Scale-Space Volume Descriptors for Automatic 3D Facial Feature Extraction

Daniel Chen, George Mamic, Clinton Fookes, Sridha Sridharan

Abstract—An automatic method for the extraction of feature points for face based applications is proposed. The system is based upon volumetric feature descriptors, which in this paper has been extended to incorporate scale space. The method is robust to noise and has the ability to extract local and holistic features simultaneously from faces stored in a database. Extracted features are stable over a range of faces, with results indicating that in terms of intra-ID variability, the technique has the ability to outperform manual landmarking.

Keywords—Scale space volume descriptor, feature extraction, 3D facial landmarking

I. Introduction

THE ability to consistently extract features is at the heart of most facial correspondence and recognition algorithms. In particular the construction of statistical models requires point correspondences to exist between different faces. The "ground truth" in correspondence is achieved via manual landmark selection. With large databases this is not only time consuming and tedious, but can also lead to substantial errors being introduced when noise is present ie. spikes caused by reflection of 3D scanner data from the retina.

The anatomical landmarks which are often used in facial correspondence algorithms are:

- Eye corners both inner and outer;
- Nose, including the tip, nostril edges, nasion and subnasal points;
- Chin;
- Lips can consist of lip corners, and also the midpoints of upper and lower lips;
- Glabella

It is important to recognise that landmarks such as the eyes, glabella and nose are reasonably invariant to variations in expression whereas, the chin and lips can vary quite markedly.

This paper presents a method for the extraction of facial feature points in 3D faces. The system is built upon integral based volume descriptors which are robust to noise. This is expanded to a scale based approach allowing accurate detection of features of various sizes, enabling automatic labelling of landmark points.

D. Chen is with the Image and Video Research Laboratory, Queensland University of Technology, Brisbane, QLD 4001, Australia (e-mail: daniel.chen@qut.edu.au).

Dr. G. Mamic is with the Image and Video Research Laboratory, Queensland University of Technology, Brisbane, QLD 4001, Australia.

Dr. C. Fookes is with the Image and Video Research Laboratory, Queensland University of Technology, Brisbane, QLD 4001, Australia (e-mail: c.fookes@qut.edu.au).

Prof. S. Sridharan is with the Image and Video Research Laboratory, Queensland University of Technology, Brisbane, QLD 4001, Australia (e-mail: s.sridharan@qut.edu.au).

The outline of this paper is as follows. Section II provides some background on facial biometrics and how they have been previously extracted and used in large databases. Section III describes feature extraction using volume descriptors, with Section IV showing how to use these features in facial biometric extraction. Section V presents results that were achieved with data extracted from the spring 2003 component of the Facial Recognition Grand Challenge (FRGC) Database [1] and this is followed by conclusions and future directions which are provided in Section VI.

II. Facial Landmark Identification

Facial Landmark algorithms can be broadly categorised into model based and non-model based methods. Hutton [2] showed how a dense surface model of the human face can be built from a database where active shape models (ASMs) are combined with the iterative closest point (ICP) algorithm to fit the model to new faces. The model is built by aligning the surfaces using a sparse set of hand-placed landmarks. Thin plate spline warping is then used for dense correspondence creation with a base mesh. All of the mesh vertices are then used as landmarks to build a 3D point distribution model.

Rueckert *et al.* [3] developed statistical deformation models (SDM) which allow the construction of average models of the anatomy and their variability. SDMs are constructed using the statistical analysis of the deformations required to map anatomical features in one subject into the corresponding features in another subject. A non-rigid registration algorithm is used to compute the deformations required to establish correspondences between the reference subject and the subjects in the population class under investigation. Although the paper presents results using the human brain the technique is easily adapted for faces.

Blanz and Vetter [4] present a method for face recognition across variations in pose and across a wide range of illuminations. To account for these variations, the algorithm fits a statistical, morphable model of 3D faces to images. These morphable face models are built by establishing dense point to point correspondences between the probe and template faces using a modified optical flow algorithm.

Non-model based algorithms attempt to extract features/regions of the face based on the shape of the face and classify these regions based upon the results. Differential geometry has been used by a number of authors. Wang *et al.* [5] determine correspondence between points on pairs of surfaces based on shape using a combination of geodesic distance and surface curvature. An initial sparse set of corresponding

points are generated using a shape based matching procedure. Geodesic interpolation is employed in order to capture the complex surface. In this case the results are applied to human cerebral cortical surfaces however a similar approach could be used for faces.

