

Shape-based Automatic Detection of a Large Number of 3D Facial Landmarks

Syed Zulqarnain Gilani Faisal Shafait Ajmal Mian
School of Computer Science and Software Engineering,
The University of Western Australia

{zulqarnain.gilani, faisal.shafait, ajmal.mian}@uwa.edu.au

Abstract

We present an algorithm for automatic detection of a large number of anthropometric landmarks on 3D faces. Our approach does not use texture and is completely shape based in order to detect landmarks that are morphologically significant. The proposed algorithm evolves level set curves with adaptive geometric speed functions to automatically extract effective seed points for dense correspondence. Correspondences are established by minimizing the bending energy between patches around seed points of given faces to those of a reference face. Given its hierarchical structure, our algorithm is capable of establishing thousands of correspondences between a large number of faces. Finally, a morphable model based on the dense corresponding points is fitted to an unseen query face for transfer of correspondences and hence automatic detection of landmarks. The proposed algorithm can detect any number of pre-defined landmarks including subtle landmarks that are even difficult to detect manually. Extensive experimental comparison on two benchmark databases containing 6,507 scans shows that our algorithm outperforms six state of the art algorithms.

1. Introduction

Due to the non-invasive nature of shape acquisition, 3D morphometric analysis based on facial landmarks is becoming a tool of choice in the fields of anthropometry, human biology and medicine [21]. Scientists have used manually annotated landmarks to ascribe an objective score to the facial gender [17], for syndrome delineation [3, 5, 19, 30], craniofacial analysis for sleep apnoea [2, 14, 24] and many related fields. Scanners are being installed at many hospitals and medical research centers to routinely collect 3D face scans of patients [30]. For example, the Facebase Consortium [23] contains thousands of 3D face scans. Similarly, the Raine database [26] contains hundreds of 3D face scans of individuals with different syndromes as well as controls. Both of these databases are under active development with new scans being constantly added.

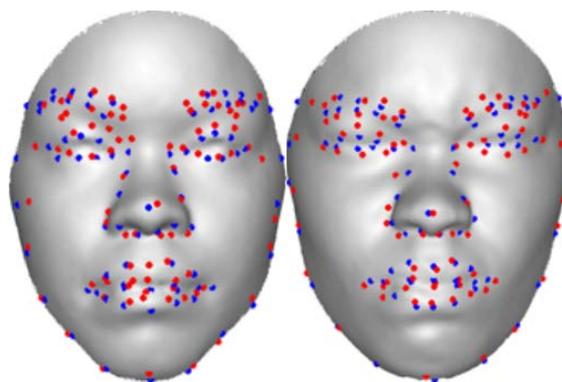


Figure 1. Our algorithm automatically detects an arbitrarily large number of facial landmarks by establishing dense correspondences between 3D faces. The figure shows 85 landmarks detected (red) on neutral and extreme anger expression of a subject from BU3DFE database [43]. The ground truth is represented by blue dots.

A major bottleneck in large-scale 3D facial morphometric analysis is the manual annotation of biologically significant landmarks [15], which is not feasible because of three reasons: 1) it is a laborious and time consuming process, 2) it requires training to master annotation and the annotations may still contain inter and intra operator errors, 3) soft tissue biological landmarks are quite subtle and almost impossible to accurately annotate manually on the 3D scan. These factors are limiting research on huge datasets [23, 26], or where the number of required landmarks is large. Thus, there is an urgent need for a robust technique that is able to automatically detect a large number of biologically significant landmarks.

Current landmarking techniques have focused on a sparse set of facial landmarks and detect these landmarks by exploiting their intrinsic characteristics. Nair and Cavallaro [29] created a Point Distribution Model (PDM) from 49 manually annotated points on 150 scans of BU3DFE [43] database. The statistical model was then fitted on the rest of the faces of the same database to detect these landmarks. Although the method works well in the presence of noisy data, the mean error in the localization of landmarks was

rather high. Furthermore, the model requires an initial manual landmarking of a significant number of points and is not designed to detect any new landmarks. Lu and Jain [25] presented a multimodal approach for facial feature extraction. Using a face landmark model, the authors detected seven facial landmarks on 3D faces using shape index from range images and cornerness from intensity images. The algorithm is specific to this sparse set of landmarks without the possibility of detecting any further points.

Segundo et al. [39] used a curve analysis approach for landmark detection in 3D faces of the FRGCv2 [36] and BU3DFE databases. They extracted features from the mean and Gaussian curvatures for detecting five facial landmarks around the nose and eye (high curvature) regions. Gupta et al. [18] proposed a sequential heuristically driven automatic landmark detection algorithm for face recognition. The authors detected ten anthropometric landmarks on 1,149 3D faces of the Texas 3D Face Recognition Database. Each face is assumed to be in neutral pose and expression. Both these algorithms are dependent on the intrinsic shape properties of a very small set of landmarks.

Creusot et al. [11] presented a machine learning approach to detect 14 corresponding landmarks in 3D faces. They trained an off-line LDA classifier on a set of 200 faces and a landmark model using 16 different local descriptors. The descriptors are curvature based and specific to the 14 landmarks. Each landmark detection is then reduced to a two class classification problem and the final result is fused. The method works well for neutral expression faces of the FRGCv2 and Bosphorus databases. Perakis et al. [32] proposed a method to detect landmarks under large pose variations using a statistical Facial Landmark Model (FLM) for the full face and another two FLMs for each side of the face. A combination of Shape Index and Spin Images is used to find key points on the face. Fiducial landmarks were determined on the basis of minimum combined normalized Procrustes and Spin Image similarity distance from all three FLMs. This method was used to detect eight correspondences in the FRGCv2 database.

We cast the problem of facial landmarking as a sub-problem of 3D dense correspondence. This enables us to detect a large number of fiducial landmarks. The challenge of establishing dense correspondence is to find a mapping of a significant number of points on one surface to their equivalent points on the second surface. The 3D shapes generally have non-linear surface dissimilarities, making sparse correspondence insufficient for landmark localization. The challenge is further compounded by the unavailability of the ground-truth shape correspondence. Dense 3D shape correspondence has many applications in computer graphics, statistical shape analysis and shape phenotyping. In the context of 3D face analysis, dense correspondences have been used for facial morphometric measurements such

as asymmetry for syndrome diagnosis [19, 20], statistical shape modelling [12, 13, 22], shape interpolation [4], non-rigid shape registration [1, 9, 10], deformation analysis [28] and face recognition [7, 16, 31, 37].

