REGION SEGMENTATION AND FEATURE POINT EXTRACTION ON 3D FACES USING A POINT DISTRIBUTION MODEL

Prathap Nair and Andrea Cavallaro

Multimedia and Vision Group – Queen Mary, University of London (United Kingdom)
Email: {prathap.nair,andrea.cavallaro}@elec.qmul.ac.uk

ABSTRACT

We present a novel approach to accurately detect landmarks and segment regions on face meshes without the use of texture, pose or orientation information. The proposed approach is based on a 3D Point Distribution Model (PDM) that is fitted to the region of interest using candidate vertices extracted from low-level feature maps. The robustness of the algorithm is evaluated in the presence of noise and at the variation of the number of scans and model points used in the learning phase. Experimental results demonstrate the accuracy of the proposed method in detecting landmarks, with an improvement of 55% over a state-of-the-art non-statistical approach.

***Index Terms*—** Region segmentation, 3D feature points, landmarks, shape model, feature maps.

1. INTRODUCTION

Facial surface scans are increasingly used in applications such as deformation analysis, animation and face recognition. Many applications require the accurate localization of feature points on the scans. When these feature points correspond to specific anthropometric locations on the human face, they are referred to as *landmarks*. Landmark detection is used for relating vertices from different scans (prior to registration), for generating signatures (for biometrics) and for segmenting regions of interest.

Most existing methods for landmark detection on meshes are dependent on prior knowledge of feature map thresholds, orientation and pose [3, 9, 10]. Deformable models such as Active Shape Models (ASM), Active Appearance Models (AAM) and 3D Morphable Models (3DMM) are extensively used for image segmentation and landmark detection [4, 8]. The shape model used in these approaches, called Point Distribution Model (PDM), aims to perform image interpretation using prior statistical knowledge of the shape to be found. In AAMs texture information is also modeled and associated with the corresponding point locations for model fitting. 3DMM is a concept closely related to AAMs where a 3D model is used to estimate the 3D parameters in a 2D image and to segment objects [1, 8]. Although recently the PDM was adapted for 3D volumetric data [5] and reconstruction of 3D meshes [2], to the best of our knowledge they have not been applied yet for segmentation and landmark detection on polygonal meshes. This is largely due to the fact that PDMs are usually used with the associated texture model. However, texture information is not always available.

In this paper, we propose a robust algorithm for detecting landmarks on face scans (also referred to as face meshes or 3D faces). The proposed algorithm uses a PDM to eliminate the need for prior knowledge of orientation and pose of the scans and relaxes the constraints on feature map thresholding. The PDM represents the shape of the region of interest to be segmented. First, suitable feature maps that highlight the curvature properties of the mesh are extracted. Next these maps are used to isolate candidate inner eye and nose tip vertices. Finally, landmark selection is performed by estimating the transformation between the model and candidate vertices that minimizes the deviation from the mean shape (Fig. 1).

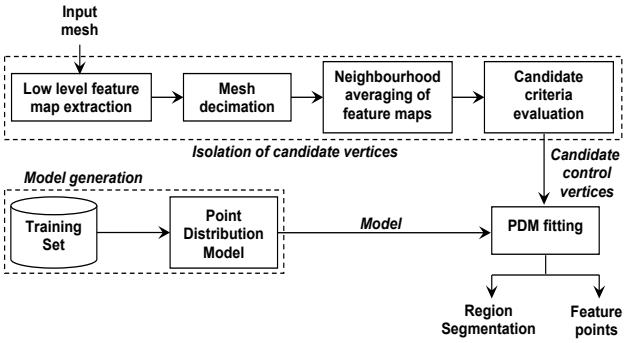


Fig. 1. Block diagram of the proposed algorithm.

The paper is organized as follows: Section 2 describes the processing for the segregation of candidate vertices. The creation of the PDM and the model fitting is described in Section 3. Section 4 discusses the experimental results and validation. Finally, in Section 5 we draw the conclusions.

2. ISOLATION OF THE CANDIDATE VERTICES

The process for the isolation of the candidate vertices is organized in two steps. First, low-level feature maps are extracted that give an indication of the shape and degree of curvature at each vertex of the mesh. Next, candidate inner eye vertices and nose tip vertices are isolated using mesh decimation and neighborhood averaging of the low-level feature maps.

