

# Dense 3D Face Correspondence

Syed Zulqarnain Gilani<sup>1b</sup>, Ajmal Mian, Faisal Shafait, and Ian Reid

**Abstract**—We present an algorithm that automatically establishes dense correspondences between a large number of 3D faces. Starting from automatically detected sparse correspondences on the outer boundary of 3D faces, the algorithm triangulates existing correspondences and expands them iteratively by matching points of distinctive surface curvature along the triangle edges. After exhausting keypoint matches, further correspondences are established by generating evenly distributed points within triangles by evolving level set geodesic curves from the centroids of large triangles. A deformable model (K3DM) is constructed from the dense corresponded faces and an algorithm is proposed for morphing the K3DM to fit unseen faces. This algorithm iterates between rigid alignment of an unseen face followed by regularized morphing of the deformable model. We have extensively evaluated the proposed algorithms on synthetic data and real 3D faces from the FRGCv2, Bosphorus, BU3DFE and UND Ear databases using quantitative and qualitative benchmarks. Our algorithm achieved dense correspondences with a mean localisation error of 1.28 mm on synthetic faces and detected 14 anthropometric landmarks on unseen real faces from the FRGCv2 database with 3 mm precision. Furthermore, our deformable model fitting algorithm achieved 98.5 percent face recognition accuracy on the FRGCv2 and 98.6 percent on Bosphorus database. Our dense model is also able to generalize to unseen datasets.

**Index Terms**—Dense correspondence, 3D face, morphing, keypoint detection, level sets, geodesic curves, deformable model

## 1 INTRODUCTION

ONE of the canonical tasks in shape analysis is to find a meaningful mapping between two or more shapes [1]. The process, called shape correspondence, is a pre-requisite for many computer vision, computer graphics and medical image analysis applications. The requisite density of correspondence is often dictated by the underlying shape and target application. Sometimes, sparse correspondence is sufficient to infer shape semantics by matching representative points, for example the four corners of a rectangle or emblematic points on key joints of a human body. However, sparse correspondence is often inadequate in case of articulated shapes [2], [3] where parts of the shape can bend independently or in the correspondence of anatomical shapes which can deform in an elastic manner [4]. In such circumstances, dense correspondence is required to guarantee representation of global shape changes, for instance in case of morphing or attribute transfer. Furthermore, very subtle changes within a class of shapes can be detected only if the correspondence between these shapes is dense [6].

In this paper we are concerned with the task of finding dense correspondences between a very large number of similar shapes; in our case 3D scans of human faces. We do so

because this further enables us to generate highly accurate 3D morphable models that can be used for information transfer between the training set and a test face or between two test faces by morphing the 3D model to fit the test face(s). For example, given the location of anthropometric landmarks [7] on the 3D morphable model, these landmarks can be automatically localized on previously unseen test faces [8]. Furthermore, dense correspondences and morphable models can be used for 3D face recognition [9], [10]. Other applications include facial morphometric measurements such as gender scoring [5] and asymmetry for syndrome diagnosis [6], statistical shape modelling [11], [12], shape interpolation [13], non-rigid shape registration [3], [14], [15], deformation analysis [16] and recognition [17], [18], [19].

While it is possible to manually annotate a small number ( $\sim 30$ ) of correspondences for a few 3D faces, it is not feasible to manually identify dense correspondences ( $\sim 6,000$ ) between hundreds of 3D faces. The literature also proposes computing dense correspondence by extending manually annotated sparse ones [9], [20]. However, with the advent of huge 3D face databases like the Facebase Consortium [21] or Raine dataset [22], [23], this strategy too has become impractical and calls for fully automatic algorithms. Automatically establishing dense correspondences between the 3D faces of two different persons is an extremely challenging task because the facial shape varies significantly amongst individuals depending on their identity, gender, ethnicity and age [7] as well as their facial expression and pose. The problem of dense 3D point-to-point correspondences can be formulated as follows. Given a set of  $N$  3D faces,  $\mathbf{F}_j = [x_p, y_p, z_p]^T$ ,  $j = 1, \dots, N, p = 1, \dots, P_j$ , the aim is to establish a dense bijective mapping  $f: \mathbf{F}_i \rightarrow \mathbf{F}_j (i \neq j)$  over  $k$  vertices where  $1 < k < \min(P_i, P_j)$ . Correspondences should cover all regions of the face for high fidelity and should follow the same triangulation for shape consistency.

- S.Z. Gilani, A. Mian, and F. Shafait are with the School of Computer Science and Software Engineering, The University of Western Australia, 35 Stirling Highway, Crawley, WA, 6009, Australia. E-mail: syedzulqarnain.gilani@research.uwa.edu.au, {ajmal.mian, faisal.shafait}@uwa.edu.au.
- I. Reid is with the School of Computer Science, University of Adelaide, Ingkarni Wardli, North Terrace Campus, Adelaide, SA 5005, Australia. E-mail: ian.reid@adelaide.edu.au.

Manuscript received 15 June 2016; revised 2 May 2017; accepted 3 July 2017.  
Date of publication 10 July 2017; date of current version 12 June 2018.

(Corresponding author: Syed Zulqarnain Gilani.)

Recommended for acceptance by R. Yang.

For information on obtaining reprints of this article, please send e-mail to: reprints@ieee.org, and reference the Digital Object Identifier below.

Digital Object Identifier no. 10.1109/TPAMI.2017.2725279

Existing dense correspondence techniques have one or more of the following limitations: (1) They need manually annotated landmarks on 3D faces for initialization. (2) They use image texture matching to find 3D shape correspondence. (3) They correspond all faces to a single reference face neglecting the global proximity of the 3D faces. (4) They have not been tested on *complete* benchmark databases such as the FRGCv2 [24] or Bosphorus [25] datasets for face recognition and landmark identification. (5) They have no explicit mechanism of updating the dense correspondence model.

In this context, we propose a fully automatic algorithm for establishing dense correspondences simultaneously between a large number of 3D faces. Our algorithm does not require any manual intervention and relies solely on 3D shape matching to encode accurate facial morphology. We organize the 3D faces into a minimum spanning tree based on bending energy required to deform one shape into the other so that correspondences can be propagated in a reliable way. We propose a mechanism for automatic initialization of a sparse set of correspondences on the outer boundary of the 3D faces. We form a triangulation of these correspondences, and iteratively add to the set of points by matching points of distinctive surface curvature along (and close to) the triangulated edges. After exhausting the possibilities for such matches, we further expand the set of matches by generating points distributed evenly within triangles by evolving level set geodesic curves from the centroids of large triangles. The outcome of our algorithm is a Keypoint-based 3D Deformable Model (K3DM).

Our second major contribution is a deformable model fitting algorithm where K3DM is used to morph into unseen query faces. Starting from the mean face, the fitting algorithm iterates between two steps. The query face is transformed rigidly to align with the model and the model is deformed using regularized least squares to fit the query face. This algorithm converges in a few iterations and is robust to noise, outlier points, missing points, pose and expression variations.

Our final contribution is an algorithm for augmenting the K3DM. Given the K3DM and a new batch of  $M$  faces, we construct a minimum spanning tree using the nearest face to the K3DM as the root node. The K3DM is augmented by adding one face at a time, starting with the root node, and each time updating the model and deforming the updated model to better fit the next face in the spanning tree.

Evaluating dense correspondence techniques is challenging due to the inherent difficulty of obtaining ground-truth data. In the existing literature, evaluations have mostly been performed on a sparse set of anthropometric facial landmarks [26], [27], [28] since these can be manually labelled. However, evaluation on only a few ( $\leq 20$ ) anthropometric points does not show how well dense correspondences have generalized to the whole face. Thus, subjective evaluations are frequently performed [29] by visually inspecting the quality of morphing between faces [4], [30]. In this paper, we show how synthetic 3D faces (Facegen Modeller) can be used to quantitatively evaluate dense correspondences on a large set of points ( $\geq 1,000$ ). Using the presented deformable face model, we perform extensive experiments for landmark localization (Section 6.1) and face recognition (Section 6.2) using real faces from the

FRGCv2 [24] and BU3DFE [31] databases. Results show that our algorithm outperforms state-of-the-art application-specific algorithms in each of these areas.

## 2 RELATED WORK

Existing 3D correspondence techniques can be grouped into descriptor based, model based and optimization based [1].