Brett [6] finds correspondences between two triangulated mesh surface representations. The algorithm produces a matching pair of sparse polyhedral approximations, one for each shape surface, using a global Euclidean measure of similarity. A method of surface patch parameterisation is presented and its use in the interpolation of surfaces for the construction of a merged mean shape with a densely triangulated surface is described, which may be used as a basis for automated landmarking.

This paper presents a non-model based strategy which is based upon the extraction of volume descriptors from a face. The volume descriptors are more robust than curvature based strategies and are very effective at consistently extracting the anatomical landmarks that exist upon a face.

III. Volume Descriptors

Feature extraction algorithms which rely on differential geometry generally require some form of smoothing to remove noisy components which are amplified when derivatives are computed. This then leads to other questions such as the type of smoothing to undertake and of course the magnitude of smoothing that is required in the given conditions.

An alternative is to use integral descriptors. Manay [7] showed that integral invariants have the desirable properties of their differential cousins, such as locality of computation (which allows matching under occlusions) and uniqueness of representation (in the limit), however, they are not as sensitive to noise in the data.

Gelfand *et al*. [8] further developed this work through the integral volume descriptor. A sphere is convolved with a given object along its surface to determine a feature value at every point on the surface of the object. For a shape $\mathbf{P}$ consisting of N points $\mathbf{p_1} \dots \mathbf{p_N}$, this is defined as follows,

$$V_r(\mathbf{p}) = \int_{B_r(\mathbf{p}) \cap S} dx. \tag{1}$$

The integration kernel $B_r(\mathbf{p})$ is a sphere of radius r centred at the point $\mathbf{p}$ and S is the interior of the surface represented by $\mathbf{P}$. The quantity $V_r(\mathbf{p})$ is the volume of the intersection between the sphere $B_r(\mathbf{p})$ and the surface defined by the input mesh. This is illustrated in Figure 1.

This quantity can be calculated efficiently by performing a multiplication of the input shape occupancy voxel grid V_P with the sphere grid V_B in the Fourier domain $V(f) = V_P(f) \times V_B(f)$. The value of the volume descriptor v at each vertex can then be calculated via an inverse Fourier transform.

A point on the object surface is regarded as a feature point depending on its 'uniqueness' compared to other points along the surface. This is done by calculating a histogram over the feature values and taking the points present in the lower occupancy bins. For the experiments performed in this paper, the lower percentile of histogram bins is taken. These feature points form a compact representation of the shape which may then be used for tasks such as registration, correspondence and recognition. For this to be performed consistently over a designated class of shapes it is necessary to be able to consistently extract the same feature points for a surface. This can be done effectively by using scale based volume descriptors.

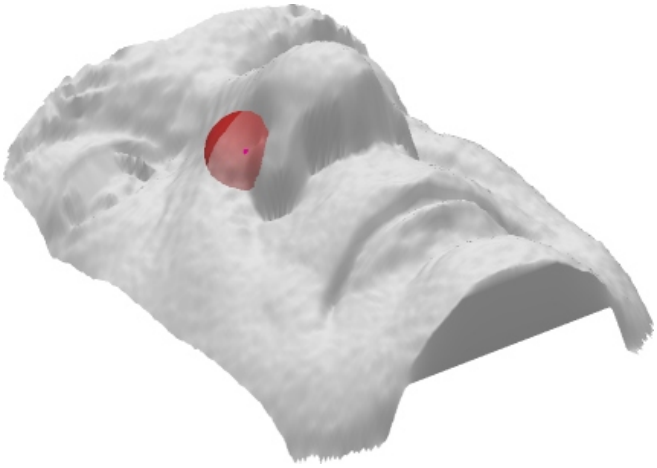

Fig. 1. Integration kernel.

IV. Scale based Volume Descriptors

Face recognition literature is populated with a range of techniques which capture holistic features, local features or a fusion of these features in order to produce the best possible recognition performance. This motivates the development of a scale based volume descriptor for use in faces.

Building upon the volumetric feature descriptor described in Section III, the radii of the spherical kernel used is varied over a range of sizes. Specifically, small scale features are persistent for small radii of the descriptor and large scale features are persistent for the large radii. Empirical tests with faces has indicated that the range of radii for feature extraction lies between $r_{\min} = 7 \times \rho$ and $r_{\max} = 46 \times \rho$ where ρ is the voxel resolution of the face, which is equivalent to 1mm in our test cases. The experiments presented in this paper used ten equi-spaced sphere radii in the range $r_{\min}$ and $r_{\max}$.

