STEREO-ASSISTED LANDMARK DETECTION FOR THE ANALYSIS OF 3-D FACIAL SHAPE CHANGES

A J Naftel¹, Z Mao¹ and M J Trenouth²
¹Department of Computation, UMIST, Manchester M60 1QD
²Department of Oral and Maxillofacial Surgery and Orthodontics, Royal Preston Hospital, Preston PR2 4HT
andrew.naftel@co.umist.ac.uk

ABSTRACT

Techniques for the three-dimensional analysis of facial morphological changes resulting from surgical treatment have recently been proposed. These approaches largely rely on expert manual intervention for placing anatomical landmarks on 3-D facial models which is time consuming and error prone. A new automated approach is presented here which combines a stereo-assisted active shape model for 3-D landmark extraction with geometric morphometrical tools for a statistical analysis of facial shape changes. A stereo-photogrammetric imaging system is used for acquiring 3-D face models. This incorporates an active shape detection phase which is used for the automatic localisation of 2-D facial features in greyscale stereo images. The detected features can then be used to generate 3-D soft tissue landmarks by stereo correlation matching and disparity map interpolation. Generalised Procrustes analysis, principal component analysis and thin plate spline decomposition are then applied to the analysis of shape changes in 2-D facial midline profiles and extracted 3-D facial landmarks. The proposed method is validated both statistically and visually by characterizing shape changes induced by mandibular repositioning using a bite block in a heterogeneous sample of 20 patients attending a weekly orthodontic clinic. It is shown that the method is capable of distinguishing between changes in facial morphology due to mandibular repositioning and changes due to other factors such as growth and normal variation within the patient cross-sectional sample.

1 INTRODUCTION

There is increasing interest in methods for acquiring accurate 3D models of the human face. Such methods have important applications in medicine (e.g. pre- and post-operative planning for facial surgery), entertainment (e.g. creating human avatars to

Page 1 of 22
populate virtual environments) and telecommunications industries (e.g. facial encoding, MPEG-4, videoconferencing). Over the years various optical techniques have been proposed for tackling this problem, including digital photogrammetry [1,2,3], light striping [4] and laser scanning [5]. Light striping is well-suited to the recovery of surfaces which are gently curving (e.g. human back shape [6]) but it tends to result in sparse data collection when applied to the face [4]. Laser scanning is accurate but scanning times are slow (~15s), and it is not truly capable of capturing the surface texture.

Results obtained by photogrammetric vision systems employing stereo matching have been steadily improving [8,9]. This approach is capable of recovering dense point maps, image capture is instantaneous and key facial landmarks can be located to an accuracy of 0.5mm using high resolution cameras. Also, detailed 3-D mesh models can be built and rendered in a few seconds on a high performance PC. Systems that exploit stereo techniques for building accurate 3-D computer models of the face are now a commercial reality (TcTi, C3D, 3dMD) although market take-up has been slow.

Clinicians, however, are not merely interested in accurately imaging facial shape but in the ability to detect changes in the image. When evaluating the success of surgical treatment or orthodontic appliance therapy, it is also important to be able to distinguish between changes in morphology due to treatment and changes due to other factors such as growth and normal variation.

Techniques for three-dimensional analysis of shape changes resulting from surgical correction have recently been proposed [1,2]. Among the existing non-landmark based methods for shape analysis, the most successful approach has been the use of surface curvature methods [3,4]. However, the calculation of surface curvature is unavoidably noise-sensitive which is a problem for 3-D facial capture techniques such as stereo imaging [5,6] and laser scanning [8]. In this case, pre-processing techniques such as smoothing and local surface fitting must be carried out beforehand, which may result in removal of important surface detail. Furthermore, the use of a curvature-based approach precludes statistical comparison of group differences and interpretation is difficult.