*Descriptor Based Techniques.* These techniques match local 3D point signatures derived from the curvatures, shape index and normals. However, they are often highly sensitive to surface noise and sampling density [33] of the underlying geometry [34]. More significantly for our purpose, the density of corresponding points is typically low resulting in correspondences between a very sparse set of anthropometric landmarks.

One of the earliest works, in this category, for establishing dense correspondence was proposed by Sun and Abidi [30], [32] who projected geodesic contours around a 3D facial point onto their tangential plane and used them as features to match two surfaces. The approach, with minor modifications, was employed by Salazar et al. [35] to establish point correspondence on 3D faces in BU3DFE database. Lu and Jain [36] presented a multimodal approach for facial feature extraction. Using a face landmark model, the authors detected seven corresponding points on 3D faces using shape index from range images and corneriness from intensity images. Segundo et al. [37] combined surface curvature and depth relief curves for landmark detection in 3D faces of the FRGCv2 and BU3DFE databases. They extracted features from the mean and Gaussian curvatures for detecting five landmarks in the nose and eye (high curvature) regions.

Creusot et al. [27] presented a machine learning approach to detect 14 corresponding landmarks on 3D faces. They trained multiple LDA classifiers on a set of 200 faces and a landmark model using a myriad of local descriptors. Each landmark detection was treated as a two class classification problem and the final results were fused. This method works well for neutral expression faces of the FRGCv2 and Bosphorus databases. Perakis et al. [26] 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 profile views of the face. Keypoints are detected using Shape Index and Spin Images and then matched 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 and UND Ear databases. Later, the authors proposed a technique [38] for fusing features from 2D and 3D data to detect these landmarks with better accuracy than [26].

Some methods have also been proposed for generating sparse correspondence for 3D face recognition [44], [45], [46], [47]. However, these methods are based on keypoint correspondences that are repeatable only on the same identity.

*Model Based Techniques.* These approaches create a morphable model using a sparse set of correspondences and then extend them to dense correspondences.

Employing a Point Distribution Model coupled with 3D point signature detection, Nair and Cavallaro [8] estimated the location of 49 corresponding landmarks on faces. They tested their algorithm on 2,350 faces of the BU3DFE [31]

database and reported a rather high mean landmark localization error.

Blanz and Vetter [29] proposed a dense correspondence algorithm using optical flow on the texture and the 3D cylindrical coordinates of the face points assuming that the faces are spatially aligned. They constructed a 3D morphable face model from 100 male and female faces each. An arbitrary face was chosen as a reference and the remaining scans were registered to it by iterating between optical flow based correspondence and morphable model fitting. One potential pitfall of the texture based dense correspondence [29] is that facial texture is not always consistent with the underlying 3D facial morphology, e.g., the shape and location of eyebrows. Moreover, this algorithm requires seven manually annotated facial landmarks for initialization. Later, in [9], [39] the authors used the 3D morphable model for face recognition. Experiments were performed on only 150 pairs of 3D faces [39] from FRGCv2 database, although the total number of scans in the database are 4,007. The seminal work of Blanz and Vetter [29] was extended by Paysan et al. [40] in the Basel Face Model (BFM) which used an improved mesh registration algorithm [41]. The authors have made their dense correspondence model publicly available which has enabled us to draw comparisons with their model.

Passalis et al. [42] proposed an Annotated Face Model (AFM) based on an average facial 3D mesh. The model was created by manually annotating a sparse set of anthropometric landmarks [7] on 3D face scans and then segmenting it into different annotated areas. Later, Kakadiaris et al. [43] proposed elastic registration using this AFM by shifting the manually annotated facial points according to elastic constraints to match the corresponding points of 3D target models in the gallery. Face recognition was performed by comparing the wavelet coefficients of the deformed images obtained from morphing. Passalis et al. [18] further improved the AFM by incorporating facial symmetry to perform pose invariant face recognition. However, the algorithm depends on detection of at least five facial landmarks on a side pose scan.

Level set curves were evolved in [28] to automatically extract seed points and correspondences were established by minimizing the bending energy between patches around seed points of different faces. A morphable model based on the dense corresponding points was then fitted to unseen query faces for transfer of correspondences. The accuracy of landmark localization in this method depends on the number and accuracy of initial seed points.

**Optimization Based Techniques.** These methods optimize an objective function to find a mapping between fiducial points. Non-rigid ICP (NICP) is one such technique which formulates deformable registration as an optimization problem consisting of a mesh smoothness term and several data fitting terms [41], [78]. These algorithms require accurate global initialization points ranging from 14 points [41] to 68 points [77]. These points are either manually annotated [29], [41] or detected automatically using texture [77]. An extension to this method removes the need for fiducial points but assumes a partial overlap of facial regions [78], [79]. The alignment between two faces is performed with a global rigid transformation followed by per-vertex affine transformations that bring the non-rigid shapes into full alignment.

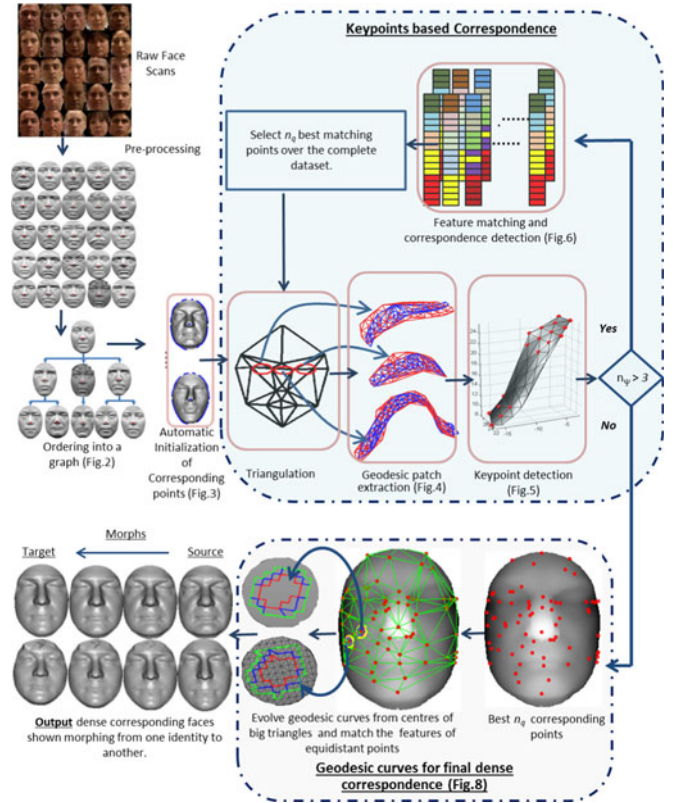


Fig. 1. Block diagram of the presented dense 3D face correspondence algorithm.

Such methods are more suited for time varying deformations of the same identity and often do not result in a bijective (one-to-one) mapping of the vertices. Booth et al. [77] constructed a dense correspondence model of several faces from a propriety dataset by registering the scans to a template mesh using NICP algorithm [41] initialized by 68 fiducial landmarks detected using texture. Bolkart et al. [80] presented dense correspondence as an optimization problem and used the Minimum Description Length (MDL) [11] as the objective function. The authors of methods that are based on NICP [40], [41], [77], [78], [79] or other alternative optimization techniques [80] have not reported facial landmark localization results. Hence, it is difficult to perform a direct objective comparison with these methods.

### 3 DENSE 3D FACE CORRESPONDENCE

The overall idea of our system for dense correspondence between 3D face scans, is to begin with a set of automatically extracted seed points that represent points matched across all faces in the dataset, and gradually densify the set of matches. Fig. 1 depicts the overall flow of our system. Here we give an overview of how this proceeds, and then expand the details in the sections below.