Our method proceeds by evolving level set geodesic curves on each 3D facial surface and samples farthest location points [34]. These sample points are then mapped to their corresponding points on a reference 3D face by matching the surface around each point. A cost function defined over bending energy is minimised to select the best corresponding points. Finally, a morphable model based on the dense corresponding points is fitted to an unseen query face for transfer of correspondences. Figure 2 illustrates the complete algorithm. Our algorithm is hierarchical and evolves level set curves with adaptive speeds to detect even the most subtle feature points.

Our key contributions are the following. (1) Unlike existing methods, our algorithm is not designed for detecting specific landmarks that have discriminative geometric properties. Instead, we can detect any number of pre-defined landmarks including subtle landmarks that are difficult to detect manually or with feature matching algorithms. (2) We combine level set curve evolution with geometric speed functions to automatically extract effective seed points for dense correspondence. (3) With a hierarchical structure, our algorithm is capable of extracting thousands of corresponding points on a large set of faces.

We have performed extensive experiments on publicly available FRGCv2 and BU3DFE databases and have compared our results with six state-of-the-art methods. Existing methods have reported results on different number of landmarks, ranging from 6 to 14. Our method has achieved lower mean localization error than the competing methods on all of those landmarks. Furthermore, we report, for the first time, landmark localization on all 83 ground-truth points provided with the BU3DFE dataset.

2. Proposed Algorithm

The block diagram of our proposed algorithm is given in Figure 2 while the individual components are explained below.

2.1. Pre-processing

The nose tip of a 3D face is detected automatically following Mian et al. [27]. Centring a sphere of 90mm at the nose tip, the face is cropped and its pose is corrected to a canonical form by registering the 3D face to a template mask. During registration, the nose tips are aligned after translation and then the 3D face is allowed to rotate only. We observed that this strategy gives a better alignment and ensures that all the 3D faces in the database have the same frontal pose. Figure 3 shows the results of preprocessing on four identities of FRGCv2 dataset.

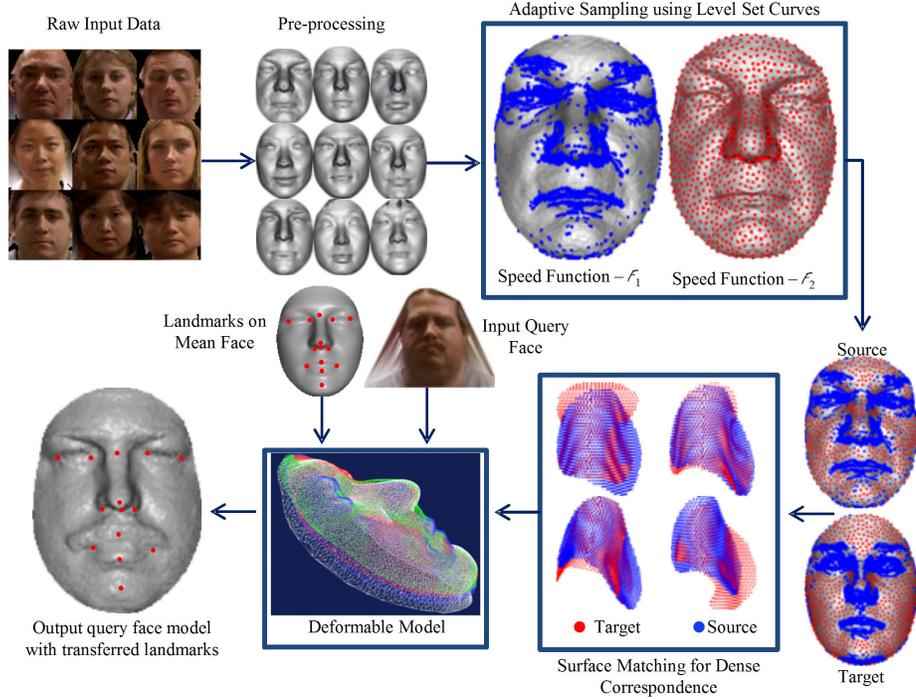


Figure 2. Block diagram of the proposed algorithm. Level set curves are evolved with adaptive speed functions to find seed points on preprocessed face scans. Seed points are matched by minimizing the bending energy of their neighbourhood patches to find dense correspondences. A deformable model is then constructed and fitted to unseen faces to transfer correspondences and hence landmarks.



Figure 3. Preprocessing shown on four identities of FRGCv2 database. The upper row corresponds to the raw 3D image with texture while the lower one is the output of our preprocessing step.

2.2. Adaptive Sampling with Level Set Curves

The point cloud of each 3D face is given as $\mathbf{F}_j = [x_i, y_i, z_i]^T$, where $j = 1, \dots, N$ and $i = 1, \dots, P_j$. The level set interface [40, 41] at point i is represented by $\phi(i) = 0$ and $|\phi(i)|$ gives the shortest distance from this point to the boundary $\phi = 0$. The level set equation is given by,

$$\phi_t + \mathcal{F}|\nabla\phi| = 0 \quad (1)$$

where \mathcal{F} is the propagation speed of the interface front. If the speed function $\mathcal{F} > 0$, (1) can be alternatively formulated by the arrival function γ . The value of $\gamma(i)$ represents the time for the interface to reach a point i from its initial

location, giving the Eikonal equation:

$$|\nabla\gamma|\mathcal{F} = 1 \quad (2)$$

We solve the Eikonal equation numerically using the Fast Marching [34] method by the first order up-winding and discretize it as under:

$$\left[\begin{array}{l} \max(D_{ijk}^{-x}\gamma, 0)^2 + \min(D_{ijk}^{+x}\gamma, 0)^2 \\ + \max(D_{ijk}^{-y}\gamma, 0)^2 + \min(D_{ijk}^{+y}\gamma, 0)^2 \\ + \max(D_{ijk}^{-z}\gamma, 0)^2 + \min(D_{ijk}^{+z}\gamma, 0)^2 \end{array} \right]^{\frac{1}{2}} = \frac{1}{\mathcal{F}_{ijk}} \quad (3)$$

where D^- is the backward difference operator in the given direction, D^+ is the forward difference operator while i, j and k are the basis on an orthogonal grid.