Recent advances in the field of geometric morphometrics [9,10] offer an effective means for medical applications of landmark-based analysis. The use of landmarks permits a variety of morphometric tests to be applied in order to detect differences in shape. One
can test a null hypothesis for differences in shape and perform a principal component analysis (PCA) to analyse the main modes of shape variation in a patient sample. However, to date, the application of morphometric tests to the dentofacial complex have largely relied on expert knowledge for the manual detection of landmarks [11,12,13,14] or curved outlines, a process which is time-consuming and error prone. There has been little attempt to develop automated tools to characterise such shape changes.

In this paper, we present a new approach towards automating the analysis of 3-D facial shape changes. The aim of this paper is to show that the proposed approach is capable of distinguishing between changes due to simulated corrective treatment and those due to other factors such as growth and normal variation. The organisation of the paper is as follows. In section 2 we describe the main features of the digital stereo-photogrammetric imaging system used for acquiring 3D face models. The method for extracting 3-D facial landmarks and midline profiles is discussed in section 3. In section 4 we give an overview of the morphometric techniques used for analysing changes in facial shape. An experimental evaluation of the method is presented in section 5, and the results of analysing morphological changes in a patient sample are discussed therein. Finally, we present some conclusions and limitations of the current method.

2 FACIAL IMAGE ACQUISITION

A standard stereo camera set-up is used to capture the facial image pairs. The two cameras are attached to a specially designed stereo base which sits on a tripod. Each camera can translate and rotate along this base so that the convergence angle, baseline separation and height are adjustable. Digitally controlled slide projectors are used to flash separate random texture and structured light patterns onto the subject's face. Random texture is used to aid the search for point correspondences in the matching phase. Without this large areas of the face are relatively featureless and stereo-correlation produces many false matches.

The CCD cameras used have a resolution of 768 x 576 pixels and these are digitised in a framegrabber able to store up to sixteen 8 bit grey level images in RAM. The cameras are pre-calibrated using precisely defined circular control point targets.
Details of the camera calibration procedure have been described elsewhere [17]. Image pairs acquired using texture \((I_L, I_R)\) and structured light \((S_L, S_R)\) projection are given in figure 1. Using digitally controlled switching between cameras and projectors, the time between capture is less than one second. Any slight subject movement is unimportant since the structured light pair is used only at the image partitioning stage.

Clearly, it is not possible to capture the entire face with only one stereo pair. We would require an additional stereo pair for the left and right sides of the face, followed by a step to register the three facial models. This work is currently underway, but for the purposes of landmark extraction described here, a fronto-parallel stereo pair is sufficient as it covers about 70% of the viewable surface.

The system is capable of generating dense 3-D surface maps by an image partitioning strategy and stereo matching phase, the details of which can be found in [17]. The accuracy of euclidean reconstruction has been established by analysing images of a plaster cast model of a human face containing 13 simulated facial landmarks. This model was placed at nine different orientations and locations within the calibration space. Stereo image pairs (figure 2) were captured at each pose and the corresponding landmark features were detected using a centroid detection algorithm. These landmark points were then reconstructed to produce nine independent 3-D point sets. The point sets were then registered to a common coordinate frame using a standard approach [18], and the mean landmark positions determined. The standard deviations of the registered landmarks are found to be 0.12, 0.11 and 0.49mm in \(x, y\) and \(z\) coordinates respectively.

The 3-D ground truth was available through precise electronic theodolite survey and this can be used to provide a quality measure for Euclidean reconstruction error. The quality measure \(D_{\text{error}}\) can then be defined as a root mean square (rms) error term

\[
D_{\text{error}} = \left(\frac{1}{N_L} \sum_{N_L} \left(\hat{x}_C - x_T\right)^2 + (\hat{y}_C - y_T)^2 + (\hat{z}_C - z_T)^2\right)^{1/2}
\]

(1)

where \((\hat{x}_C, \hat{y}_C, \hat{z}_C)\) = computed mean landmark coordinates, \((x_T, y_T, z_T)\) = ground truth coordinates and \(N_L\) = number of landmarks. For this experiment it was found that \(D_{\text{error}} = 0.55\)mm. Although the error measures quoted above are obtained under somewhat idealised conditions – high contrast circular shaped markers on a light coloured phantom
object – it is reasonable to assume that $D_{error}$ of 0.5 to 1mm are achievable when modelling human face point sets.