We first organize the faces into a tree (Section 3.1) based on similarity. We then seek a set of reliable seed matches (Section 3.2) from which to begin an iterative densification process. Each iteration of the densification process (Section 3.3) begins by selecting the current best set of matches (comprising  $n_q$  of the full set of  $n$  matches) and forming a triangulation of these points. Taking each edge of the resulting triangulation in turn, we extract a narrow patch

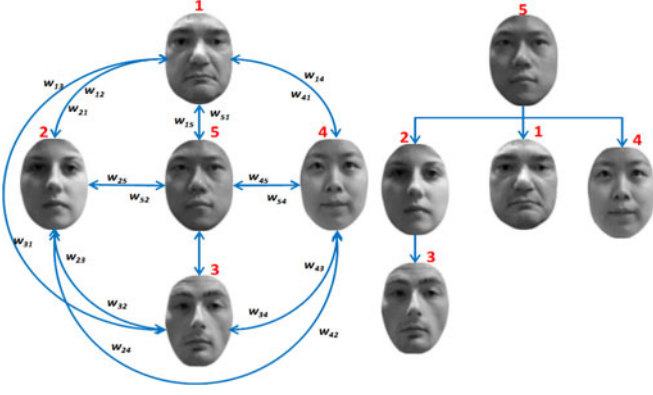


Fig. 2. The directed graph  $G = (V_g, E_g)$  (Left) and the Minimum Spanning Tree (MST)  $\Pi = (V_t, E_t)$  (Right) constructed from five example images of FRGCv2.

centered on the edge from a pair of faces that are adjacent in the tree. For each of these patches we find points of distinctive curvature (Section 3.4) - these will be new candidate matches, or *keypoints* - and compute a 38-dimensional descriptor of the local surface around each keypoint. Using constrained nearest neighbor we then determine points that match well between the pair of patches (i.e., their descriptors match and they are within a proximity threshold in the patch). We repeat this process for all parent/child pairs throughout the tree, and eliminate all keypoints that are not consistently matched throughout the tree. The remaining keypoints that are successfully matched across all faces in the dataset are added to the current set of matches. At the end of one iteration, when we have cycled through all the triangulated edges, we choose a new best set of  $n_q$  matches and repeat the process.

Once the search for keypoints is exhausted, further correspondences on facial areas devoid of discriminative points are established by first evolving level set curves and sampling equidistant vertices (See Section 3.6). Feature vectors of these vertices on the reference face are then matched with the remaining faces to establish correspondence as previously stated.

The outcome of this process is a set of densely corresponding 3D faces which we call the Keypoint-based 3D Deformable Model (K3DM).

### 3.1 Preprocessing and Organizing Faces

The nose tip of a 3D face is detected automatically following Mian et al. [67]. Centering a sphere at the nose tip, the face is cropped. The pose of the 3D face is iteratively corrected to a canonical form using the Hotelling transform [53]. Next, holes are filled and noise is removed using the gridfit algorithm [76].

Next, we pre-organise the face dataset into a graph (in fact, a tree) in which similar faces are “close” to one another. Let  $G = (V_g, E_g)$  be a directed graph where each node  $V_g$  is a 3D face  $F$  from the dataset and each edge  $E_g$  connects two nodes  $(v_i, v_j)$  of the graph. Each edge of the graph has weight  $w$

$$w(v_i, v_j) = \frac{\beta_{ij} + \beta_{ji}}{2}, \quad (1)$$

where  $\beta_{ij}$  is the amount of bending energy required to deform face  $F_i$  to  $F_j$  and is measured using the 2D thin-plate

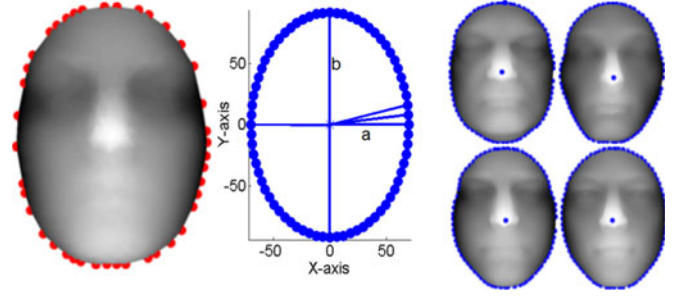


Fig. 3. (a) Vertices of the 2D-convex hull of the projection. (b) Points sampled at angular intervals of  $\pi/36$ . (c) Initial sparse correspondence projected on four identities of the FRGCv2 dataset.

spline model [48]. Note that  $\beta_{ij} \neq \beta_{ji}$  and  $\beta_{ii} = 0$ . Since, the faces are already roughly aligned, their nearest neighbor points are taken as approximate correspondences for the purpose of calculating the bending energy. From  $G$ , we construct a minimum spanning tree  $\Pi = (V_t, E_t)$  using Kruskal’s algorithm. The node with the maximum number of children is taken as the root node.

The purpose of this pre-organisation is to increase the likelihood of finding point matches between pairs of faces. A naive approach would be to arbitrarily choose a single (or average) face as reference and find its correspondences to others in the dataset. But such an approach ignores the proximity between the face instances and the global information underlying the population. The process and a sample graph are shown in Fig. 2.

### 3.2 Sparse Correspondence Initialization

We initialize the correspondences by first automatically establishing a sparse set of seed points. We restrict these seed points to those that lie on the roughly ellipse-shaped 2D convex hull of the face, i.e., the 2D-hull when the 3D mesh is projected into the  $x - y$  plane. We sample these points at regular angular intervals of  $\delta = \pi/36$  (see Fig. 3), where the angle  $\delta$  is measured at the nose tip. There is of course no guarantee that in the finite resolution mesh of the face there will be a point at an exact multiple of  $\pi/36$ , but for each face we choose the nearest point. This yields a set of 72 3D seed points for each 3D face in the dataset which are used in the first iteration of the triangulation and densification process, as described in the next section.

### 3.3 Triangulation and Geodesic Patch Extraction

The main part of our algorithm is an iteration that takes the best set of matches that have been established to date, and grows the number of correspondences. For the first iteration, we use the sparse set of correspondences established as in the previous section, while for subsequent iterations we determine the best set of  $n_q$  matches from the full set of  $n$  matches as described in Section 3.5.

In each iteration, given  $n_q$  correspondences between  $N$  faces, we perform a 2D Delaunay triangulation of the mean  $x - y$  locations of the  $n_q$  current best matches. This triangulation is then used consistently across all faces. We then pick a pair of parent/child nodes from the Minimum Spanning Tree  $\Pi$ ,  $F_j$  and  $F_k$ . For both faces in the pair, we extract

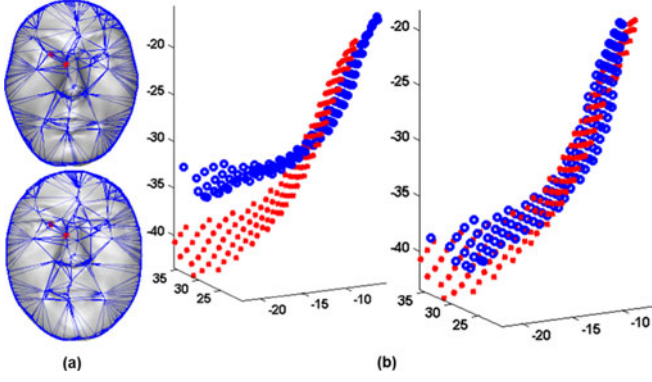


Fig. 4. Illustration of geodesic patch extraction. (a) Two 3D faces with triangulation over a few corresponding points from the 2<sup>nd</sup> iteration. Geodesic surface patch is extracted between two sample points shown in red colour. (b) Pointclouds of the geodesic surface patches before and after registration.

a narrow surface patch  $\mathbf{S} = \{[x_i, y_i, z_i]^T, i = 1, \dots, m\} \subset \mathbf{F}$ , centered on a geodesic curve defined by each triangle edge (see Fig. 4). For the sake of simplicity we call this a geodesic patch.

The (projected) length of the patch is the same as the length of the edge. The “narrow” width is set with reference to the scale of the original face mesh resolution. More specifically, we set the width to be  $5\rho$  where  $\rho$  is the average mesh-edge length in the vicinity of the endpoint of the edge (note that here the mesh-edges refer to the edges in the original dataset, not the edges of the triangulation used for the densification). This makes the extraction of the geodesic patch scale invariant. The values of  $\rho$  for real 3D faces captured with the Minolta or the 3dMDface scanners typically range from 1-3 mm.

Finally, we bring the patches  $\mathbf{S}_j, \mathbf{S}_k$  into approximate alignment using non-rigid registration [49], [50]. The process is shown in Fig. 4.

### 3.4 Keypoint Detection on Geodesic Patches

Our aim now is to establish accurate correspondences between a patch on one face  $\mathbf{S}_j$  and its corresponding patch on the other face  $\mathbf{S}_k$ . We do this in a fairly standard manner by finding distinctive keypoints, generating a descriptor of the local surface around each point, and establishing matches between points on each patch whose descriptors are sufficiently close.