The speed function \mathcal{F} sets the speed of propagation of the front and hence the density of point sampling. We perform adaptive sampling of the 3D face \mathbf{F}_j by using two separate speed functions; a curvature dependent and a uniform function. The mean curvature H is given by [35],

$$H = \nabla \cdot \frac{\nabla\phi}{|\nabla\phi|} = \frac{\{(\phi_{yy} + \phi_{zz})\phi_x^2 + (\phi_{xx} + \phi_{zz})\phi_y^2 + (\phi_{xx} + \phi_{yy})\phi_z^2 - 2\phi_x\phi_y\phi_{xy} - 2\phi_x\phi_z\phi_{xz} - 2\phi_y\phi_z\phi_{yz}\}}{(\phi_x^2 + \phi_y^2 + \phi_z^2)^{\frac{3}{2}}}$$

while the Gaussian curvature K is,

$$K = \frac{\left\{ \begin{array}{l} \phi_x^2(\phi_{yy}\phi_{zz} - \phi_{yz}^2) + \phi_y^2(\phi_{xx}\phi_{zz} - \phi_{xz}^2) + \phi_z^2(\phi_{xx}\phi_{yy} - \phi_{xy}^2) \\ + 2[\phi_x\phi_y(\phi_{xz}\phi_{yz} - \phi_{xy}\phi_{zz}) + \phi_y\phi_z(\phi_{xy}\phi_{xz} - \phi_{yz}\phi_{xx}) \\ + \phi_x\phi_z(\phi_{xy}\phi_{yz} - \phi_{xz}\phi_{yy})] \end{array} \right\}}{(\phi_x^2 + \phi_y^2 + \phi_z^2)^2}$$

The two principle curvatures at each point i are related to the Mean and Gaussian curvatures such that $\kappa_1 = H_i + \sqrt{H_i^2 - K_i}$ and $\kappa_2 = H_i - \sqrt{H_i^2 - K_i}$. The Curvedness at each point is a function of the two principles curvatures and is defined as:

$$C = \sqrt{\frac{\kappa_1^2 + \kappa_2^2}{2}} \quad (4)$$

We use the Curvedness property of the points on 3D face and define the curvature dependent speed function as under,

$$\mathcal{F}_1 = \begin{cases} 1 & \text{for } C > (\bar{C} + \sigma_C) \\ 0 & \text{otherwise.} \end{cases} \quad (5)$$

The speed function \mathcal{F}_1 slows down the propagation of the front in areas of high curvature (Figure 4 row-1). In order to sample 3D points from quasi-planar regions, we use the uniform speed function, \mathcal{F}_2 and set its value to 1 at every point on the 3D face. Consequently, we evolve two level set curves with \mathcal{F}_1 and \mathcal{F}_2 and perform farthest point adaptive sampling [33] of each 3D face \mathbf{F}_j to obtain $\widetilde{\mathbf{F}}_j^1 = [x_k, y_k, z_k]^T$ ($k = 1, \dots, P^{\mathcal{F}_1}$) and $\widetilde{\mathbf{F}}_j^2 = [x_l, y_l, z_l]^T$ ($l = 1, \dots, P^{\mathcal{F}_2}$) pertaining to the two speed functions \mathcal{F}_1 and \mathcal{F}_2 respectively. Note that $P^{\mathcal{F}_1}$ and $P^{\mathcal{F}_2}$ are user defined number of sampled points, $(P^{\mathcal{F}_1}, P^{\mathcal{F}_2}) < \min(P_j)$ and $(\widetilde{\mathbf{F}}_j^1, \widetilde{\mathbf{F}}_j^2) \in \mathbf{F}_j$. This means that the desired number of corresponding points can be controlled. In fact, by increasing the value of $P^{\mathcal{F}_1}$, we can choose to detect more correspondences in high curvature regions and by increasing the value of $P^{\mathcal{F}_2}$, we will detect more correspondences in low curvature regions.

The 3D points $\widetilde{\mathbf{F}}_j^2$ sampled using the uniform speed function \mathcal{F}_2 cover the entire 3D face uniformly (see Figure 4 row-2) which include the discriminative points with high curvature as well as the non-discriminative ones that lie on planar surfaces. Since, $\widetilde{\mathbf{F}}_j^1$ already covers the highly discriminative 3D points to be used later for surface matching (discussed in Section 2.3), we aim to to exclude them from $\widetilde{\mathbf{F}}_j^2$. This is done by removing all points pertaining to $\widetilde{\mathbf{F}}_j^1$ from each 3D face $\widetilde{\mathbf{F}}_j^2$, such that $\widetilde{\mathbf{F}}_j^2 = \widetilde{\mathbf{F}}_j^2 - (\widetilde{\mathbf{F}}_j^2 \cap \widetilde{\mathbf{F}}_j^1)$. Figure 4 shows the adaptively sampled points on the pre-processed faces of five identities of FRGCv2. The top row shows points pertaining to $\widetilde{\mathbf{F}}_j^1$, the middle row shows the points for $\widetilde{\mathbf{F}}_j^2$ while the last row depicts the points for the reduced $\widetilde{\mathbf{F}}_j^2$.

2.3. Coarse to Fine Surface Matching for Dense Correspondence

The adaptively sampled points $\widetilde{\mathbf{F}}_j^1$ and $\widetilde{\mathbf{F}}_j^2$ are not in correspondence yet. The goal of coarse surface matching is to

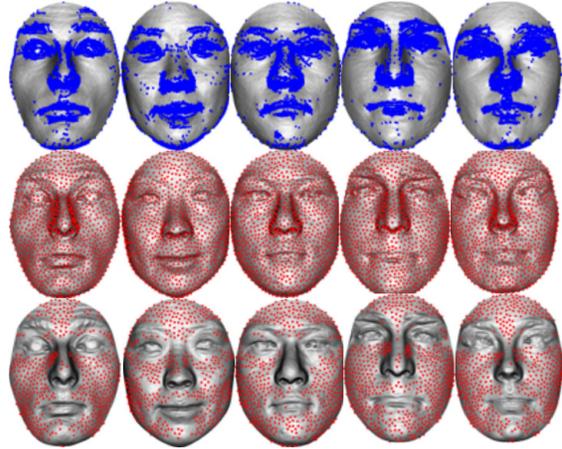


Figure 4. Adaptively sampled points on the pre-processed faces of five identities of FRGCv2. The top row shows points pertaining to $\widetilde{\mathbf{F}}_j^1$, the middle row shows the points for $\widetilde{\mathbf{F}}_j^2$ while the last row depicts the points for the reduced $\widetilde{\mathbf{F}}_j^2$.

find a reasonable mapping of points between the 3D faces $\widetilde{\mathbf{F}}_j^1$. We first select one of the faces from the training set as the source and find its correspondence to the rest of the database in a pair wise scheme. We start by cropping a surface \mathbf{S}_s of radius r_p around each point p on the sampled source face. Note that while p is selected from the sampled face $\widetilde{\mathbf{F}}_j^1$, the surface is cropped from the pre-processed input face \mathbf{F}_j . This ensures a rich sampling of points in the surface.