This implementation of scale space differs from that of [8] where they deem a point as a 'persistent' feature if it is selected as a feature point over consecutive scales. Their method did not lend well for our application, however, as the feature points extracted failed to line up consistently with landmark locations. We developed a different scale space approach, changing the classification of 'persistent' points, along with the addition of a clustering step. Our novel approach was designed to maximise the consistency in locating facial feature points.

The first stage of the scale based algorithm, is an iterative process where each sphere $B_{r_i}(\mathbf{p})$ with radius $r_{\min} < r_i < r_{\max}$ is passed over the surface and the captured feature points are used to cast votes in an $M \times N$ matrix $\mathbf{V_f}$. The matrix $\mathbf{V_f}$ is identical in size to the X, Y, Z data input matrices.

Once all the votes have been entered over the different scales, we define a persistent feature as being those which

exists over three or more scales at the same location, and use this information to build a map $\mathbf{V_{fp}}$ of persistent features as they exist across the face. By having the features exist over three scales we are able to have both local and holistic features retained in the representation.

The resulting map $\mathbf{V_{fp}}$ is used to 'activate' points on the original surface. These points are then clustered using Euclidean distances with a threshold of 10mm. This figure was derived via empirical tests performed on the face class. The centroids of the resulting clusters form the scale-based volume descriptors which accurately and consistently extract the landmarks of the human face.

The method for the extraction of the scale based volume features is as follows:

Initialise the matrix $\mathbf{V_f}$.
for all radius r_i, **do**
  Extract feature points as described in Section III,
  Cast votes in the appropriate elements of $\mathbf{V_f}$.
end for
Calculate persistent points in $\mathbf{V_f}$ to form $\mathbf{V_{fp}}$.
Use $\mathbf{V_{fp}}$ to index points on the original surface.
Perform clustering on the recovered points and use the centroids of the clusters as features.

The entire feature extraction process is summarised in Figure 2.

V. Experimental Results

This section will examine: the features that are typically extracted on faces; the intra-facial variance that is obtained and how this compares with manual landmarking; and finally the number of features that are extracted at each of the different scales of the spheres that are run across the face. The scale based volume descriptors were tested using the FRGC 1.0 database. The 3D data of the FRGC database contained 640×480 images and our experiments used 943 images of over 300 different people.

Figure 3 presents an example of four faces from the FRGC database with the resulting voting matrix and extracted features from the scale based volume descriptors. The voting matrix is colour mapped according to the number of votes received at each point, ranging from blue (zero votes) to red (10 votes).

In all of the cases, the nose-tip and eye corners accurately match up to 3 of the extracted scale based volume descriptors, with the nostrils and lip corners showing up fairly consistently. These points were earmarked in Section II as crucial landmarks which are required for face based applications. By extracting this number of points, combined with a reasonable correspondence algorithm it is not hard to see how this would form the basis of a registration/active shape model system. Given that this is the intended use of this algorithm the next step is to investigate the reproducibility of the landmark extraction by the volume descriptors.

Using a very simple algorithm, the nose-tip and inner eye corners were identified. The nose is assumed to be the front-most point on the face, with the eyes being the dominant (most populous cluster) point in each of the top two quadrants of the face. This method was effective, with only a few cases where