More specifically, to find keypoints we consider the surface distinctiveness at each point in the patch. We do so by calculating the covariance of all the points within a neighborhood of  $5\rho$  of the current point, and marking as keypoints any points for which the ratio of the largest two eigenvalues of the covariance exceeds a threshold. Note that if the neighborhood is uniform these eigenvalues will be equal, and therefore the point is unsuitable as a keypoint.

Fig. 5 shows keypoints detected by our algorithm in the tenth iteration on four different identities of the FRGCv2 database.

We use the keypoints detected on surface patch  $\mathbf{S}_j$  for feature extraction and matching only if an adequate number of keypoints are detected (we use a minimum of three), otherwise,  $\mathbf{S}_j$  is not considered to be sufficiently descriptive and the matches are not sought within the patch.

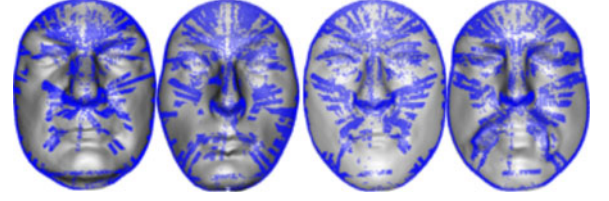


Fig. 5. Illustration of keypoints (not corresponding points) detected along geodesic patches in the tenth iteration of our algorithm. Notice the repeatability of keypoints across the identities.

### 3.5 Feature Extraction and Matching

We denote by  $\Psi_j = [x_i, y_i, z_i]^T, i = 1, \dots, n_\Psi$  the set of keypoints detected on the surface  $\mathbf{S}_j$ , where  $n_\Psi$  is the number of keypoints (likewise for  $\Psi_k$ ). For each keypoint we extract a feature vectors  $\mathbf{x}$  which describe the local surface (within  $5\rho$ ) using a set of 3D signature and histogram based descriptors. These descriptors have been widely used in the literature [27], [51], [52] for automatic object recognition and for landmark detection. We use a combination of many descriptors since the surface patch is quite small and a single descriptor may not capture sufficient information. The list of descriptors is given below:

- The spatial location  $[x_i, y_i, z_i]^T$ .
- The surface normal  $[n_x, n_y, n_z]^T$ .
- The seven invariant moments [53] of the  $3 \times 3$  histograms of the  $XY, YZ$  and  $XZ$  planes.
- The central moment  $\mu_{mn}$  of order  $m + n$  of the histogram matrix  $\mathbf{H}$

$$\mu_{mn} = \sum_{i=1}^{\varphi} \sum_{j=1}^{\varphi} (i - \bar{i})^m (j - \bar{j})^n \mathbf{H}(i, j), \quad (2)$$

where  $\varphi$  is the total number of points in  $\mathbf{H}$ ,  $\bar{i} = \sum_{i=1}^{\varphi} \sum_{j=1}^{\varphi} i \mathbf{H}(i, j)$  and  $\bar{j} = \sum_{i=1}^{\varphi} \sum_{j=1}^{\varphi} j \mathbf{H}(i, j)$ .

- The mean of the two principle curvatures  $k_1$  and  $k_2$  calculated at each point on the extracted local surface
- The Gaussian Curvature  $K = k_1 k_2$
- The Mean Curvature  $H = \frac{k_1 + k_2}{2}$
- The Shape Index. We use two variants of the shape index which vary from 0 to 1 and  $-1$  to 1 respectively,  $s_a = \frac{1}{2} - \frac{1}{\pi} \arctan \frac{k_1 + k_2}{k_1 - k_2}, \quad 0 \leq s_a \leq 1$  and  $s_b = \frac{2}{\pi} \arctan \frac{k_1 + k_2}{k_1 - k_2}, \quad -1 \leq s_b \leq 1$ .
- The Curvedness  $c = \sqrt{\frac{k_1^2 + k_2^2}{2}}$
- The Log-Curvedness  $c_l = \frac{2}{\pi} \log \sqrt{\frac{k_1^2 + k_2^2}{2}}$
- The Willmore Energy  $e_w = H^2 - K$ ,
- The Shape Curvedness  $c_s = s_b \cdot c_l$
- The Log Difference Map  $m_l = \ln(K - H + 1)$ .

Using these descriptors, the dimensionality of the final feature vector  $\mathbf{x}$  is 38. These features are extracted over a small enough local surface centered at the keypoint such that they are repeatable across identities. In contrast, the feature vector extracted by Mian et al. [44], [54] takes the range values of a larger surface (typically 20 mm radius) surrounding each keypoint. Hence, their features are repeatable only over the same identity. One of the prerequisites of the techniques that use depth values as features [44], [51],

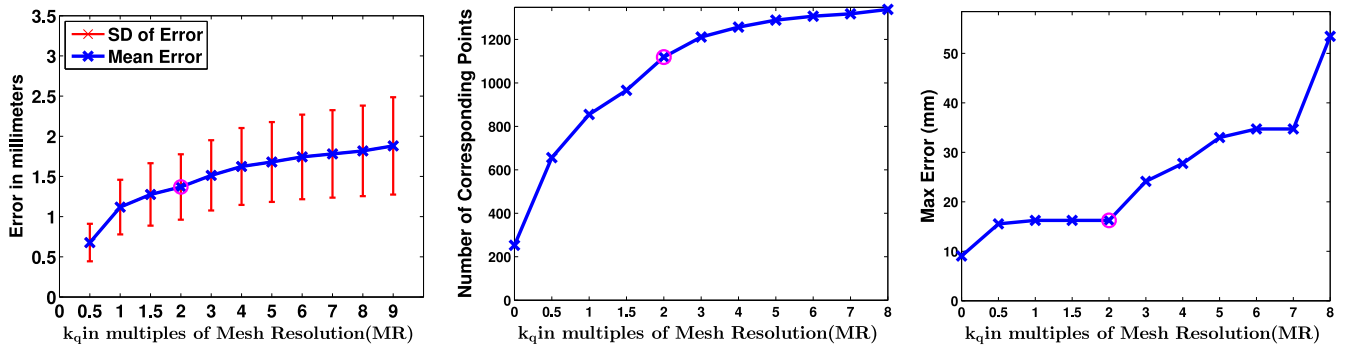


Fig. 6. The effect of correspondence quality threshold  $k_q$  in the synthetic dataset in the first iteration. (Left) Graph of  $k_q$  versus the mean and SD of correspondence localization error. (Middle)  $k_q$  versus the number of correspondences established. (Right)  $k_q$  versus the maximum localization error. For all our experiments we have set  $k_q = 2\rho$  shown in the graphs in a magenta circle.

[52], [54] is to define a local reference frame for pose invariant matching. In our case, the features are quasi pose invariant and hence do not require a local reference frame. This is because the pose of each training faces has been iteratively corrected to a canonical form during preprocessing and the features are extracted from a very small patch.

Next, we perform constrained-NN search between the feature vectors  $\mathbf{x}_j$  from  $\mathbf{S}_j$  and  $\mathbf{x}_k$  from  $\mathbf{S}_k$ , such that the corresponding points lie within a proximity of  $2\rho$  to each other, and their matching score  $d(\mathbf{x}_j, \mathbf{x}_k)$  - taken to be Euclidean distance between their feature descriptors - is less than a threshold  $k_q$ . The quality of correspondence varies directly with  $k_q$ . Higher values of  $k_q$  will result in poor corresponding points with large errors, whereas lower values of  $k_q$  may reject valid correspondences and hence adversely effect the correspondence density. Fig. 6 shows the effect of  $k_q$  on the correspondence found in our experiments on the synthetic dataset. As we increase the value of  $k_q$ , the mean localization error and its standard deviation (SD) increases. Fig. 7 shows the outcome of this step on two identities.

This process is repeated for all surface patches in a pair of faces, and for all pairs of faces in the MST. Only points that are matched throughout the MST in the pairwise scheme are retained and these are added to the set of correspondences obtained for the previous iteration. We then select from the full set of correspondences those which have the smallest matching score  $d(\mathbf{x}_j, \mathbf{x}_k)$ . We denote the number of selected correspondences by  $n_q$  and use a value of  $n_q = 80$  in our experiments. In order to adequately cover

the whole face for the subsequent iteration, we add the original seed points to the  $n_q$  points. Next, we obtain a triangulation of these points on the mean face of the dataset and extract geodesic surface patches as described in Section 3.3, repeating the process.