The task at hand now is to find the correspondence between a point p on source face to a point q on a given target face. This can be done by matching the surface \mathbf{S}_s with surfaces extracted around every point in the target. However, this step would be both redundant and computationally expensive. Since all faces in the database are centred at their nose tips, it can be assumed that the point q on a target face corresponding to the point p on source face can be found on the target face within a small region \mathbf{R} around p . We empirically choose a conservatively large radius of $4cm$ (roughly a quarter of the 3D face) for the region \mathbf{R} and we select all sampled points on the target face within this region. We then crop out surfaces \mathbf{S}_t of radius r_p around each point in \mathbf{R} . Next we measure the non-rigid shape difference $\Delta(s, t)$ between the source surface \mathbf{S}_s and each target surface \mathbf{S}_t in \mathbf{R} . The shape difference is given by;

$$\Delta(s, t) = \frac{\beta_{st} + \beta_{ts}}{2} \quad (6)$$

Here, β_{st} is the amount of bending energy required to deform \mathbf{S}_s to \mathbf{S}_t and is measured using the 2D thin-plate spline model [8]. The bending energy can be calculated as,

$$\beta(s, t) = \mathbf{x}^T \Omega \mathbf{x} + \mathbf{y}^T \Omega \mathbf{y} + \mathbf{z}^T \Omega \mathbf{z} \quad (7)$$

where \mathbf{x} , \mathbf{y} and \mathbf{z} are the vectors containing the x , y and z coordinates of m points in the target surface \mathbf{S}_t and Ω is the

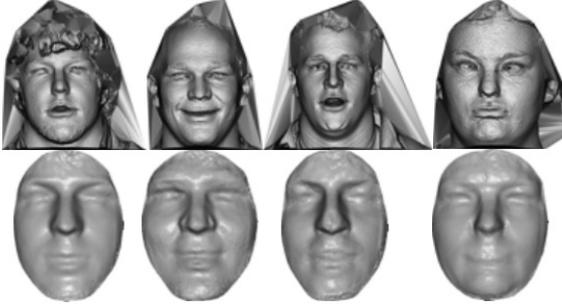


Figure 5. 3D deformable model fitting to four identities from the FRGCv2 database. First row contains the original scans while the second row contains the fitted models.

bending matrix. It is defined as the $m \times m$ upper left matrix of $\begin{bmatrix} \mathbf{K} & \mathbf{B} \\ \mathbf{B}^T & \mathbf{0} \end{bmatrix}^{-1}$.

Here $\mathbf{B} = [\mathbf{1}, \hat{\mathbf{x}}, \hat{\mathbf{y}}, \hat{\mathbf{z}}]$, where $\hat{\mathbf{x}}, \hat{\mathbf{y}}$ and $\hat{\mathbf{z}}$ are the x, y and z coordinates of m points on the source surface \mathbf{S}_s , $K(i, j) = \|\mathbf{s}_s^i - \mathbf{s}_s^j\|^2 \log\|\mathbf{s}_s^i - \mathbf{s}_s^j\|$ with $i, j = 1, \dots, m$ and $\mathbf{0}$ is an $m \times 4$ matrix of zeros. Since, $\beta(s, t)$ and $\beta(t, s)$ are not equal, their average gives the non-rigid shape difference between the two surfaces.

We say that a point q within \mathbf{R} coarsely corresponds to point p in \mathbf{S}_s if the shape difference $\Delta(s, t)$ in their corresponding patches is minimum. To find a finer correspondence between the two points we repeat the surface matching process between the point p on source face and all points within radius r_q of the coarse corresponding point q on the pre-processed target face. We have empirically found $r_q = 5mm$ to be a good choice for this search space. Correspondence between all 3D faces $\widetilde{\mathbf{F}}_j^1$ are established by repeating the coarse to fine surface matching process for all points $P^{\mathcal{F}1}$. The output of this step is a set of corresponding faces $\widetilde{\mathbf{F}}_j^c = [x_p, y_p, z_p]^T$, where $j = 1, \dots, N$ and $p = 1, \dots, P^{\mathcal{F}1}$.

The coarse to fine surface matching establishes correspondence between the 3D faces in regions of high curvature. We now present a hierarchical scheme to propagate this correspondence further and establish dense correspondence between the 3D face using the uniformly sampled non-descriptive points $\widetilde{\mathbf{F}}_j^2$. We first select a source and a target face and register them using the corresponding points $\widetilde{\mathbf{F}}_j^c$. The same transformation is used to register the $\widetilde{\mathbf{F}}_j^2$ points of source face to those of the target. Next, all points in $\widetilde{\mathbf{F}}_j^2$ on the two faces in the pair are projected on a $2D$ grid and the correspondences are transferred from the source to the target face. This ensures speedy propagation of correspondences with minimum localization error. These new correspondences are updated in $\widetilde{\mathbf{F}}_j^c$ to give $\mathbf{F}_j^c = [x_p, y_p, z_p]^T$, where $j = 1, \dots, N$ and $p = 1, \dots, P$.

2.4. Model Fitting and Landmark Localization

The output of surface matching step is a set of N densely corresponding faces, \mathbf{F}_j^c which form a 3D Deformable Model (3DM), $\Psi = [\mathbf{f}_1^c, \mathbf{f}_2^c, \dots, \mathbf{f}_N^c]$, where $\mathbf{f}^c = [x_1, \dots, x_p, y_1, \dots, y_p, z_1, \dots, z_p]^T$ and $p = 1, \dots, P$. The row mean μ_Ψ of the 3DM is given by,

$$\mu_\Psi = \frac{1}{N} \sum \mathbf{f}_i^c \quad (8)$$

Without loss of generality, assume that the row means of Ψ_m are all 0. The 3DM Ψ_m can be modelled by a multivariate Gaussian distribution and its eigenvalue decomposition is given by,

$$\mathbf{U}\mathbf{S}\mathbf{V}^T = \Psi_m \quad (9)$$

where $\mathbf{U}\mathbf{S}$ are the principal components (PCs), and the columns of \mathbf{V} are their corresponding loadings. \mathbf{S} is a diagonal matrix of eigenvalues and we retain 98% of the energy corresponding to the first n columns of \mathbf{U} . Furthermore, the mean face of the 3DM is given by $\overline{\mathbf{F}}^c = \frac{1}{N} \sum_{j=1}^N \mathbf{F}_j^c$.