3 AUTOMATIC 3-D LANDMARK EXTRACTION

The method developed for automatic facial shape change analysis is a combination of stereo-assisted feature detection and morphometric techniques. One advantage of using a stereo-based image acquisition system is that the surface texture is captured which facilitates robust 3D landmark extraction. Coupled with this, the use of morphometric techniques provides a geometrically and statistically meaningful way to characterise the shape change.

An active shape model (ASM) is a statistically-based technique for building geometric models of the shape and grey level appearance of a variable object in order to automatically locate new instances of the object in 2D images [15]. It has been used successfully to locate the main facial features in images obtained under widely varying conditions [16]. This technique is now extended to 3-D facial landmark extraction by using stereo correspondence search and surface map interpolation.

One-hundred and seven (107) landmark points in the left stereo image were defined to represent the facial features as shown in figure 3. Some points corresponded to true landmarks, such as corners of eyes and mouth, and other points were generated automatically by sampling at equally spaced intervals along the hand digitised contour. They were manually digitised on 25 training images from the facial database and aligned using the Procrustes algorithm [19] to compute a mean shape and build a point distribution model (PDM). Principal component analysis on the covariance matrix of deviations from the mean shape was carried out to yield a set of basis vectors describing the main modes of shape variation in the training data. At the same time, the grey level appearance model at each shape point was constructed in the same way as for the PDM. This completes the ASM training procedure. The average grey level profiles were then used to drive and deform the shape model iteratively to locate 2-D facial features in stereo images that lay outside the training set. Full details of this method have been described elsewhere [15].
Initial placement of the shape model on a new image plays an important role to ensure correct convergence of the ASM search. Since the facial orientation was standardised at the data capture (inside the same control volume and facing the similar direction), an initial position of the PDM, defined by the scale and translation parameters, was calculated by locating the horizontal and vertical edges of the facial borders (figure 4) using the integral projection [20] of the intensity image. A Sobel horizontal edge filter [21] was applied to the intensity image followed by the vertical projection of the filtered image. Figure 5 depicts the vertical projection of the horizontal edge, where \( L \) and \( R \) indicate the left and right facial borders. The first peak detects the edge of a dark background drape common to all images. By finding the minimum and maximum values along the horizontal axis, the left and right facial borders were located. A similar operation was carried out for the upper facial border and the vertical position of the eyes and eyebrows as illustrated in figure 4. This information was then used to determine the initial scale and pose. Based on this initialization, 2-D facial landmark detection in the left stereo image was carried out. A maximum limit of 50 iterations was imposed when searching for the optimum position. Figure 6 shows an example of facial feature detection by ASM in a 2-D image. The left picture indicates the initial position of the shape model and the right picture is the optimum position after an iterative search procedure. Once the position of the 2-D feature was located it was then converted into 3-D space by the stereo correspondence search described next.

For every feature point detected by the ASM in the left image, a search for its corresponding point in the right image was carried out by a stereo correlation matching algorithm [17]. Since the matching algorithm was performed on integer image pixel positions whereas the positions of the detected features can have float point values and also the correspondence search for a particular feature point may fail, an interpolation algorithm is required to produce the disparities for these feature points from their neighbouring points. The TPS mapping function was selected for this purpose because it has certain desirable interpolatory properties [22]. The interpolation was performed locally and a variable window size was used to control the number of image points for calculating the TPS parameters. A smaller window size can be selected in regions where the disparity map is dense, whilst a larger window is needed for regions containing sparse data. Once the disparity values for all the features had been obtained, these point
correspondences were converted into 3-D coordinate space using a set of camera parameters obtained from a metric calibration procedure [18]. Figure 7 shows an example of automatically detected 3-D facial landmarks by stereo-assisted ASM. Here, 90 facial landmarks are displayed and will be used in the further analysis of shape change. The 2-D landmarks on the facial border (figure 3) are not reliable due to occlusion in the stereo pair and hence are used purely for the purpose of segmenting the face from the background.