### 3.6 Densifying Matches in Uniform Regions

Keypoints, by their very definition concentrate around regions of high curvature/discrimination, such as the mouth, nose, and eyes. In this section we describe how we establish correspondences in more uniform regions where keypoints cannot be found. A simple approach to establish dense correspondence in these areas would be to sample them uniformly within triangles of the Delaunay triangulation. This approach has been used in 2D by Munsell et al. [55] who pre-organized the shape instances based on a similarity measure and then established correspondence between pairs of shapes by mapping the points from the source instance to the target instance after minimizing a bending energy. However, a uniform sampling in the triangle only results in uniform sampling on the face in planar regions. Instead, we adopt a sampling strategy that respects the underlying surface distances (geodesics) on each face.

After triangulation of the final set of best quality corresponding points, we select large triangles with area greater than  $t_a$ . We set  $t_a$  to be the mean area of all triangles in the connectivity, an effective and expeditious choice. From the centroid of each triangle, we evolve a level-set curve, in which the front speed is set to be uniform along a (radial) geodesic. For convenience we refer to these curves as “level-set geodesics”. We follow the Fast Marching Method [56] and use the implementation given by Peyre [57]. We then sample the points along the curve at regular intervals to ensure equidistant points (see Fig. 8). Because the evolution

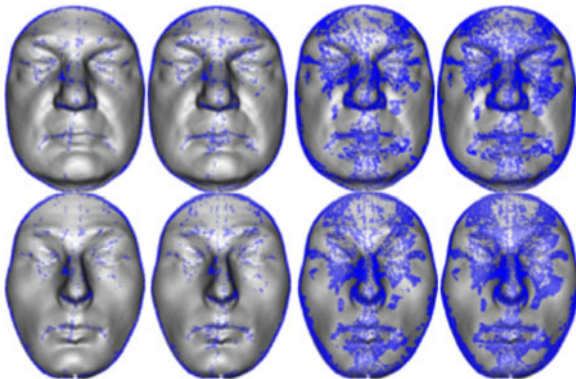


Fig. 7. Correspondence established in 1st, 4th, 13th and 18th iteration of our algorithm on the first two identities of FRGCv2. Notice how well the points correspond across the identities.

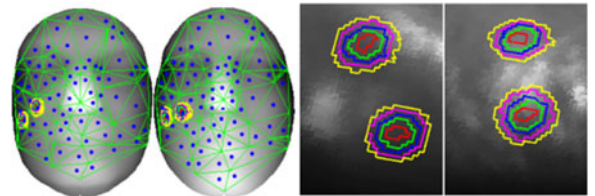


Fig. 8. Correspondence establishment on smooth surfaces. Two faces from an ordered pair with triangulation over  $n_q$  best quality corresponding points. Blue dots indicate the centroids of large triangles. Level set based evolution of geodesic curves for the two sample triangles, magnified on the right.

speed of the curve is uniform along geodesics, we obtain a uniform sampling on the surface; this is in contrast to uniform sampling within the triangle which would not necessarily be uniform on the surface itself. Although these points are not keypoints, they are repeatable on all 3D faces across identities because they are extracted from triangles whose vertices are corresponding to each other across the dataset. Furthermore, these points are extracted at equal intervals over a small region which is smooth.

Given this set of points sampled uniformly on the surface, we extract feature vectors and perform pairwise matching as before. Points whose feature vectors are close enough are retained as matches, with the rest discarded. This is not an iterative process and points are sampled only once from each triangle meeting the threshold criterion. Fig. 8 visually illustrates the process.

An alternative method for densifying the matches in the uniform regions could be to register the source and target faces using the NICP algorithm [41], [78] initialized by the correspondences established in the previous section. Once the target face has deformed to the source face, densification of correspondences is achieved by mapping the vertices in uniform regions of the source face to that of the target face. This approach requires tweaking the NICP parameters and the iterative optimization process for non-rigid face deformation tends to be computationally expensive. Our results in Section 6.1 also show that our feature matching approach achieves higher accuracy and therefore, we use this approach for the remaining part of the paper.

#### 4 K3DM FITTING AND AUGMENTATION

The output of the dense correspondence algorithm is the set of  $N$  densely corresponding 3D faces  $\tilde{\mathbf{F}}_j$ . Our objective now is to develop a compact deformable model based on these densely corresponding faces. To do so we take a standard PCA-based approach, and we call the result our Keypoint-based 3D deformable model. More formally, let  $\Upsilon = [\tilde{\mathbf{f}}_1, \tilde{\mathbf{f}}_2, \dots, \tilde{\mathbf{f}}_N]$ , where  $\tilde{\mathbf{f}} = [x_1, \dots, x_p, y_1, \dots, y_p, z_1, \dots, z_p]^T$  and  $p = 1, \dots, P$ . The row mean  $\mu_\Upsilon$  of the K3DM is given by,

$$\mu_\Upsilon = \frac{1}{N} \sum_{i=1}^N \tilde{\mathbf{f}}_i. \quad (3)$$

The row-normalized model  $m = \Upsilon - \mu_\Upsilon$  can be modelled by a multivariate Gaussian distribution and its eigenvalue decomposition is given by,

$$\mathbf{USV}^T = \Upsilon_m, \quad (4)$$

where  $\mathbf{US}$  are the principal components (PCs), the columns of  $V$  are their corresponding loadings, and  $\mathbf{S}$  is a diagonal matrix of eigenvalues. We use only the first  $n$  columns of  $\mathbf{U}$  which correspond to 98 percent of the energy.

We propose to deform the statistical model given in (4) into a query face  $\mathbf{Q}$  in a two step iterative process, i.e., registration and morphing. Algorithm 1 gives the details of fitting the deformable model to a query face. Note that we use  $\mathbf{Q}$  and  $\mathbf{M}$  for the point clouds of the query face and model and use  $\mathbf{q}$  and  $\mathbf{m}$  for their vectorized versions respectively. The query face after vectorization can be parametrized by the statistical model such that  $\mathbf{m}^i = \mathbf{U}\alpha^i + \mu_\Upsilon$ , where the

vector  $\alpha^i$  contains the parameters which are used to vary the shape of the model in the  $i$ th iteration and  $\mathbf{m}^i$  is the vectorized form of the model representing the query face. In the initialization step  $\alpha^i$  is set to zero and the deformable model  $\mathbf{M}^i$  is characterized by the mean face of the K3DM. Each iteration begins with a registration step where the input face  $\mathbf{Q}$  is registered to the model  $\mathbf{M}^i$ . This step essentially entails finding an approximate correspondence between the model and the query face and a rigid transformation. Correspondence is established by searching for the Nearest Neighbor (NN) of each point of  $\mathbf{M}^i$  in  $\mathbf{Q}$  using the k-d tree data structure [58]. Let  $\mathbf{d}$  represent the NN euclidean distance between the corresponded query face and the model such that  $d_j = \|\tilde{\mathbf{Q}}_j^i - \mathbf{M}_j^i\|_2$ . We define outliers as points on  $\tilde{\mathbf{Q}}$  whose NN distance with  $\mathbf{M}^i$  is greater than a threshold  $t_c$  where  $t_c = \bar{\mathbf{d}} + 3\sigma_d$  and exclude them from registration. This step ensures that the outliers do not affect the registration process. Next, the query face is translated to the mean of the model and is rotated to align with  $\mathbf{M}^i$ . We denote the corresponded and registered query face by  $\mathbf{Q}_r$ .

In the next step, the model  $\mathbf{M}^i$  is deformed to fit the registered query face  $\mathbf{Q}_r$  such that,

$$\alpha^i \leftarrow \min_{\alpha^i} \|\mathbf{U}^* \alpha^i + \mu_\Upsilon - \mathbf{q}_r\|_2 + \lambda \|\alpha^i - \alpha^{i-1}\|_2, \quad (5)$$

and  $\mathbf{m}^i = \mathbf{U}\alpha^i + \mu_\Upsilon$ . The  $*$  denotes that only those points (rows of  $\mathbf{U}$  and  $\mu_\Upsilon$ ) are considered which satisfy the threshold  $t_c$ . The second term in (5) puts a constraint on deforming the model. The applied condition is intuitive because we want to partially deform the model in each iteration such that the model approximates the query face in small steps. The iterative procedure is terminated when the residual error  $\|\mathbf{m}^i - \mathbf{q}_r\|_2 \leq \epsilon_f$ . In all of our experiments  $\lambda$  was set to 0.8 and  $\epsilon_f = 10^{-4}$ . Fig. 9 shows the K3DM fitting results on three datasets.