In order to characterize the curvature property of each vertex on the mesh, two features maps are computed, namely the *shape index* and the *curvedness index* [6]. These features maps are derived based on the principal curvature values, $\kappa_1(.)$ and $\kappa_2(.)$, at all the vertices of the surface mesh. The shape index, ρ, at a vertex v_i, is defined as

$$\rho(v_i) = \frac{1}{2} - \frac{1}{\pi}\tan^{-1}\left(\frac{\kappa_1(v_i)+\kappa_2(v_i)}{\kappa_1(v_i)-\kappa_2(v_i)}\right), \qquad (1)$$

where $\kappa_1(v_i) \geq \kappa_2(v_i)$; $\rho(.) \in [0, 1]$. The feature map generated by $\rho(.)$ can describe subtle shape variations from concave to convex thus providing a continuous scale between salient shapes. However, $\rho(.)$ does not give an indication of the scale of curvature present in the shapes. For this reason, an additional feature is introduced,

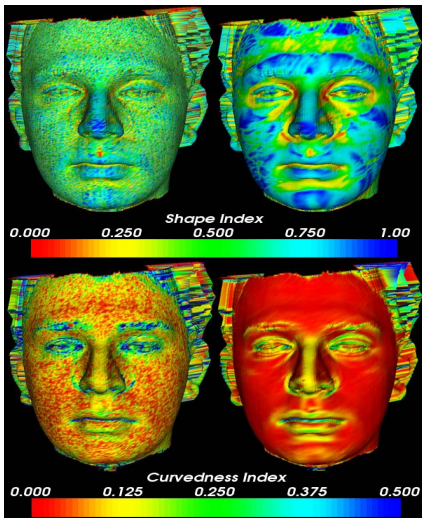


Fig. 2. Comparison of feature maps generated with a non-smoothed surface (left) and smoothed surface (right)

the curvedness of a surface. The curvedness of a surface, $\gamma(.)$, at a vertex v_i, is defined as

$$\gamma(v_i) = \frac{\sqrt{\kappa_1^2(v_i) + \kappa_2^2(v_i)}}{2}. \quad (2)$$

The low level feature maps are computed after Laplacian smoothing that reduce outliers arising from the scanning process. A comparison between the feature maps generated with a smoothed and non-smoothed surface scan is shown in Fig. 2.

To reduce the computational overhead through the reduction of outlier candidate vertices, the original mesh is first decimated. Then the feature maps are averaged across vertex neighbors according to,

$$\tilde{\rho}(v_i) = \frac{1}{P} \sum_{p \in \mathcal{P}(v_i)} \rho(v_p), \quad \tilde{\gamma}(v_i) = \frac{1}{P} \sum_{p \in \mathcal{P}(v_i)} \gamma(v_p) \quad (3)$$

where $\mathcal{P}(v_i)$ is the set of P neighboring vertices of v_i.

If $\tilde{\gamma}(.) > \gamma_s$, then v_i is in a salient high-curvature region. The condition $\tilde{\rho}(.) < \rho_e$ identifies concave regions; while $\tilde{\rho}(.) > \rho_n$ identifies to convex regions. We can therefore segregate by thresholding candidate inner eye vertices from the nose tip ones. The thresholds $\gamma_s = 0.8$, $\rho_e = 0.2$ and $\rho_n = 0.9$ were found to be adequate for the entire database.

Second order neighborhoods for feature averaging and a decimation of 80% was also used. Note that decimation needs to be done after the extraction of the feature maps, otherwise the resulting features would not characterise the original surface. Likewise, the neighborhood averaging of the feature maps is done post decimation. If it is done before decimation, the consistency of features in a neighbourhood would remain and outlier candidate vertices would not be eliminated. Note that the smoothed and decimated mesh is only used for the isolation of the candidate vertices, whereas the original mesh is used for the detection of the landmarks. Examples of scans with candidate vertices isolated are shown in Fig 3; regions in green are candidate nose tip vetices and regions in red are candidate eye tip vertices.

