \tag{5}$$
$$V^- = \{\mathbf{p}_i - \mathbf{p}_j | \mathbf{p}_i, \mathbf{p}_j \text{ of different type}, i \neq j\} \tag{6}$$

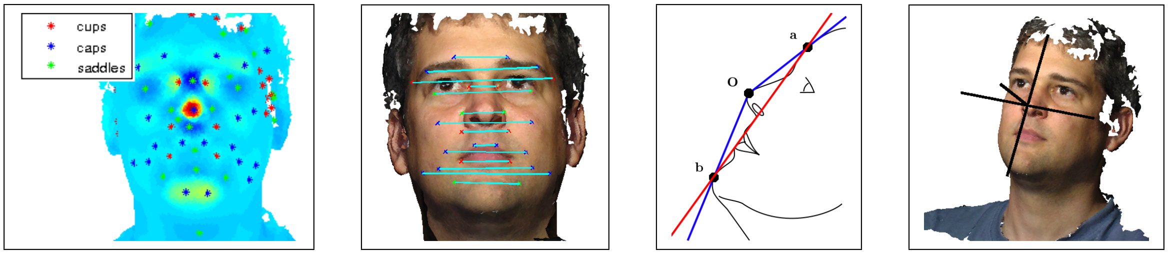
The intuition is that symmetry causes a small set of tightly aligned vectors in V^+, whereas all other vectors both in V^+ and V^- are more diffuse. We formalise this intuition by a simple estimation of the density of vectors pointing in each direction.

$$\rho^+(\boldsymbol{u}) = \frac{\sum_{\boldsymbol{v}\in V^+} f(\boldsymbol{u}, \boldsymbol{v})}{2\pi(1 - \cos\alpha)} \tag{7}$$

computes the density of V^+ over a spherical cap of radius α (in radians) centred on $\boldsymbol{u}$, with

$$f(\boldsymbol{u}, \boldsymbol{v}) = \begin{cases} 1 & \text{if } \cos^{-1}\frac{\boldsymbol{u}\cdot\boldsymbol{v}}{|\boldsymbol{u}||\boldsymbol{v}|} < \alpha \\ 0 & \text{otherwise} \end{cases} \tag{8}$$

Similarly we denote ρ^- to compute the density on V^-.



(a) Curvature and local extrema (b) Estimation of $\boldsymbol{x}$ (c) Estimation of $\boldsymbol{y}$ (d) Intrinsic coordinate system

Figure 1. From the local extrema of Gaussian curvature (a) we derive a collection of horizontal line segments (b). The plane of symmetry is estimated from their midpoints. The points in this plane maximizing the angle on O provide a robust estimate of the vertical axis (c). The resulting intrinsic coordinate system is shown in (d).

The plane of symmetry A is characterised by its normal and its distance to the origin (in world coordinates). The normal is the horizontal axis of the face

$$\boldsymbol{x} = \arg\max_{\boldsymbol{u}} \rho^{+}(\boldsymbol{u}) - \rho^{-}(\boldsymbol{u}) \tag{9}$$

where the subtraction ρ^- makes it more robust to incidental alignment of arbitrary vectors. The distance along $\boldsymbol{x}$, which we shall denote h_0, is estimated from the midpoints of line segments in V^+ subject to f. Let $M = \{\mathbf{m}_1, \mathbf{m}_2, \mathbf{m}_3, \ldots\}$ be those midpoints, then

$$h_0 = \mu_{1/2}(\{\mathbf{m}_i^T \boldsymbol{x}\}) \tag{10}$$

is their median projected onto $\boldsymbol{x}$. Asserting $|\boldsymbol{x}| = 1$ this is directly the distance of A to the origin.

2.2 An origin

Most common in literature is to position the origin at the nose tip directly [8, 4, 12]. An overlooked fact is however that most people have a nose that points either left or right. Forcing the origin to coincide with this point would therefore break symmetry. A better choice is to project the point of the nose tip, $\mathbf{p}_t$, onto A:

$$\mathbf{O} = \mathbf{p}_t + (h_0 - \boldsymbol{x}^T \mathbf{p}_t)\boldsymbol{x} \tag{11}$$

$\mathbf{p}_t$ is obtained from the point with maximum curvature (type cap) close to A.