---

##### Algorithm 1. K3DM Fitting

---

**Require:**  $\Upsilon_m = [\tilde{\mathbf{f}}_1, \tilde{\mathbf{f}}_2, \dots, \tilde{\mathbf{f}}_N] - \mu_\Upsilon$  and Query Face  $\mathbf{Q} = [x_p, y_p, z_p]^T$  where  $p = 1, \dots, P_q$ .

**Initialization:**

- 1: Iteration:  $i = 0$  and  $\epsilon^0 = 1$
  - 2:  $\mathbf{USV}^T = \Upsilon_m$
  - 3:  $\alpha^i = 0$  and  $\mathbf{m}^i = \mathbf{U}\alpha^i + \mu_\Upsilon$
  - 4: **while**  $\epsilon^i > \epsilon_f$  **do**
  - 5:   Update iteration:  $i = i + 1$
  - 6:    $\tilde{\mathbf{Q}} = \mathbf{Q} \leftarrow \mathbf{M}^i$  (NN using k-d tree)
  - 7:    $\tilde{\mathbf{Q}}' = \{\tilde{\mathbf{Q}} \mid \|\tilde{\mathbf{Q}}_j^i - \mathbf{M}_j^i\|_2 < \bar{\mathbf{d}} + 3\sigma_d\}$
  - 8:    $\mathbf{Q}_r = \tilde{\mathbf{Q}}' \mathbf{R} + \mathbf{t}$  (Registration step)
  - 9:    $\mathbf{U}^* \leftarrow \{\mathbf{U} \mid \text{rows of } \mathbf{U} \text{ correspond to } \tilde{\mathbf{Q}}'\}$
  - 10:    $\alpha^i \leftarrow \min_{\alpha^i} \|\mathbf{U}^* \alpha^i + \mu_\Upsilon - \mathbf{q}_r\|_2 + \lambda \|\alpha^i - \alpha^{i-1}\|_2$
  - 11:    $\mathbf{m}^i = \mathbf{U}\alpha^i + \mu_\Upsilon$
  - 12:    $\epsilon^i = \|\mathbf{m}^i - \mathbf{q}_r\|_2$
  - 13: **end while**
  - 14: **return**  $\mathbf{Q}_r, \hat{\alpha}, \mathbf{m}$
- 

From a practical perspective, there is usually a need to augment an existing dense correspondence model with new 3D faces. In the following, we present a K3DM augmentation algorithm to achieve this objective. Given the K3DM and a batch of  $M$  new 3D faces, we compute the bending energy required to deform the mean face of the K3DM to

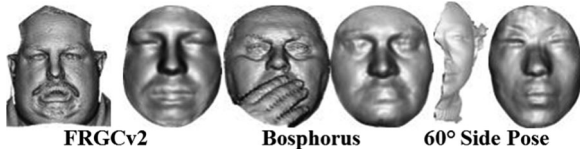


Fig. 9. K3DM fitting results on three datasets. The first scan for each dataset is the raw input while the second scan is the fitted model. The 60 degree side pose scan has been rotated to highlight the partial data.

each of the new faces. This information is employed to organize the  $M$  faces in a Minimum Spanning Tree as outlined in Section 3.1. Traversing from the root node (mean face), the K3DM is morphed into each child node using the model fitting procedure given in Algorithm 1. The resulting corresponded 3D face of the input identity is added to the K3DM. Algorithm 2 gives the details of our model augmentation technique.

#### Algorithm 2. K3DM Augmentation

**Require:**  $\Upsilon = [\tilde{\mathbf{f}}_1, \tilde{\mathbf{f}}_2, \dots, \tilde{\mathbf{f}}_N]$  and a batch of input 3D faces  $F_M = \{\mathbf{f}_1, \mathbf{f}_2, \dots, \mathbf{f}_M\}$ ,  $M \geq 1$ .

**Initialization:**

- 1: Pre-organize the  $M$  faces in a Minimum Spanning Tree  $\Pi = (V_t, E_t)$
- 2: **for** each 3D face  $\mathbf{f}_i$  in  $\Pi$  **do**
- 3:  $\tilde{\mathbf{f}}_i = \text{fit\_K3DM}(\Upsilon, \mathbf{f}_i)$
- 4:  $\Upsilon = [\tilde{\mathbf{f}}_1, \tilde{\mathbf{f}}_2, \dots, \tilde{\mathbf{f}}_N, \mathbf{f}_i]$
- 5: Increment number of faces in the model
- 6:  $\mu_\Upsilon = \frac{1}{N} \sum_{n=1}^N n$
- 7: **end for**
- 8: **return**  $\Upsilon = [\tilde{\mathbf{f}}_1, \tilde{\mathbf{f}}_2, \dots, \tilde{\mathbf{f}}_{N+M}]$

## 5 EXPERIMENTAL SETUP

We have carried out extensive experiments on synthetic and real data. Below are the details of the datasets used, evaluation criteria and the experiments performed.

### 5.1 Datasets Used

Our synthetic dataset consists of 100 3D faces generated from the Facegen software.<sup>1</sup> Facegen has been used by scientists in the field of neuroscience and social cognition to generate synthetic faces for replicating human stimuli [59], [60]. The 100 faces are in perfect correspondence with each other and hence provide the ground truth. Each face has 3,727 vertices and 7,179 triangles. For experiments on real 3D faces, we used the FRGCv2 [24], Bosphorus [25], BU3DFE [31] and side pose scans of the UND Ear database Collections F [61] and G [62]. Some sample images and details of these datasets are given in Fig. 10. The purpose of using such diverse datasets was to evaluate the performance of our proposed technique for partial data, occlusion, expression and pose invariance.

### 5.2 Evaluation Criteria

Fig. 11 shows qualitative results of our dense correspondence algorithm. The smooth transition between different faces is indicative of accurate correspondences [4], [20]. We have included a video of morphings in the supplementary

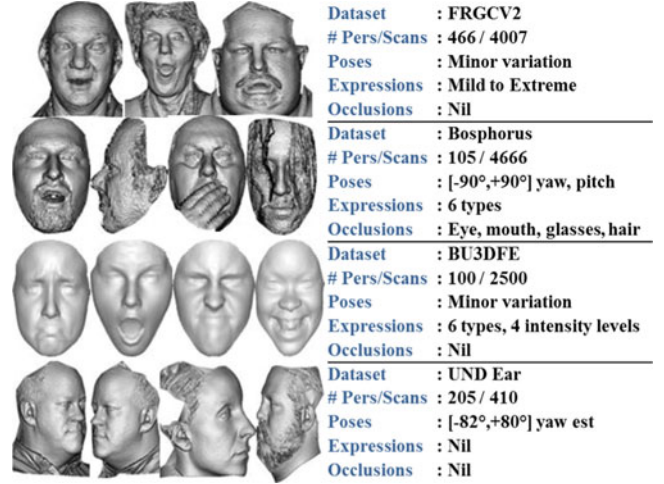


Fig. 10. Sample images and details of our four experimental datasets.

material, which can be found on the Computer Society Digital Library at <http://doi.ieeecomputersociety.org/10.1109/TPAMI.2017.2725279>.

Objective evaluation of dense correspondence algorithms on real data is difficult due to the unavailability of the ground-truth shape correspondences [63]. One solution is to use synthetic data where correspondences are known a priori. We used the synthetic 3D face dataset as ground truth for our evaluations. To the best of our knowledge, this is the first time synthetic 3D face images have been used to evaluate results of a dense correspondence algorithm in terms of mean localization error of the correspondences. This dataset and protocol was also used to evaluate the efficacy of individual modules of our algorithm.

In the case of real data, the accuracy of the dense correspondence can be measured together with the deformable model fitting algorithm by measuring the accuracy of landmark localization and face recognition. Results are expected to be better when the underlying models have accurate dense correspondences. Hence, we used our dense correspondence models and fitting algorithm in these applications and evaluated the results. In all tables, we have highlighted the best and the second best result in that category. Note that the main focus of this paper is to propose a dense 3D face correspondence algorithm. Experiments on landmark localization and face recognition have been carried out to validate the accuracy of the correspondences.