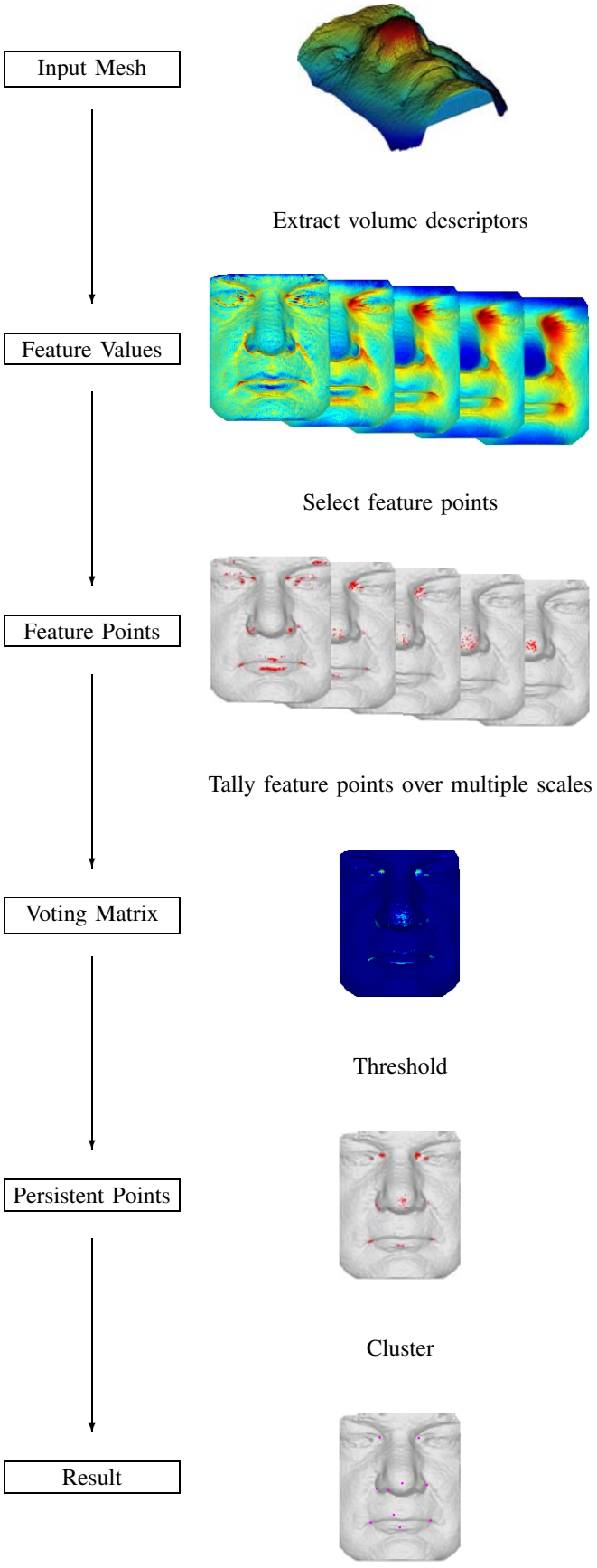

Fig. 2. Feature extraction process.