We deform the statistical model given in (9) into a query face \mathbf{Q} which is first registered and roughly corresponded to the mean face $\overline{\mathbf{F}}^c$ by searching for the Nearest Neighbour (NN) of each point of $\overline{\mathbf{F}}^c$ in \mathbf{Q} using the k-d tree data structure [6]. After vectorization, the query face can be parametrized by the statistical model such that $\mathbf{m}_q = \mathbf{U}\alpha + \mu_\Psi$, where the vector α contains the parameters which are used to vary the shape of the model and \mathbf{m}_q is the vectorized form of the query model \mathbf{M}_q generated by the 3DM. The vector α is given by,

$$\alpha = \mathbf{U}^T(\mathbf{q}^c - \mu_\Psi) \quad (10)$$

where \mathbf{q}^c is the vectorized corresponded query face.

Finally, dense correspondence is propagated to the query face \mathbf{Q} by registering it to \mathbf{M}_q . This is done by minimizing the distance between NN points (rigid transformation), then deforming the model using Equation (10) (non-rigid morphing) and a final NN search to transfer the correspondences. We manually annotate an arbitrary set of landmarks on $\overline{\mathbf{F}}^c$ and transfer them to the query model \mathbf{M}_q since the point clouds of both 3D faces correspond to each other. Note that the manual annotation is done only once and that too on $\overline{\mathbf{F}}^c$ only. Figure 5 shows the 3DM fitted to four identities of the FRGCv2 database.

3. Results and Analysis

We have performed extensive landmarking experiments on FRGCv2 [36] and BU3DFE [43] databases because they are readily available, have considerable variation in ethnicity, age, pose and expression and also because benchmarking is possible due to the availability of ground truth landmarks. FRGCv2 contains 4,007 scans of 466 individuals. Manual annotations provided by Szeptycki et al. [42] and Creusot et al. [11] were used as ground truth for comparison. The BU3DFE dataset comprises of 2,500 scans of 100

Table 1. Comparison of the mean and standard deviation of localization error ϵ_L (in mm) with the state-of-the-art algorithms on FRGCv2 dataset. A ‘-’ denotes that the authors did not detect this particular landmark.

Author	Lu et al. [25]		Segundo et al. [39]		Perakis et al. [32]		Creusot et al. [11]		This Paper	
#Images	676		4007		975		4007		4007	
ϵ_L	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Ex(L)	9.50	17.10	-	-	5.58	3.33	5.87	3.11	4.50	2.97
En(L)	8.30	17.20	3.69	2.26	4.15	2.35	4.31	2.44	3.12	2.09
N	-	-	-	-	-	-	4.20	2.07	3.63	2.02
Ex(R)	10.30	18.10	-	-	5.83	3.42	6.00	3.03	3.74	2.79
En(R)	8.20	17.20	3.35	2.33	4.41	2.49	4.29	2.03	2.73	2.14
Prn	8.30	19.40	2.73	1.39	4.09	2.41	3.35	2.00	2.68	1.48
Ac(L)	-	-	4.83	2.03	-	-	4.73	3.68	3.66	2.18
Ac(R)	-	-	5.84	1.73	-	-	4.86	3.54	3.43	1.80
Ch(L)	6.00	16.90	-	-	5.56	3.93	5.47	3.45	5.31	2.05
Ch(R)	6.20	17.90	-	-	5.42	3.84	5.64	3.58	4.38	2.08
Ls	-	-	-	-	-	-	4.23	3.21	3.31	2.65
Li	-	-	-	-	-	-	5.46	3.92	4.02	3.80
Pg	-	-	-	-	4.92	3.74	7.28	7.41	3.95	3.27
Sn	-	-	-	-	-	-	3.67	3.11	4.05	3.08
Ps(R)	-	-	-	-	-	-	-	-	4.01	2.71
Pi(R)	-	-	-	-	-	-	-	-	3.10	1.97
Ps(L)	-	-	-	-	-	-	-	-	3.85	2.68
Pi(L)	-	-	-	-	-	-	-	-	2.98	2.18
Mean	8.11	17.69	4.09	1.95	5.00	3.19	4.95	3.33	3.69	2.44

Table 2. Comparative results of our proposed algorithm with SISI-NPSS [32] reported on the identities defined in DB00F [32]. Our dense correspondence algorithm is able to localize all landmarks with a lower error.

Expression	Neutral (n=443)				Mild (n=355)				Extreme (n=177)				All (n=975)			
	SISI-NPSS		This Paper		SISI-NPSS		This Paper		SISI-NPSS		This Paper		SISI-NPSS		This Paper	
ϵ_L	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Ex(R)	5.38	3.14	3.80	2.78	5.76	3.42	4.43	3.31	5.71	3.57	4.19	3.01	5.58	3.33	4.10	3.04
EN(R)	3.95	2.19	3.24	2.16	4.28	2.35	3.04	2.25	4.38	2.68	3.60	2.63	4.15	2.35	3.23	2.29
En(L)	4.37	2.51	2.61	1.78	4.48	2.33	2.76	1.98	4.40	2.74	3.13	2.41	4.41	2.49	2.76	1.99
Ex(R)	5.66	3.37	3.54	2.69	5.95	3.38	3.70	2.89	6.02	3.59	3.56	2.58	5.83	3.42	3.60	2.74
Prn	3.99	2.24	2.37	1.98	3.92	2.06	2.75	2.48	4.67	3.25	3.25	3.32	4.09	2.41	2.67	2.47
Ch(R)	4.25	2.30	4.08	2.33	5.36	3.10	5.28	2.96	9.26	5.88	7.39	4.74	5.56	3.93	5.12	3.34
Ch(L)	4.35	2.40	3.43	2.31	5.21	3.14	4.21	3.07	8.55	5.87	6.41	3.87	5.42	3.84	4.25	3.12
Pg	4.21	2.36	3.97	2.38	4.66	2.70	4.32	2.61	7.27	6.45	4.19	3.01	4.92	3.74	4.10	3.04
Mean	4.52	1.51	3.38	0.94	4.95	1.46	3.81	1.28	6.28	2.60	4.46	1.77	5.00	1.85	3.72	1.15

individuals. Each subject has been imaged in neutral and six different expressions each having four intensity levels. The dataset comes with a ground truth of 83 manually annotated corresponding points. The ground truth is devoid of Nasion (N) and the Pronasale (Prn). We have manually added them to the ground truth bringing the total to 85.