We create separate dense correspondence models from the FRGCv2, Bosphorus and BU3DFE datasets and denote



Fig. 11. Qualitative results of our dense correspondence algorithm on the first three identities of FRGCv2. The first face in each row is the source and the last face is the target.

1. Singular Inversions, "Facegen Modeller", [www.facegen.com](http://www.facegen.com)

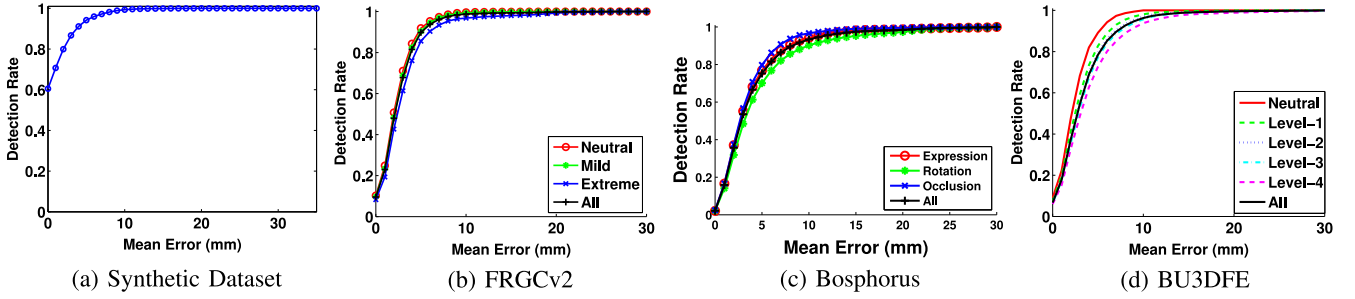


Fig. 12. Results of dense correspondence: (a-d) Cumulative localization error distribution plots on the Synthetic (2,246 vertices), FRGCv2 (14 landmarks), Bosphorus (14 landmarks) and BU3DFE (12 landmarks) datasets.

them with  $K3DM_{FR}$ ,  $K3DM_{BO}$  and  $K3DM_{BU}$  respectively. We compare our results to the Basel Face Model (BFM) proposed by Paysan et al. [40]. We also establish dense correspondence using our NICP variant for densifying the initial keypoints based feature matches (see Section 3.6). The algorithm is initialized by the correspondences found in Section 3.5 and the variant is referred as  $K3DM$ -NICP.

## 6 RESULTS AND ANALYSIS

### 6.1 Landmark Localization

*Synthetic Dataset.* First, we present the evaluation of our algorithm on synthetic data. We establish dense correspondence on 100 synthetic 3D faces using our proposed algorithm and report the mean and the standard deviation (SD) of the localization error with respect to the ground truth. The original synthetic dataset contains 3,727 vertices for each 3D face. Our proposed method is successful in establishing dense correspondence over 2,246 vertices (60 percent of the original) with a mean localization error of 1.28 mm and SD  $\pm 2.2$  mm. Correspondence within 10 mm is established on 99.33 percent vertices. Fig. 12a shows a plot of the cumulative distribution of correspondences within a given error distance. We also establish dense correspondence over 2,341 vertices of the synthetic dataset using our  $K3DM$ -NICP variant (see Section 3.6). That method results in a mean localization error of 1.30 mm with  $\pm 2.3$  mm standard deviation.

To ascertain the contribution of different components, we repeat our experiments by removing different components (see Fig. 1) from our algorithm. The results in Table 1 show that the combination of all components/modules gives the best results.

*FRGCv2 Dataset.* We construct a dense correspondence  $K3DM$  from the first neutral scan of the first 200 identities (100 males and females each) of this dataset. The

remaining 1,956 scans of 266 identities are used as test data. Next, we construct a  $K3DM$  from the neutral scans of the next 200 identities (100 male and female each) and use the 2,051 scans corresponding to the first training set for testing. This way, we are able to perform landmark detection on all 4,007 scans of FRGCv2, each time ensuring that the identity used for making the  $K3DM$  is not present in the test data.

We establish dense correspondences between 9,309 vertices on the FRGCv2 dataset ( $K3DM_{FR}$ ) and report the mean and SD of the Landmark Localization Error ( $\epsilon_L$ ) on 14 fiducial points considered to be biologically significant [65]. These anthropometric landmarks are annotated only on the mean face and transferred to each densely corresponded scan in the dataset. Manual annotations provided by Szeptycki et al. [66] and Creusot et al. [27] were used as ground truth for comparison.

A comparison of the mean and SD of landmark localization error of our proposed algorithm with the state-of-the-art in Table 2 shows that our results outperform them by a significant margin.  $K3DM_{BU}$  was constructed from 100 neutral expression scans and 100 angry expression level-1 scans.  $K3DM_{BO}$  was constructed from the first neutral scan of 105 identities. The  $K3DM_{FR}$  achieves the best performance and even the cross domain  $K3DM$ s and the  $K3DM$ -NICP variant outperform existing state-of-the-art. Cumulative localization error plots using  $K3DM_{FR}$  are shown graphically in Fig. 12b.

*Bosphorus Dataset.* We construct two  $K3DM_{BO}$  (100 faces each) from the neutral scans of the Bosphorus dataset [25] such that the model and test identities are mutually exclusive. Note that there are 299 neutral expression scans in the dataset. We manually annotate 14 fiducial landmarks on the mean face of  $K3DM$  and transfer the information to other scans after model fitting. Fig. 12c shows the cumulative detection rate of the 14 landmarks. Table 3 details landmark localization results on the three categories of the Bosphorus dataset. It is evident that our algorithm performs significantly better than the state-of-the-art under occlusions, rotation and expression variation. Creusot et al. [27] and Sukno et al. [64] trained their algorithms on 99 neutral scans. They did not report results on these 99 scans and the scans with yaw rotation of  $\pm 90^\circ$ . On the contrary we report the landmarking results on all 4,666 scans of the database including the scans with large yaw variation. Landmark annotations provided by [25], [27] were used as ground truth. For this experiment,  $K3DM_{FR}$  was created from the neutral scans of first 100 male and female (each)

TABLE 1  
Module Wise Mean and SD of Localization Error (mm) on 2,246 Vertices of the Synthetic Dataset

Excluded Module(s)	mean $\pm$ std
Organising faces into a graph	2.16 $\pm$ 2.8
Keypoint detection	3.06 $\pm$ 5.1
Feature matching	3.61 $\pm$ 6.8
Keypoint detection and feature matching	4.78 $\pm$ 7.3
Selecting best matches in each iteration	2.61 $\pm$ 3.4
No modules excluded	1.28 $\pm$ 2.2

TABLE 2  
Comparative Results of the Mean and SD (mm) of Landmark Localization Error on FRGCv2 Dataset