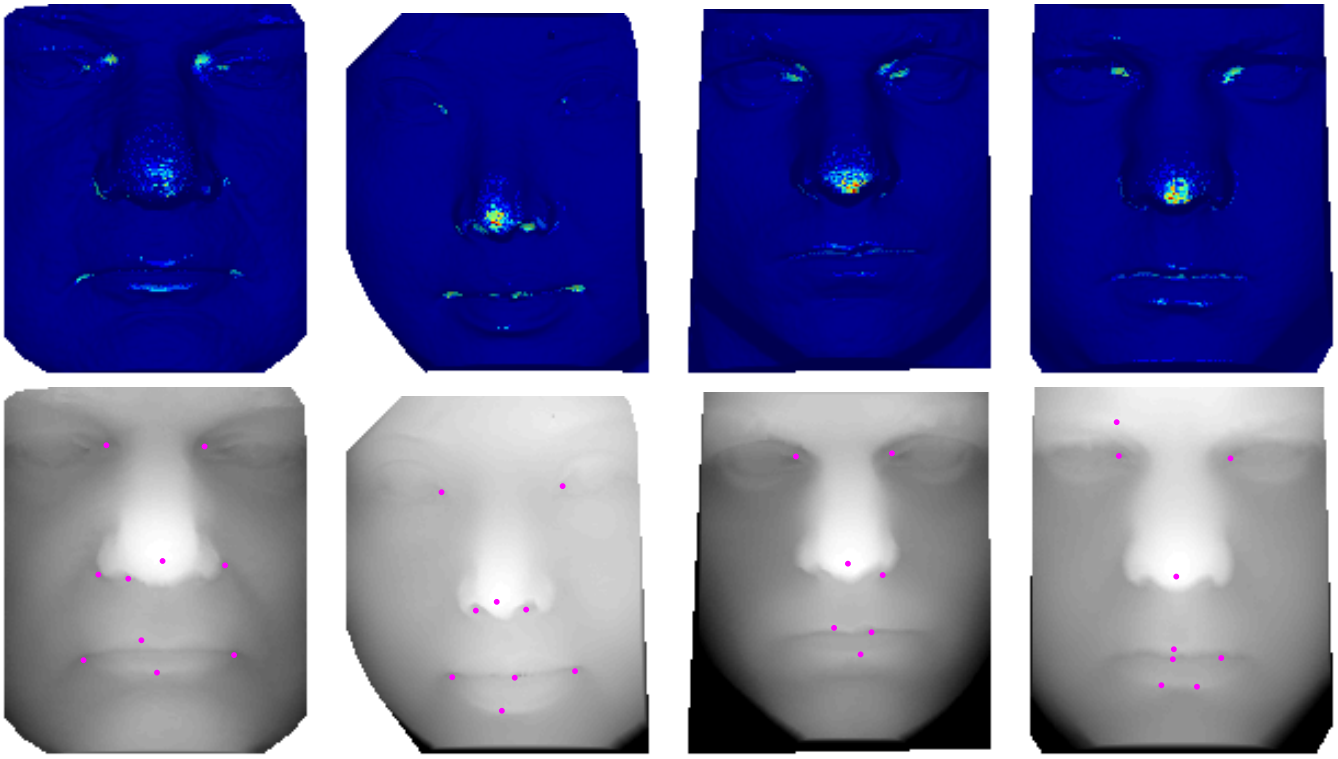

Fig. 3. Examples of the voting matrix (top) and resulting feature centroids (bottom).

the eyes were not correctly identified. In such instances, points on the eyebrows or the outer eye corner were usually identified instead, though some cases did fail due to excessive distortions caused by reflections of the scanning laser on the retina. This could be overcome by applying further pre-processing to the data [9].

A singular vector decomposition (SVD) was performed to align these points over all instances of the same ID (same person). Given X_1 and X_2 are $k \times n$ matrices for k points in n dimensions, and the SVD of $X_1^{\mathrm{T}} X_2$ is USV^{T}, X_1 can be aligned to X_2 by the rotation matrix VU^{T}. Some examples of the alignment can be seen in Figure 4. The plots are coloured such that points belonging to the same person have the same colour.

To quantify the accuracy of the alignment, the variance of each of the three aligned marker locations within the same ID is calculated. Figure 5 presents the histograms of these variance values for the eyes and nose where two or more scans of the same person exists in the database.

The median intra-ID variance for the inner eye corners were 0.5168mm and 0.5599mm, and the nose tip variance were 0.7437mm. This compares favourably to manual selection of these points which has been shown in studies to produce intra-ID variations of the order of 1-2mm [10]. The histogram did produce some outliers due to the incorrect identifications mentioned previously. Having the ability to correctly identify 4 or more points consistently from the front view of a face also means that accurate registration results can be obtained from simple SVD calculations rather than performing the computationally intensive ICP algorithm.

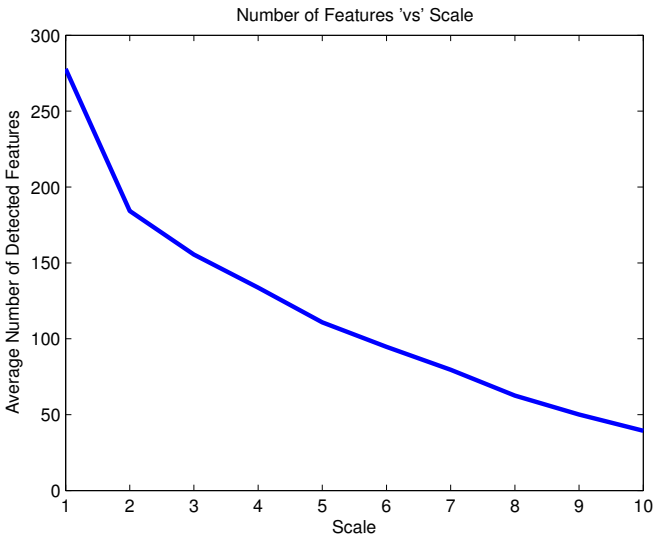

Fig. 7. Number of features extracted at each scale.

Ten scales of radii were used in the experiments presented in this paper. The average number of features extracted at each scale is plotted in Figure 7. An example of the detected features can be seen in Figure 6, using the first face shown in Figure 3.