3. MODEL GENERATION AND FITTING

A PDM is used to represent the shape of the region of interest that includes the required landmarks, along with statistical information of the shape variation across the training set. This statistical information is used to test candidate positions for the most plausible fit for the model. With the use of a PDM, we aim to build a parameterized model, $\mathbf{\Omega} = \Upsilon(\mathbf{b})$, where $\mathbf{b}$ is a vector of parameters. To this end, a training set of L face scans were manually landmarked with N points representing the region of interest. Each training shape is a $3 \times N$ element vector, $\mathbf{\Omega} = \{\omega_1, \omega_2, ..., \omega_N\}$, where $\omega_n = (x_n, y_n, z_n)$ represents each landmark.

Training shapes are then aligned to the same co-ordinate frame (registered) so that global transformations are eliminated and statistical analysis is carried out only on shape variations. We use procrustes analysis [7] to align the training shapes to their mutual mean in a least-squares sense, via similarity transformations. This minimizes $\boldsymbol{D}$, the sum of distances of each shape $\mathbf{\Omega^k}$ to the mean $\overline{\mathbf{\Omega}} = \frac{1}{L}\sum_{k=1}^{L} \mathbf{\Omega^k}$, i.e $D = \sum_{i=1}^{N} |\boldsymbol{\omega^k(i)} - \overline{\boldsymbol{\omega}}(i)|^2$ [4]. At each iteration, $\overline{\mathbf{\Omega}}$ is scaled such that $|\overline{\mathbf{\Omega}}| = 1$. Using PCA, the variations of the shape cloud formed by the training shapes in the $(L \times 3 \times N)$ - dimensional space are estimated along the principal axes of the point cloud. The principal axes and corresponding variations are represented by the eigenvectors and eigenvalues obtained from the covariance $\boldsymbol{Z}$ of the data, computed using

$$\mathbf{Z} = \frac{1}{L-1} \sum_{k=1}^{L} (\mathbf{\Omega_k} - \overline{\mathbf{\Omega}})(\mathbf{\Omega_k} - \overline{\mathbf{\Omega}})^T. \quad (4)$$

Let $\boldsymbol{\phi}$ contain the t eigenvectors corresponding to the largest eigenvalues. Then any shape similar to those in the training set can be approximated using

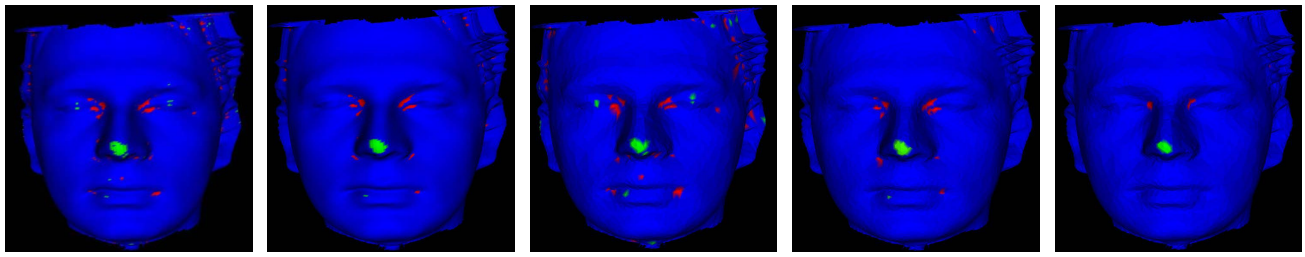
$$\mathbf{\Omega} \approx \overline{\mathbf{\Omega}} + \boldsymbol{\phi b} \quad (5)$$

where $\boldsymbol{\phi} = (\boldsymbol{\phi_1}|\boldsymbol{\phi_2}|\ldots|\boldsymbol{\phi_t})$ and $\mathbf{b}$ is a t dimensional vector given by $b = \phi^T(\omega - \overline{\omega})$. The value of t is chosen such that the model represents 98% of the shape variance, ignoring the rest as noise.The vector $\mathbf{b}$ defines a set of parameters of the deformable model, which are used to vary the shape. The mean shape is obtained when all parameters are set to zero.