Using our presented algorithm, it is possible to establish dense correspondence on all 4,007 scans of FRGCv2 database, in a pair wise manner. However, we constructed a dense morphable model from the first available neutral scan of the first 200 identities of FRGCv2. This model was then fitted to all the scans of the remaining 266 identities (1,961 in total) in order to test the quality of dense correspondence and the fitting procedure. To perform experiments on the rest of the database, we constructed a separate model from the neutral scans of the next 200 identities and used all the scans (2,046) corresponding to the first training

set as probes. In this way we report Landmark Detection results on all 4,007 scans of FRGCv2, ensuring that the identities in the gallery and probe sets are disjoint.

We annotated 18 biologically significant anthropometric landmarks [15] on the mean face \bar{F}^c of both models. These landmarks were then projected back on each 3D face and we report the mean of landmark localization error (ϵ_L) in millimetres (mm). Table 1 shows the mean and standard deviation of ϵ_L over 18 landmarks as well as a comparison with the state-of-the-art algorithms on the same dataset. Note that Creusot et al. [11] tested their algorithm on 4,750 images and provide results for each individual. For comparison, we select the same 4,007 faces that were used in this paper. It is evident from the results that our proposed algorithm performs better than its counterparts. A histogram of error distribution, separately for the neutral and non-neutral scans of FRGCv2, is shown in Figure 6.

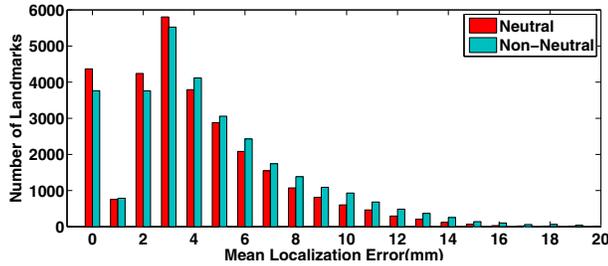


Figure 6. Histogram of mean localization error for 18 landmarks on 4,007 scans of FRGCv2 dataset (18×4007 Landmarks).

Table 3. Comparison of landmark localization results with the state-of-the-art on BU3DFE dataset.

Author	Nair et al. [29]	Segundo et al. [39]	Salazar et al. [38]	This Paper
#Images	2350	2500	350	2500
ϵ_L	Mean	Mean SD	Mean SD	Mean SD
Ex(L)	20.46	-	9.63 6.12	4.43 2.74
En(L)	12.11	6.33 4.82	6.75 4.54	4.75 2.64
N	-	-	-	3.50 2.70
Ex(R)	19.38	6.33 5.04	8.49 5.82	4.35 2.70
En(R)	11.89	-	6.14 4.21	3.29 2.67
Prn	8.83	1.87 1.12	5.87 2.70	2.91 2.03
Ac(L)	-	6.66 3.36	6.47 3.30	4.30 2.73
Ac(R)	-	6.49 3.40	7.17 3.07	4.28 2.71
Sbal(L)	-	-	-	4.86 2.80
Sbal(R)	-	-	-	3.57 2.59
Sn	-	-	-	3.90 3.26
Ch(L)	-	-	-	6.00 3.94
Ch(R)	-	-	-	5.45 3.12
Ls	-	-	-	3.20 2.68
Li	-	-	-	6.90 5.31
Mean	14.53	5.54 3.55	7.22 4.25	4.38 2.98

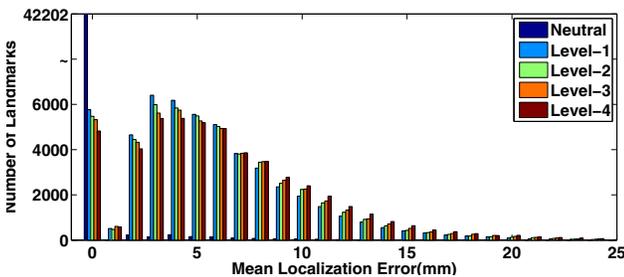


Figure 7. Error distribution histogram of 85 landmarks on 2,500 scans of BU3DFE dataset (83×2500 Landmarks). Notice that the mean error for majority of landmarks on neutral expression is less than $2mm$.

Perakis et al. [32] reported their landmarking results using the SISI-NPSS algorithm on neutral, mild and extreme expression scans of FRGCv2. Since their experimental protocol is available on-line, we have compared our results with them on matching identities and show the comparison in Table 2. Note that the authors used 300 scans with varying expressions to train their models whereas we have used only 200 neutral expression scans to generate the dense correspondence model. Our mean localization error for the eight selected landmarks is $3.72mm$ in contrast to $5mm$ as reported by [32].

From a practical perspective, it may not always be possi-



Figure 8. Effects of expression on landmark localization. The six expressions of BU3DFE dataset are shown in the columns while the rows depict the increasing intensity levels. The error distribution of 12 landmarks for the four intensity levels is shown in the last row. The error is maximum in expressions which involve gap in the lips.

ble to generate dense correspondence models from the data set available at hand. In such scenarios, it would be beneficial to fit an existing dense correspondence model to the available query faces. The outcome of landmark detection in this case would depend on the quality of dense correspondence of the original model. We have used the BU3DFE database to test this feature of our dense correspondence model. The model is created from the first available neutral scan of the first 200 identities of FRGCv2 and fitted to each 3D face from the BU3DFE dataset. Note that the FRGCv2 images were acquired from the Minolta™ scanner while the BU3DFE images were acquired using the 3DMDFace™ system.