Apart from the 3-D facial landmark points, facial midline profile was generated by the intersection of the 3-D surface model with the facial symmetry plane (figure 8). The symmetry plane was automatically computed using a surface-based registration method [24]. The obtained midline profile was then resampled to produce 39 pseudo-landmark positions for each 3-D image model (figure 9) and these landmarks will be used in the shape analysis because they contain important facial shape information.

### 4 MORPHOMETRIC ANALYSIS OF FACIAL SHAPE CHANGE

Three main approaches for analysing change in facial shape were employed; (i) generalized Procrustes analysis (GPA), (ii) PCA in the tangent space and (iii) TPS decomposition. These techniques were applied to the landmark data extracted from the previous stage to test a null hypothesis of shape change, analyse the structure of shape variability and visualize the change as a deformation over local and global scales.

GPA is primarily used for estimating an average shape and small scale variation about that average. It registers a series of landmark configurations by removing translational and rotational differences and scaling them until they most closely match. The registered landmark configurations constitute Kendall’s shape space [26] with Procrustes distance as metric. PCA in the tangent plane to this space provides an effective means of investigating the structure of shape variability and reducing the dimensionality of the data prior to multivariate analysis. Inference testing can be carried out in the tangent space coordinates provided that shape variation is small [9]. The orthonormal eigenvectors of the covariance matrix of a subset of tangent coordinates define the principal components (PCs). When these PCs are sorted in descending order of
eigenvalues for the covariance matrix, the main modes of shape variation can be described by the first few significant PCs. A TPS can be used together with a Cartesian transformation grid to exactly map the biological shape change between two homologous data sets with minimum bending energy. The associated principal warps decompose shape change into local and global components of successively increasing scale, analogous to a Fourier analysis, and is an effective means for visualization. A detailed exposition of these techniques with further applications to anthropometric landmark data can be found in the text by Dryden and Mardia [10].

5 EXPERIMENTAL EVALUATION AND RESULTS

The imaging system was used to acquire fronto-parallel stereo images of 61 patients (aged 10 to 26 years; mean 16 years) attending an orthodontic clinic at the Royal Preston Hospital. To test the efficacy of the proposed image analysis method, a deliberate gross change in the position of the patient’s lower jaw was induced by means of a bite block. This induces a three-dimensional morphological change in the soft tissue profile having a similar order of magnitude to that which could be achieved clinically by functional orthodontic appliances or by surgical correction. Two sets of stereo images were captured for each patient, one set before bite block insertion and the other afterwards. The reconstructed 3D facial models were then divided into two groups; group I denoting patient profiles prior to insertion of the bite block and group II after insertion. A morphometric analysis of facial shape change was then carried out to evaluate shape differences between the two groups and to attempt to separate out changes due to mandibular repositioning from those due to normal shape variation in the homogeneous sample.

The stereo-assisted ASM landmark detection algorithm was tested on 25 facial images outside the training database and a success rate of over 80 percent was obtained. Manual landmark extraction is required on the images where stereo-assisted ASM fails. The successful rate can be further improved by increasing the size of the training set using a bootstrapping procedure [16] and improving the image quality and resolution of stereo capture.
The morphometric analysis was carried out on a sub-sample of 20 patients from each group to characterise the shape difference. These 20 patients were selected outside of the ASM training data set and whose facial images had reasonably good capture quality. Facial shape analysis was first carried out on the 2-D midline profiles of 20 randomly selected patients to test the proposed method. Figure 10 illustrates the scatter plots of the midline profiles which were partially Procrustes aligned using the unchanged (upper 25) landmarks. The solid lines indicate the Procrustes averaged profiles for groups I and II. As expected, the lower jaw for the group II sample has been displaced forward by the bite block. This deformation can be depicted more clearly in Figure 11 by a Cartesian transformation grid calculated using thin plate spline mapping. The left and right pictures are the average profiles for group I and group II respectively. Figure 11 shows that when shape change is viewed as a deformation, the biggest change is localised in the region of the lower jaw.