The PDM $\mathbf{\Omega}$ is fitted onto a new mesh Ψ_i by performing similarity transformations of the model using three control points of the mean shape, which are the inner eye points ($\boldsymbol{\omega_r}$ and $\boldsymbol{\omega_l}$) and the nose tip point ($\boldsymbol{\omega_f}$), with $\{\boldsymbol{\omega_r}, \boldsymbol{\omega_l}, \boldsymbol{\omega_f}\} \in \mathbf{\Omega}$. Combinations of the candidate inner eye vertices and candidate nose tip vertices on Ψ_i are used as target points to transform the model. Next the remaining points of $\mathbf{\Omega}$ are moved to the closest vertices on Ψ_i. $\mathbf{\Omega}$ is then projected back into the model space and the parameters of the model, $\mathbf{b}$, are updated. Based on this selective search, the transformation exhibiting the minimum deviation from the mean shape is chosen as the fit for the model. The steps of the algorithm are summarized in *Algorithm 1*. Sample snapshots of the evolution of the model with different combinations of candidate vertices are shown in Fig. 4

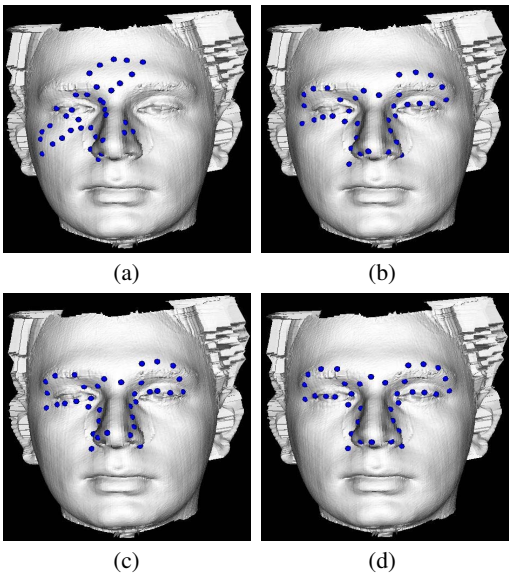
4. EXPERIMENTAL RESULTS

In this section we demonstrate the performance of the proposed algorithm for region segmentation and landmark detection. The landmark detection is evaluated at the variation of the number of scans and number of model points used in learning the model, and with the addition of white noise. A database with 75 face scans was used for



(a) (b) (c) (d) (e)

Fig. 3. Comparison of different strategies for detection of candidate vertices: (a) without averaging and decimation, (b) with averaging and no decimation, (c) with decimation and no averaging, (d) with averaging and then decimation, (e) with decimation and then averaging



(a) (b) (c) (d)

Fig. 4. Example of evolution of the model $\mathbf{\Omega}$ during fitting (a)-(d)

Algorithm 1 Global Fitting

E : Set of candidate eye vertices; F : Set of candidate nose vertices
x : number of candidate eye vertices; y : number of candidate nose vertices
$\breve{C}_{\Psi}(x)$: Closest point to x on Ψ

for $i \leftarrow 1, x$ **do**
  $\alpha_r = E(i)$
  for $j \leftarrow 1, x$ **do**
    $\alpha_l = E(j)$
    for $k \leftarrow 1, y$ **do**
      $\alpha_n = F(k)$
      $Estimate\ T_{\theta,t,s} : min_T \leftarrow D = |\alpha_r - \omega_r|^2 + |\alpha_l - \omega_l|^2 + |\alpha_n - \omega_n|^2$
      $\acute{\mathbf{\Omega}} = T_{\theta,t,s}(\overline{\mathbf{\Omega}})$
      for $p \leftarrow 1, N$ **do**
        $\omega(p) = \breve{C}_{\Psi}(\acute{\omega}(p))$
      end for
      $\tilde{\mathbf{\Omega}} = T^{-1}_{\theta,t,s}(\mathbf{\Omega})$
      $\mathbf{b} = \boldsymbol{\phi}^T(\tilde{\mathbf{\Omega}} - \overline{\mathbf{\Omega}})$
    end for
  end for
end for

Transformation with minimum ν, where $\nu = \sum_i \mathbf{b_i}$ is chosen as best fit

Table 1. Landmark detection accuracy error as a function of N, with $L = 25$. (Key: α_r, right inner eye; α_l, left inner eye; α_{or}, right outer eye; α_{ol}, left outer eye; α_n, nose tip)

N	α_r	α_l	α_{or}	α_{ol}	α_n
5	6%	4%	34%	40%	6%
12	2%	2%	28%	28%	0%
22	2%	2%	26%	22%	0%
32	0%	0%	26%	16%	0%
36	0%	0%	22%	14%	0%

the evaluation, with 25 training scans and 50 test scans, all meshes having roughly 60K vertices.