The results clearly show the need for the scale based extension. At the smaller scales, many features are detected though larger structures, such as the nose, are missed. The

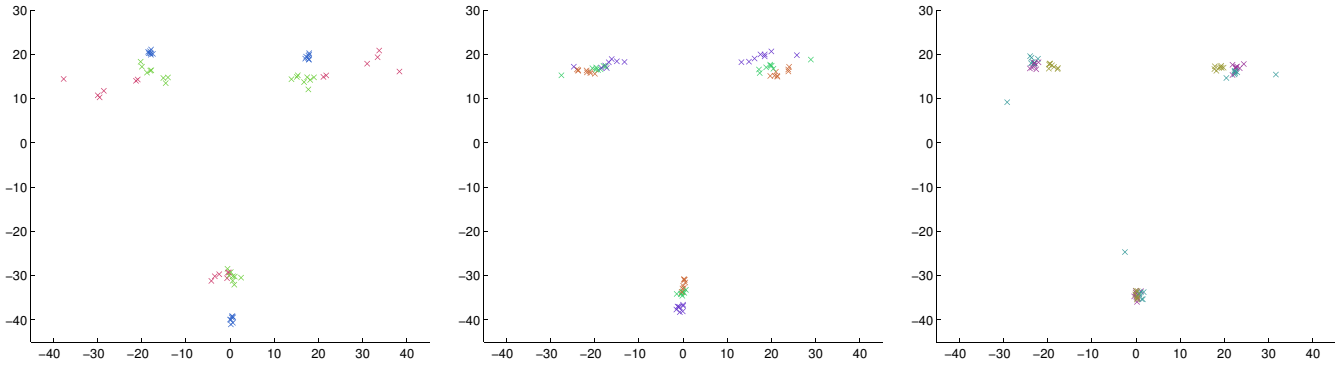

Fig. 4. Example alignments of nose tip and eye corner points.

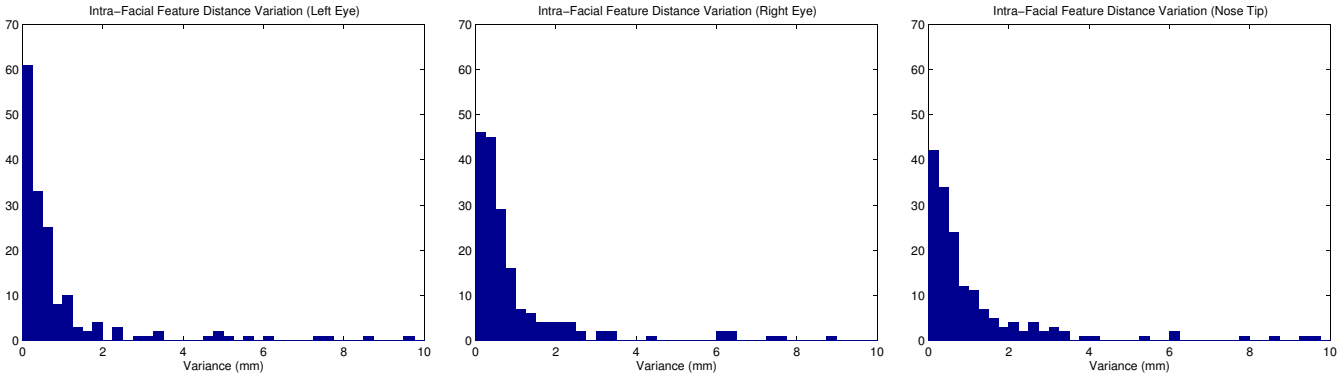

Fig. 5. Histogram of intra-facial feature variance.