Procrustes superposition was used to align this set of midline profiles, followed by PCA on the Procrustes tangent coordinates, to reveal the main modes of shape variation. Figure 12 illustrates the shape variation represented by the first three principal components (PCs). Row $i$ denotes the $i$th PC ($i = 1, 2, 3$) and each column from left to right displays a multiple $c$ of the standardised PC score (where $c = -3, -2, -1, 0, 1, 2, 3$). This is equivalent to a midline plot of the mean shape ($c = 0$) and a sequence of plots up to $\pm 3$ standard deviations from the mean outline. The effect of shape variation can be seen by scanning the outlines along each row from left to right. The percentages of shape variability accounted for by the first 3 PCs are 35.1%, 16.9% and 15.1%. The largest shape variation captured in PC 1 was found in the lower and upper jaw region due to the insertion of a bite block. PC 2 largely accounts for the variation (growth) in nose shape (together with some jaw repositioning) which is to be expected given the age range in our sample (10-26 years). The first two PCs resulting from this analysis are plotted against each other in Figure 13. It can be seen that the two groups have been partially separated by the plot with group I (‘o’) largely occupying the lower half and group II (‘*’) the upper half of the figure. Selecting 8 landmarks in the mouth region from the midline profile (landmarks 25, 27, 29, 31, 33, 35, 37 and 39 in figure 9) and performing a two-sample Hotelling’s $T^2$ test using Procrustes aligned coordinates, the mean shapes of the two groups are significantly different at $P < 0.05$ ($P \sim 0.028$).
However, the midline profiles only utilise 2-D coordinate information. For a more revealing picture, the full 3-D landmark data set derived from the stereo-assisted ASM search are compared. Assigning landmark data to groups I and II as before, the extracted data were submitted to three-dimensional GPA and PCA. The procrustes scatter of the facial landmarks in anterior \((x-y)\) and anterior-posterior \((z-y)\) planes are shown in Figure 14. The group II outline is displaced downwards and forwards consistent with mandibular repositioning due to insertion of a bite block. Hotelling’s \(T^2\) test was performed on the first 3 PCs which accounted for 19.5%, 15.1% and 9.9% of the total variation and the difference in mean shape is statistically significant at \(P \sim 0.00006\). The evidence of difference in mean shape for 3-D landmark data is not in doubt and is clearly stronger than for 2-D data.

Three-dimensional GPA and PCA was then applied to a subset of the 3D landmarks in the mouth region alone. Hotelling’s \(T^2\) test was performed on the first 3 PCs which accounted for 30.2%, 11.9% and 10.1% of the total variation in the sample \((n = 20)\) and the difference in shape was found to be statistically significant at \(P < 0.00001\).

6 DISCUSSION

Previous morphometric studies of cephalometric data [11,12,14,25] have relied on manual digitization of landmark coordinates sparsely distributed over the sample image. This approach was adopted by Singh and has proved successful in locating differences between class I and III morphologies [25,28,29] and different ethnic groups with class III relationships [27]. However, they only analysed 2-D cephalometric X rays and relied on a small number of anatomical landmarks to perform Procrustes analysis followed by finite element or thin plate spline analysis. Their method relied upon accurate location of anatomical landmarks which have shown to be more variable than mathematical centroid points or fixed relations between anatomical points [30]. Ayoub and Stirrups [31] clearly demonstrated that finite element analysis is sensitive to small errors in landmark location even if these are located with a high degree of reproducibility \((r > 0.98)\).