2.3 The vertical and depth axes

We will now continue with the determination of the vertical axis, $\boldsymbol{y}$. The axis of depth, $\boldsymbol{z}$, will then follow from the cross product of the first two.

Actually, "the" vertical axis of a face is somewhat ill-defined. One approach is to put the nose bridge under a fixed angle [12], but due to the large variation in nose shapes we found this to give only moderate alignment. Instead we propose to define $\boldsymbol{y}$ along the two points on the vertical section that make a maximal angle with $\mathbf{O}$ (see Fig.1(c)), each at least 4cm from $\mathbf{O}$:

$$(\mathbf{a}, \mathbf{b}) = \arg\max_{\mathbf{a},\mathbf{b}} 1 - \frac{(\mathbf{a}-\mathbf{O})^T(\mathbf{b}-\mathbf{O})}{|\mathbf{a}-\mathbf{O}||\mathbf{b}-\mathbf{O}|} \tag{12}$$

$$\boldsymbol{y} = \frac{\mathbf{b}-\mathbf{a}}{|\mathbf{b}-\mathbf{a}|} \tag{13}$$

$$\boldsymbol{z} = \boldsymbol{x} \times \boldsymbol{y} \tag{14}$$

2.4 Aligning two coordinate systems

The alignment of two shapes, faces or landmarks, is based on their respective intrinsic coordinate systems. Suppose we have two coordinate systems, $(\mathbf{O}, \mathbf{R})$ and $(\mathbf{O}', \mathbf{R}')$, with $\mathbf{R} = \begin{bmatrix}\boldsymbol{x} & \boldsymbol{y} & \boldsymbol{z}\end{bmatrix}$ and similar for $\mathbf{R}'$, then the transformation to align the second onto the first, for any point $\mathbf{p}'$, is defined as:

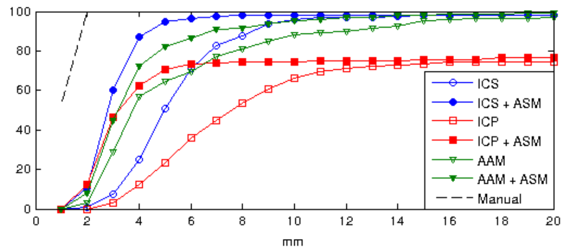
$$\mathbf{p} = \mathbf{R}\mathbf{R}'^T(\mathbf{p}' - \mathbf{O}') + \mathbf{O} \tag{15}$$

Often, a set of landmarks is defined in world coordinates, *i.e.*, $\mathbf{O}' = (0, 0, 0)$ and $\mathbf{R}' = I_3$ the identity matrix. In that case Eqn.15 simplifies to $\mathbf{p} = \mathbf{R}\mathbf{p}' + \mathbf{O}$.

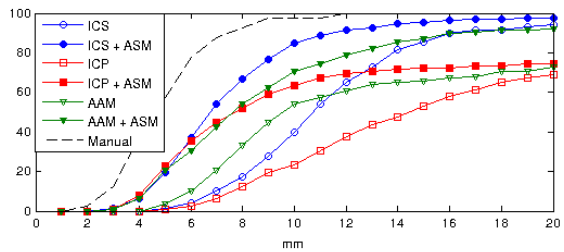
3. Experiments

A set of 322 faces have been captured with a 3dMD Face camera system, which produces images of roughly 100,000 triangles spanning 50,000 vertices, but these numbers vary between scans. The resolution of texture images is around 2,100×1,100 pixels.

Each image was manually annotated, and this annotation is used as the ground truth for evaluation. The same 14 landmark positions were used as in [10] (eye



(a) Mean error



(b) Maximum error

Figure 2. Cumulative error distributions

corners, nose bridge between the eyes, nose tip, nose wings, mouth corners and central top and bottom, chin tip, and the dip between chin and mouth). A separate set of five scans was captured and annotated by five different people to assess the theoretical error in ground truth labels. In this case their average position provides the ground truth.