As mentioned earlier, 85 manually annotated ground truth points are available for the BU3DFE dataset. We annotated the 85 points on the mean face of our morphable model and projected them back on each 3D face to calculate the mean localization error. Our proposed algorithm is able to localize the 85 ground truth points on the complete dataset within $5.85 \pm 4.26mm$. The mean and standard deviation of localization error of each point is given in the accompanied supplementary material. The histogram of error distribution for each intensity level of the six facial expressions is shown in Figure 7. The 85 landmarks annotated on one of the subjects in extreme anger expression are depicted in Figure 1. In order to compare our results with the state-of-the-art, we report the landmark localization results on 15 anthropometric landmarks, in Table 3. To the best



Figure 9. Thirty five anthropometric landmarks [15] automatically detected on seven identities of the FRGCv2 database. The top left face gives the ground truth locations and names of these landmarks. Notice, the consistency and accuracy of detection.

of our knowledge, this is the highest number of automatically detected anthropometric landmarks reported so far in the literature.

The existing algorithms are landmark specific and mostly driven by the intrinsic properties of these landmarks. It is evident from the results that the landmarks which do not display sharp curvature changes have a larger localization error. In contrast, our proposed algorithm is not tuned to a specific set of landmarks and this factor has contributed to the superior quality of our results. Furthermore, since we build a dense correspondence model over the complete 3D face, we are able to detect any number of landmarks specific to the needs of the concerned research and application community.

Another strength of our proposed approach is that we present our results on the original data. Consequently, our densely corresponding faces relate directly to the original 3D faces. We have established dense correspondence on the first available neutral scans of the 466 identities of FRGCv2 and included their indices in the supplementary material. These indices along with indices of densely corresponding 3D faces of BU3DFE dataset will be available on our website for the benefit of researchers to instantly construct a densely corresponding statistical model.

Figure 9 visually shows the detection results of 35 biologically significant anthropometric landmarks [15] on seven identities of FRGCv2 database. It is pertinent to mention that the 3D facial scans for medical research purposes are from cooperative subjects and captured in neutral expression. There is a subtle mild expression on the faces of subjects suffering from certain syndromes, e.g.

Autism [3, 20, 30]. It is evident from Figure 9 that our proposed algorithm is able to detect a large number of anthropometric landmarks on neutral and mild expressions with a reasonable accuracy. Objective evaluation of such a large number of landmarks is not feasible due to lack of ground truth and the challenge of manually annotating them on a huge database. A visual inspection shows the superior quality of detection.

4. Conclusion

Automatic detection of a large number of anthropometric facial landmarks on huge 3D face databases is challenging. To date, a subset of these landmarks is being manually annotated on the 3D faces by researchers and medical practitioners. We presented an algorithm for automatic detection of an arbitrarily large number of facial landmarks. We achieved this by establishing dense 3D face correspondences by minimizing the bending energy between seed points extracted by adaptive level set curves. Correspondences and hence landmarks are transferred to unseen faces by fitting a morphable model. Our results show that we not only outperform the state-of-the-art in terms of landmark localization error for commonly detected landmarks, but also reliably detect soft tissue landmarks on quasi-planar facial regions that are even hard to annotate manually.

5. Acknowledgements

Syed Zulqarnain Gilani is funded by the International Postgraduate Research Scholarship (IPRS). This research was also supported by ARC grant DP110102399.