The present study did not depend on manually locating homologous anatomical landmarks but instead applied the active shape model (ASM) for automatic facial landmark
extraction. The ASM is a statistical-based model using shape and appearance (grey-level) statistics extracted from a training set of images. As such it provides a more objective method of image location.

In clinical orthodontics it is not enough just to accurately image the facial shape but essential to be able to detect changes in the image. When evaluating the success of appliance therapy it is also important to be able to distinguish between changes in morphology due to treatment and changes due to other factors such as growth and normal variation. The present method described appears to be successful in separating out the changes due to bite block insertion equivalent to treatment and changes due to longitudinal growth and cross-sectional normal variation.

However, this is a pilot study and the developed methodology only applied to the patients with simulated facial shape change introduced by inserting bite blocks. When applying this method in clinical practice, several other factors should be considered where improvements could be made:

− an extra camera pair should be added to image capture to increase the facial coverage and reduce occlusion,
− the size of the ASM training data set should be increased to improve the successful rate of automatic landmark extraction,
− the method based on 3-D surface geometry [32] may be incorporated to the feature extraction algorithm for those images where stereo-assisted ASM is unsuccessful.

7 CONCLUSIONS

It has been shown that reliable 3-D landmark digitization can be automated in facial video images using a combination of feature detection by active shape model and stereophotogrammetric analysis. The proposed method is sufficiently sensitive to detect statistically significant changes in facial soft tissue shape due to mandibular repositioning in a cross-sectional patient sample.

The experimental results signify a relative displacement of the average shape of the mouth and lower jaw forwards and slightly downwards after insertion of the bite block. This would be the desired effect of surgical treatment on patients presenting class
II and class III abnormalities. It has been shown that such changes can be validated both visually and statistically. It suggests that the proposed method may be useful for auditing clinical treatment of class II and class III abnormalities with respect to its effects on facial soft tissue morphology.

ACKNOWLEDGEMENTS

Partial results of this work have been presented in MICCAI'99, Cambridge, UK, September 1999 [33].
REFERENCES


Figure 1. Stereo pairs (a) $I_L$ and (b) $I_R$ acquired with texture projection (indicated regions highlight scale distortion), (c) $S_L$ and (d) $S_R$ acquired with structured light projection.

Figure 2. Stereo pairs of plaster cast model of human face bearing 13 simulated landmarks using for estimating reconstruction accuracy.

Figure 3. Point Distribution Model for 2-D feature detection.
Figure 4. *Intensity image used for 2-D feature detection and stereo matching.*

Figure 5. *The vertical integral projection of the horizontal edge. L and R indicate the left and right face borders.*

Figure 6. *An example of ASM-assisted facial feature detection (a) Initialization of the ASM, (b) final position after convergence of search procedure.*
Figure 7. Automatically detected landmarks on a 3-D facial model.

Figure 8. Approximate symmetry plane for a 3-D facial model.

Figure 9. Midline soft tissue profile represented by 39 pseudo-landmarks.
Figure 10. Procrustes scatter of midline profiles for groups I and II. Solid lines indicate the mean profiles.

Figure 11. The deformation of procrustes average midline profile from group I (left) to group II (right) using a thin-plate spline mapping. Mandibular repositioning was achieved by inserting a bite block.
Figure 12. Shape variation explained by first 3 PCs evaluated at a multiple $c$ of the standardized PC scores ($c=-3,-2,-1,0,1,2,3$ corresponding to columns from left to right), row $i$ corresponds to $i$th PC ($i=1,2,3$).

Figure 13. Plot of PC 1 (horizontal axis) vs PC 2 for differences in midline profiles between group I ('o') and group II ('*').
Figure 14. The Procrustes scatter of landmarks for group I and II in the (a) x-y and (b) z-y planes. The red colour indicates the scatter of group I and the green colour is used for group II. Major difference between the groups can be found in mouth region due to the insertion of bite blocks.