The PDM is generated by annotating the training set of L scans (with L=10, 15, 20, 25) with N landmarks (with N=5, 12, 22, 32, 36) representing the eyes, eyebrows and nose regions, and including 5 key landmarks, i.e. outer eye points (α_{or}, α_{ol}), inner eye points (α_r, α_l) and nose tip point (α_n).

To evaluate the model fitting, a ground-truth was generated by manually annotating the test set with the 5 key landmarks to measure the error between ground-truth and the corresponding detected landmarks. To account for different head sizes, the error is normalized by the distance between α_r and α_l in each scan. The normalized error for each landmark is cumulated across the test set and used in the final comparison. A detection failure criterion is introduced, wherein if the distance between a landmark and the ground-truth is larger than a certain threshold ($\tau_P = 0.3$), it is deemed to be a failure.

Figure 5 (top) highlights the influence of the size, L, of the training set on the overall performance; whereas Fig. 5 (middle) shows the influence of varying the number of model points N. The percentage of failed detections is shown in Table 1. From these results, we can notice that the accuracy improves with the use of a larger training set L, as more shape variability is captured in the model, without incurring in over-training. An improvement is also seen on increasing the number of model points N, as a better description of the shape of interest is captured. We restricted the number of model points to 36, to reduce complexity in computation and manual annotation. The evaluation of the robustness of the proposed landmark detection method can be seen in Fig. 5 (bottom). The figure shows the influence of additive white noise with variance σ. It can be seen that the algorithm achieves stable detections up to $\sigma = 0.5$.

Finally, Fig. 6 (right) shows a visualization of the fitted model (with N=36) and the corresponding segmented region. A comparison of the proposed method with a state-of-the-art non-statistical method replicating [3, 10] is shown in Fig. 7. The landmark detection results show an overall improvement of 55% in localization accuracy.

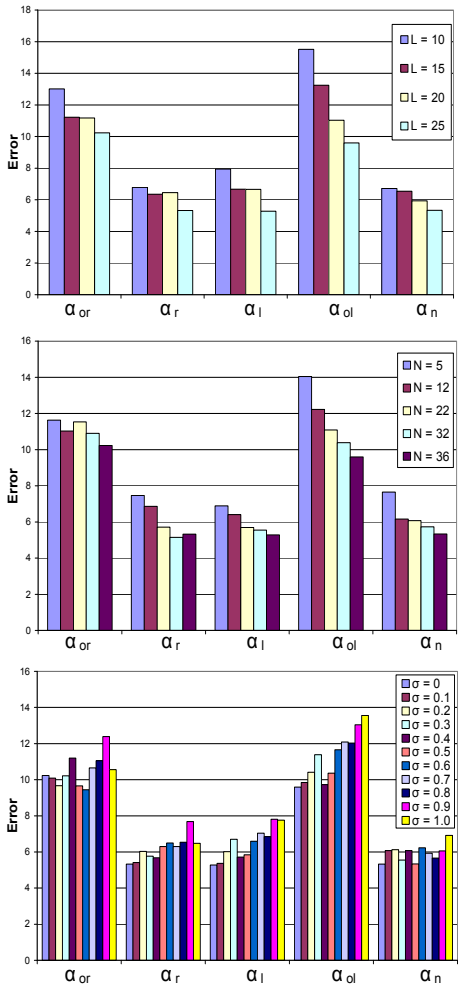


Fig. 5. Normalized distance (error) between automatically detected landmarks and ground-truth landmarks. (Top) comparison with varying L (N=32); (middle) comparison with varying N; (bottom) comparison with additive white noise with variance σ (L=25, N=36)