References

- [1] D. Aiger, N. Mitra, and D. Cohen. 4-Points congruent sets for robust pairwise surface registration. *Transactions on Graphics*, 27(3):85, 2008. [2](#)
- [2] O. B. Albajalan, A. Samsudin, and R. Hassan. Craniofacial morphology of malay patients with obstructive sleep apnoea. *The European Journal of Orthodontics*, 33(5):509–514, 2011. [1](#)
- [3] K. Aldridge, I. D. George, K. K. Cole, J. R. Austin, T. N. Takahashi, Y. Duan, and J. H. Miles. Facial phenotypes in subgroups of prepubertal boys with autism spectrum disorders are correlated with clinical phenotypes. *Mol Autism*, 2(1):15, 2011. [1](#), [8](#)
- [4] M. Alexa. Recent advances in mesh morphing. *Computer Graphics Forum*, 21(2):173–198, 2002. [2](#)
- [5] S. Bejerot, J. M. Eriksson, S. Bonde, K. Carlström, M. B. Humble, and E. Eriksson. The extreme male brain revisited: gender coherence in adults with autism spectrum disorder. *The British Journal of Psychiatry*, 201(2):116–123, 2012. [1](#)
- [6] J. L. Bentley. Multidimensional binary search trees used for associative searching. *Communications of the ACM*, 18(9):509–517, 1975. [5](#)
- [7] V. Blanz and T. Vetter. Face recognition based on fitting a 3D morphable model. *IEEE TPAMI*, 25(9):1063–1074, 2003. [2](#)
- [8] F. L. Bookstein. Principal warps: Thin-plate splines and the decomposition of deformations. *IEEE TPAMI*, 11(6):567–585, 1989. [4](#)
- [9] B. Brown and S. Rusinkiewicz. Global non-rigid alignment of 3-D scans. *Transactions on Graphics*, 26(3):21, 2007. [2](#)
- [10] W. Chang and M. Zwicker. Automatic registration for articulated shapes. *Computer Graphics Forum*, 27(5):1459–1468, 2008. [2](#)
- [11] C. Creusot, N. Pears, and J. Austin. A machine-learning approach to keypoint detection and landmarking on 3D meshes. *International Journal of Computer Vision*, 102(1-3):146–179, 2013. [2](#), [5](#), [6](#)
- [12] R. Davies, C. Twining, and C. Taylor. *Statistical models of shape: Optimisation and evaluation*. Springer, 2008. [2](#)
- [13] R. H. Davies, C. J. Twining, T. F. Cootes, J. C. Waterton, and C. J. Taylor. 3D statistical shape models using direct optimisation of description length. In *ECCV*. 2002. [2](#)
- [14] M. Eder, G. Brockmann, A. Zimmermann, M. A. Papadopoulos, K. Schwenzer-Zimmerer, H. F. Zeilhofer, R. Sader, N. A. Papadopoulos, and L. Kovacs. Evaluation of precision and accuracy assessment of different 3-D surface imaging systems for biomedical purposes. *Journal of Digital Imaging*, 26(2):163–172, 2013. [1](#)
- [15] L. Farkas. Anthropometry of the head and face in clinical practice. *Anthropometry of the head and face*, 2nd Ed, pages 71–111, 1994. [1](#), [6](#), [8](#)
- [16] T. Funkhouser and P. Shilane. Partial matching of 3D shapes with priority-driven search. In *Eurographics Symposium on Geometry Processing*, volume 256, pages 131–142, 2006. [2](#)
- [17] S. Z. Gilani, K. Rooney, F. Shafait, M. Walters, and A. Mian. Geometric facial gender scoring: Objectivity of perception. *PLoS one*, 9(6), 2014. [1](#)
- [18] S. Gupta, M. K. Markey, and A. C. Bovik. Anthropometric 3D face recognition. *International Journal of Computer Vision*, 90(3):331–349, 2010. [2](#)
- [19] P. Hammond et al. The use of 3D face shape modelling in dysmorphology. *Archives of Disease in Childhood*, 92(12):1120, 2007. [1](#), [2](#)
- [20] P. Hammond, C. Forster-Gibson, A. Chudley, et al. Face-brain asymmetry in autism spectrum disorders. *Molecular Psychiatry*, 13(6):614–623, 2008. [2](#), [8](#)
- [21] C. L. Heike, M. L. Cunningham, A. V. Hing, E. Stuhau, and J. R. Starr. Picture perfect? Reliability of craniofacial anthropometry using three-dimensional digital stereophotogrammetry. *Plastic and Reconstructive Surgery*, 124(4):1261–1272, 2009. [1](#)
- [22] T. Heimann and H.-P. Meinzer. Statistical shape models for 3D medical image segmentation: A review. *Medical Image Analysis*, 13(4):543–563, 2009. [2](#)
- [23] H. Hochheiser, B. J. Aronow, K. Artinger, T. H. Beaty, J. F. Brinkley, Y. Chai, D. Clouthier, M. L. Cunningham, M. Dixon, L. R. Donahue, et al. The facebase consortium: a comprehensive program to facilitate craniofacial research. *Developmental Biology*, 355(2):175–182, 2011. [1](#)
- [24] R. W. Lee, K. Sutherland, A. S. Chan, B. Zeng, R. R. Grunstein, M. A. Darendeliler, R. J. Schwab, and P. A. Cistulli. Relationship between surface facial dimensions and upper airway structures in obstructive sleep apnea. *Sleep*, 33(9):1249, 2010. [1](#)
- [25] X. Lu and A. K. Jain. Automatic feature extraction for multiview 3D face recognition. In *Automatic Face and Gesture Recognition*, 2006. [2](#), [6](#)
- [26] C. McKnight, J. Newnham, F. Stanley, J. Mountain, L. Landau, L. Beilin, I. Puddey, C. Pennell, and D. Mackey. Birth of a cohort: The first 20 years of the Raine study. *The Medical Journal of Australia*, 197:608–610, 2012. [1](#)
- [27] A. Mian, M. Bennamoun, and R. Owens. An efficient multimodal 2D-3D hybrid approach to automatic face recognition. *IEEE TPAMI*, 29(11):1927–1943, 2007. [2](#)
- [28] H. Mirzaalian, G. Hamarneh, and T. Lee. A graph-based approach to skin mole matching incorporating template-normalized coordinates. In *IEEE CVPR*, 2009. [2](#)
- [29] P. Nair and A. Cavallaro. 3-D face detection, landmark localization, and registration using a point distribution model. *IEEE Tran. on Multimedia*, 11(4):611–623, 2009. [1](#), [7](#)
- [30] T. Obafemi-Ajayi, J. H. Miles, T. N. Takahashi, W. Qi, K. Aldridge, M. Zhang, S.-Q. Xin, Y. He, and Y. Duan. Facial structure analysis separates autism spectrum disorders into meaningful clinical subgroups. *Journal of Autism and Developmental Disorders*, pages 1–16, 2014. [1](#), [8](#)
- [31] G. Passalis, P. Perakis, T. Theoharis, and I. A. Kakadiaris. Using facial symmetry to handle pose variations in real-world 3D face recognition. *IEEE TPAMI*, 33(10):1938–1951, 2011. [2](#)
- [32] P. Perakis, G. Passalis, T. Theoharis, and I. A. Kakadiaris. 3D facial landmark detection under large yaw and expression variations. *IEEE TPAMI*, 35(7):1552–1564, 2013. [2](#), [6](#), [7](#)
- [33] G. Peyré. The numerical tours of signal processing-advanced computational signal and image processing. *IEEE Computing in Science and Engineering*, 13(4):94–97, 2011. [4](#)

- [34] G. Peyré and L. Cohen. Geodesic computations for fast and accurate surface remeshing and parameterization. In *Elliptic and Parabolic Problems*, pages 157–171. Springer, 2005. 2, 3
- [35] G. Peyré and L. D. Cohen. Geodesic remeshing using front propagation. *International Journal of Computer Vision*, 69(1):145–156, 2006. 3
- [36] P. Phillips, P. Flynn, T. Scruggs, K. Bowyer, et al. Overview of the face recognition grand challenge. In *IEEE CVPR*, 2005. 2, 5
- [37] U. Prabhu, J. Heo, and M. Savvides. Unconstrained pose-invariant face recognition using 3D generic elastic models. *IEEE TPAMI*, 33(10):1952–1961, 2011. 2
- [38] A. Salazar, S. Wuhler, C. Shu, and F. Prieto. Fully automatic expression-invariant face correspondence. *Machine Vision and Applications*, 25(4):859–879, 2014. 7
- [39] M. Segundo, L. Silva, P. Bellon, and C. C. Queirolo. Automatic face segmentation and facial landmark detection in range images. *IEEE Tran. on Systems, Man, and Cybernetics, Part B: Cybernetics*, 40(5):1319–1330, 2010. 2, 6, 7
- [40] J. A. Sethian. *Level set methods and fast marching methods: evolving interfaces in computational geometry, fluid mechanics, computer vision, and materials science*, volume 3. Cambridge University Press, 1999. 3
- [41] J. A. Sethian. Evolution, implementation, and application of level set and fast marching methods for advancing fronts. *J. of Computational Physics*, 169(2):503–555, 2001. 3
- [42] P. Szeptycki, M. Ardabilian, and L. Chen. A coarse-to-fine curvature analysis-based rotation invariant 3D face landmarking. In *Biometrics: Theory, Applications, and Systems*, 2009. 5
- [43] L. Yin, X. Wei, et al. A 3D facial expression database for facial behavior research. In *Automatic Face and Gesture Recognition*, pages 211–216, 2006. 1, 5