We divided the scans into 14 equally sized sets. Over 14 rounds the algorithms were evaluated on one set with training performed on the remaining sets. We report performance over all rounds combined.

Evaluation is carried out by measuring per face for all landmarks, $P = [\mathbf{p}_1, \mathbf{p}_2, \dots, \mathbf{p}_{14}]$, the distance in millimetres to their respective ground truth points, $X = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_{14}]$. We compute the mean and maximum

$$e_\mu(P) = \frac{1}{14}\sum_{i=1}^{14} |\mathbf{p}_i - \mathbf{x}_i| \quad (16)$$

$$e_{\max}(P) = \max_i |\mathbf{p}_i - \mathbf{x}_i| \quad (17)$$

and present their cumulative distribution over all faces.

The results of different registration methods are shown in Fig.2. It shows the percentage of samples that have an error below a threshold in millimetres. The ideal annotation would touch the point $(0\text{ mm}, 100\%)$. ICS (Intrinsic Coordinate System) is the method proposed in this paper.

The figures show that ASM is able to refine landmark positions for any method. In particular ICS+ASM seems to be a good combination, annotating more than half of the scans entirely within 7mm accuracy and with a mean error below 4mm in 90% of the cases.

4. Conclusions

We have presented a novel algorithm for estimating the position and orientation of a 3D face scan. We have demonstrated its power in registering to a set of landmarks compared to other well known methods.

ICP and ICS combine well with ASM refinement. Apparently ASM corrects the type of error specific to rigid initialisation methods. Active appearance models perform well on their own, but ASM is not a suitable method for refinement here. Overall the best automatic registration is achieved by ICS+ASM.

Manual annotation is still much more accurate, but not perfect either. The results were mainly included to get a feeling for the accuracy of the ground-truth, and to allow comparison between different datasets.

References

[1] I. Atmosukarto, K. Wilamowska, C. Heike, and L. Shapiro. 3d object classification using salient point patterns with application to craniofacial research. *Pattern Recogn*, 43:1502–1517, 2010.

[2] J. Cartoux, J. Lapreste, and M. Richetin. Face authentification or recognition by profile extraction from range images. In *3DSCEN, Workshop*, 1989.

[3] J. Cox-Brinkman, A. Vedder, C. Hollak, L. Richfield, A. Mehta, K. Orteu, F. Wijburg, and P. Hammond. Three-dimensional face shape in fabry disease. *Eur J Hum Genet*, 15:535–542, 2007.

[4] T. Faltemier, K. Bowyer, and P. Flynn. Using multi-instance enrollment to improve performance of 3d face recognition. *CVIU*, 112:114–125, 2008.

[5] Z. Feng, J. Kittler, W. Christmas, X. Wu, and S. Pfeiffer. Automatic face annotation by multilinear aam with missing values. In *ICPR*, 2012.

[6] A. Fitzgibbon. Robust registration of 2d and 3d point sets. *Image Vision Comput*, 21:1145–1153, 2003.

[7] I. Kakadiaris, G. Passalis, T. G., E. Efraty, P. Perakis, D. Chu, S. Shah, and T. Theoharis. *Handbook of Face Recognition*, chapter 17. Springer Publishing, 2011.

[8] A. Mian, M. Bennamoun, and R. Owens. An efficient multimodal 2d-3d hybrid approach to automatic face recognition. *TPAMI*, 29:1927–1943, 2007.

[9] G. Passalis, P. Perakis, T. Theoharis, and I. Kakadiaris. Using facial symmetry to handle pose variations in real-world 3d face recognition. *TPAMI*, 33:1938, 2011.

[10] M. Ruiz and J. Illingworth. Automatic landmarking of faces in 3d - alf3d. In *VIE, 5th Int Conf*, 2008.

[11] C. Solomon, S. Gibson, and M. Maylin. A new computational methodology for the construction of forensic, facial composites. In *Comput Forensics*. 2009.

[12] L. Spreeuwers. Fast and accurate 3d face recognition. *IJCV*, 93:389–414, 2011.

[13] X. Zou, J. Kittler, and J. Tena. A morphing system for effective human face recognition. In *VIE*, 2008.