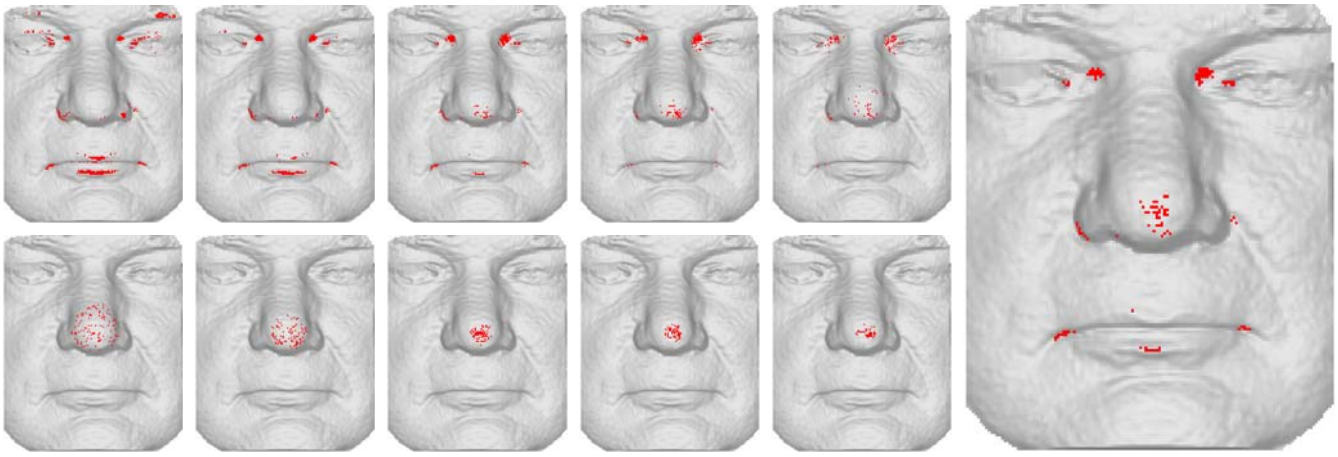

Fig. 6. Detected features at each scale. Kernel size increasing from left to right with scales 1-5 on top row and scales 6-10 on bottom. Persistent features (right).

larger kernel sizes are able to detect the nose though little else. Combining the results of the various scales enables the detection of different sized features. By taking only the persistent features, a clean feature representation of the face is able to be obtained. In these tests, features that are extracted over 3 or more scales are considered persistent.

VI. Conclusion

We have proposed an automatic method for the extraction of feature points for 3D faces. The scale based volume descriptors used are robust to noise and have the ability to extract local and holistic features simultaneously. Extraction of features are stable across multiple instances of the same face, with the variations comparing favourably to manual landmarking.

In the future we plan to develop a scale based volume descriptor driven correspondence algorithm which will have the ability to automatically determine corresponding points across different faces. This can be further enhanced by having a system which has the potential to develop partial correspondences, thus extending practical use to situations where occlusion may be present. Obviously, having the ability to generate such correspondences would also lay the foundation for a recognition system to be built based upon corresponding feature points that may exist between different faces.

References

[1] P. Phillips, P. Flynn, T. Scruggs, K. Bowyer, J. Chang, K. Hoffman, J. Marques, J. Min, and W. Worek, "Overview of the face recognition grand challenge," in *Proceedings IEEE Conference on Computer Vision and Pattern Recognition*, 2005, pp. 947–954.

[2] T. J. Hutton, B. F. Buxton, and P. Hammond, "Dense surface point distribution models of the human face," in *Proceedings IEEE Workshop on Mathematical Methods in Biomedical Image Analysis*, 2001, pp. 153–160.

[3] D. Rueckert, A. F. Frangi, and J. A. Schnabel, "Automatic construction of 3d statistical deformation models using non-rigid registration," *IEEE Transactions on Medical Imaging*, vol. 22, pp. 1014–1025, 2003.

[4] V. Blanz and T. Vetter, "Face recognition based on fitting a 3d morphable model," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 25, pp. 1063–1074, 2003.

[5] Y. Wang, B. S. Peterson, and L. H. Staib, "Shape-based 3d surface correspondence using geodesics and local geometry," in *Proceedings IEEE Conference on Computer Vision and Pattern Recognition*, 2000, pp. 644–651.

[6] A. D. Brett, A. Hill, and C. J. Taylor, "A method of 3d surface correspondence for automated landmark generation," in *British Machine Vision Conference*, 1997, pp. 709–718.

[7] S. Manay, B.-W. Hong, and A. J. Yezzi, "Integral invariant signatures," in *Proceedings 8th European Conference on Computer Vision*, 2004, pp. 87–99.

[8] N. Gelfand, N. J. Mitra, L. J. Guibas, and H. Pottmann, "Robust global registration," in *Proceedings 3rd Eurographics Symposium on Geometry Processing*, 2005, pp. 197–206.

[9] C. Fookes, G. Mamic, C. McCool, and S. Sridharan, "Normalisation and recognition of 3d face data using robust hausdorff metric," in *Proceedings Digital Image Computing: Techniques and Applications*, 2008, pp. 124–129.

[10] F. Steinke, B. Schölkopf, and V. Blanz, "Learning dense 3d correspondence," in *Proceedings 20th Annual Conference on Neural Information Processing Systems*, 2006.