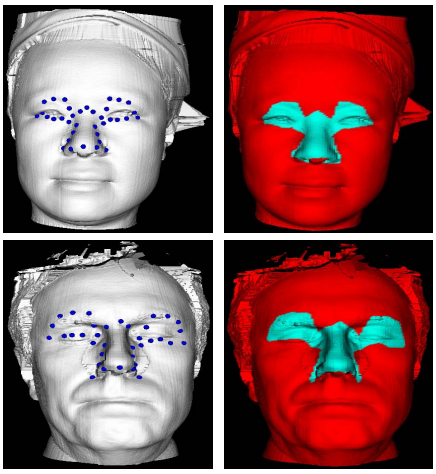


Fig. 6. Examples of landmark detection and region segmentation on facial scans: (left) detected landmarks; (right) region segmentation

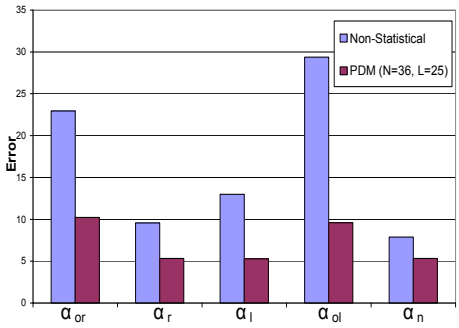


Fig. 7. Comparison of detection accuracy with a non-statistical approach replicating [3, 10], using normalized distance (error) between automatically detected landmarks and ground-truth landmarks.

5. CONCLUSIONS AND FUTURE WORK

We presented a novel method for detecting landmarks on 3D faces that uses a *point distribution model* and removes the need for assumptions on initial orientation and pose. Candidate control vertices are detected and used to fit the shape model by minimizing its deviation from the mean shape. The performance of the proposed approach was evaluated with different parameters, models and number of training samples. The algorithm shows efficiency in the fitting of the model and significant improvement over a non-statistical approach. Current work includes optimization using local neighborhood constraints (post-fitting) and the validation of the proposed landmark detection algorithm in biometrics and deformation analysis.

6. REFERENCES

[1] V. Blanz and T. Vetter. A morphable model for the synthesis of 3D faces. In *Proc. 26th annual conference on Computer graphics and interactive techniques*, pages 187–194, Los Angeles, CA, Aug. 1999.

[2] S. Buchaillard, S.H. Ong, Y. Payan, and K.W.C. Foong. Reconstruction of 3D tooth images. In *Proc. IEEE International Conference on Image Processing*, pages 1077– 1080, Singapore, Oct. 2004.

[3] D. Colbry, G. Stockman, and A. Jain. Detection of anchor points for 3D face verification. In *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, pages 118–125, New York, NY, Jun 2006.

[4] T. F. Cootes, C. J. Taylor, D. H. Cooper, and J. Graham. Active shape models: their training and application. *Computer Vision and Image Understanding*, 61(1):38–59, Jan. 1995.

[5] M. M. Dickens, S. S. Gleason, and H. Sari-Sarraf. Volumetric segmentation via 3D active shape models. In *Proc. IEEE Southwest Symposium on Image Analysis and Interpretation*, pages 248–252, NM, 2002.

[6] C. Dorai and A. K. Jain. Cosmos - a representation scheme for 3D free-form objects. *IEEE Trans. Pattern Anal. Machine Intell.*, 19(10):1115–1130, Oct. 1997.

[7] C. Goodall. Procrustes methods in the statistical analysis of shape. *Journal of the Royal Statistical Society*, 53(2):285–339, 1991.

[8] I. Matthews and S. Baker. Active appearance models revisited. *Int. Journal of Computer Vision*, 60(2):135 – 164, Nov. 2004.

[9] A.B. Moreno, A. Sanchez, J.F. Velez, and F.J. Diaz. Face recognition using 3D surface-extracted descriptors. In *Irish Machine Vision and Image Processing Conference*, Coleraine, Ireland, Sept. 2003.

[10] P. Nair, L. Zou, and A. Cavallaro. Facial scan change detection. In *Proc. European Workshop on the Integration of Knowledge, Semantic and Digital Media Technologies*, pages 77–82, London, UK, Dec. 2005.