Author	Images	Ex(L)	Ex(R)	En(L)	En(R)	N	Prn	Ac*	Ch*	Ls	Li	Pg	Sn	Mean
Lu [36]	676	9.5 ± 17.1	10.3 ± 18.1	8.2 ± 17.2	8.3 ± 17.2	-	8.3 ± 19.4	-	6.1 ± 17.4	-	-	-	-	8.1 ± 17.7
Segundo [37]	4007	-	-	3.7 ± 2.3	3.4 ± 2.3	-	2.8 ± 1.4	5.3 ± 1.9	-	-	-	-	-	4.1 ± 1.9
Perakis [26]	975	5.6 ± 3.1	5.8 ± 3.4	4.2 ± 2.2	4.4 ± 2.5	-	-	4.1 ± 2.2	5.5 ± 2.4	-	-	4.9 ± 3.7	-	5.0 ± 2.7
Cruesot [27]	4007	5.9 ± 3.1	6.0 ± 3.0	4.3 ± 2.4	4.3 ± 2.0	4.2 ± 2.0	3.4 ± 2.0	4.8 ± 3.6	5.5 ± 3.5	4.2 ± 3.2	5.5 ± 3.9	7.3 ± 7.4	3.7 ± 3.1	5.0 ± 3.3
Perakis [38]	975	4.7	5.4	4.0	4.1	3.7	4.3	-	4.1	-	-	-	-	4.3
Sukno [64]	4007	4.7 ± 2.7	4.6 ± 2.7	3.5 ± 1.7	3.6 ± 1.7	<b>2.5 ± 1.6</b>	2.3 ± 1.7	<b>2.6 ± 1.4</b>	3.9 ± 2.8	3.3 ± 1.8	4.6 ± 3.4	4.9 ± 3.5	<b>2.7 ± 1.1</b>	3.5 ± 2.4
Gilani [28]	4007	4.5 ± 2.9	3.7 ± 2.8	3.1 ± 2.1	2.7 ± 2.1	3.6 ± 2.0	2.7 ± 2.5	4.2 ± 3.2	4.8 ± 2.1	3.3 ± 3.7	4.0 ± 3.8	4.2 ± 3.3	4.1 ± 3.1	3.9 ± 2.8
BFM [40]	4007	2.2 ± 2.5	2.7 ± 1.8	2.5 ± 2.1	2.9 ± 2.2	3.2 ± 2.2	2.3 ± 2.0	8.3 ± 2.9	<b>2.6 ± 2.9</b>	2.6 ± 2.2	<b>3.8 ± 3.7</b>	4.2 ± 3.8	3.8 ± 3.6	3.7 ± 2.7
K3DM-NICP	4007	2.8 ± 2.2	2.5 ± 1.8	2.7 ± 1.8	2.6 ± 1.1	2.6 ± 1.7	2.4 ± 1.9	3.3 ± 2.5	2.7 ± 1.8	2.6 ± 3.2	4.2 ± 3.4	4.2 ± 3.3	3.5 ± 1.4	3.3 ± 2.3
<b>K3DM<sub>FR</sub></b>	4007	<b>2.6 ± 2.1</b>	<b>2.4 ± 1.7</b>	<b>2.4 ± 1.6</b>	<b>2.4 ± 0.9</b>	<b>2.5 ± 1.5</b>	<b>2.2 ± 1.8</b>	<b>3.0 ± 2.4</b>	<b>2.5 ± 1.8</b>	<b>2.4 ± 3.1</b>	4.1 ± 3.3	<b>4.1 ± 3.3</b>	<b>3.4 ± 1.1</b>	<b>2.9 ± 2.1</b>
K3DM <sub>BU</sub>	4007	2.7 ± 2.4	<b>2.3 ± 1.9</b>	<b>2.4 ± 1.9</b>	<b>2.5 ± 1.8</b>	2.8 ± 1.8	2.6 ± 1.8	6.1 ± 2.7	4.2 ± 3.1	2.9 ± 3.3	4.6 ± 3.9	<b>4.1 ± 3.4</b>	3.6 ± 2.9	3.6 ± 2.6
K3DM <sub>BO</sub>	4007	2.6 ± 2.2	2.4 ± 1.9	2.8 ± 2.0	2.9 ± 2.0	3.2 ± 2.2	<b>2.3 ± 2.1</b>	8.3 ± 3.4	3.1 ± 2.7	<b>2.5 ± 2.4</b>	<b>3.5 ± 3.7</b>	4.1 ± 3.9	3.8 ± 3.6	3.8 ± 2.7

\* Results have been averaged for left and right corners of nose and mouth.

A '-' denotes that the authors have not detected this particular landmark. Ex/En-outer/inner eye corner, N-nosebridge saddle, Prn-nosetip, Ac-nose curvature, Ch-mouth corner, Ls/Li upper/lower lip midpoint, Pg-chintip, Sn-nasal base.

identities of FRGCv2 while K3DM<sub>BU</sub> was the same as used in experiment on FRGCv2 dataset.

**BU3DFE Dataset.** We construct dense correspondence models from the neutral as well as intensity level-1 anger expression scans of 100 identities of the BU3DFE dataset [31]. We ensure mutually exclusive test and training identities while landmark localization using a K3DM<sub>BU</sub>. Comparative results on 12 anthropometric landmarks [7] on all 2,500 scans of the dataset are given in Table 4. Ground truth landmark locations are provided with the dataset [31].

Fig. 12d shows the commutative error detection rate of the 12 landmarks. K3DM<sub>FR</sub> was created from the scans of the first 100 male and 100 female identities of FRGC. Our results are better than the state-of-the-art for both the models.

**UND Ear Dataset.** To evaluate the landmark localization performance of our algorithm on side pose scans containing self occlusions, we perform experiments on the UND Ear Database [61], [62]. We follow the exact protocol outlined by [18], [38] for a fair comparison. The dataset is divided into 45 and 60 degree left and right pose scans namely DB45L,

TABLE 3  
Comparison of Landmark Localization Results with the State-of-the-Art on Bosphorus Dataset

Mean of Localization Error (mm)																	
	Author	Images	Ex(L)	Ex(R)	En(L)	En(R)	N	Prn	Ac(L)	Ac(R)	Ch(L)	Ch(R)	Ls	Li	Pg	Sn	Mean
Expression	Cruesot et al. [27]	2803	6.20	4.10	5.09	4.18	6.33	4.47	4.22	4.07	4.06	8.00	7.66	5.36	8.83	15.23	6.27
	Sukno et al. [64]	2803	5.19	4.92	2.94	2.76	2.22	2.33	3.03	3.01	6.12	6.03	4.00	6.54	7.58	2.81	4.25
	BFM [40]	2920	3.58	3.62	2.77	2.65	2.17	2.90	3.80	4.63	5.86	6.01	3.99	6.77	8.12	3.59	4.32
	<b>K3DM<sub>BO</sub></b>	2920	3.57	4.01	2.35	2.40	2.32	2.82	2.50	2.99	4.85	4.91	3.32	5.03	6.02	2.35	3.53
Rotation	Cruesot et al. [27]	1155	5.42	4.12	5.18	3.65	5.17	4.89	3.52	3.43	4.05	4.29	3.84	3.81	4.68	9.47	4.68
	Sukno et al. [64]	1155	4.48	4.95	2.97	3.23	3.40	4.36	3.36	3.37	3.76	3.75	3.47	5.01	7.77	4.19	4.15
	BFM [40]	1365	4.63	4.96	5.30	5.16	3.81	5.08	4.81	5.49	4.49	5.28	5.43	6.40	7.10	3.14	5.08
	<b>K3DM<sub>BO</sub></b>	1365	4.84	5.09	3.31	3.85	2.68	3.19	2.73	3.20	4.53	4.91	4.13	5.84	6.22	3.80	4.14
Occlusion	Cruesot et al. [27]	381	8.13	5.45	5.60	4.99	7.78	4.72	5.34	4.85	4.10	5.62	4.81	4.30	5.44	11.05	5.87
	Sukno et al. [64]	381	6.63	6.28	3.82	3.87	4.12	3.83	4.40	4.67	4.75	5.07	3.61	4.81	7.63	3.76	4.80
	BFM [40]	381	4.95	4.42	3.96	3.52	2.49	3.32	4.57	4.77	3.61	3.75	3.36	4.40	5.54	2.45	3.94
	<b>K3DM<sub>BO</sub></b>	381	4.64	4.51	3.10	2.95	2.69	3.18	2.55	3.01	4.36	4.22	2.89	4.14	5.00	2.90	3.58
All	Cruesot et al. [27]	4339	6.09	4.18	5.14	4.08	6.10	4.60	4.15	3.94	<b>4.05</b>	6.83	6.37	<b>4.81</b>	7.35	13.20	5.78
	Sukno et al. [64]	4339	5.13	5.05	<b>3.03</b>	<b>2.98</b>	2.70	3.00	<b>3.24</b>	<b>3.25</b>	5.37	<b>5.34</b>	<b>3.82</b>	5.98	7.63	3.26	4.27
	BFM [40]	4666	<b>3.93</b>	<b>4.03</b>	3.41	3.34	2.68	3.57	4.07	4.86	5.34	5.65	4.37	6.50	7.64	3.38	4.48
	<b>K3DM<sub>BO</sub></b>	4666	<b>3.94</b>	<b>4.15</b>	<b>2.62</b>	<b>2.80</b>	<b>2.46</b>	<b>2.96</b>	<b>2.55</b>	<b>3.04</b>	<b>4.73</b>	<b>4.86</b>	<b>3.53</b>	<b>5.21</b>	<b>6.01</b>	<b>2.75</b>	<b>3.70</b>
	K3DM <sub>BU</sub>	4666	4.04	4.25	3.21	3.12	<b>2.50</b>	3.27	3.65	4.34	5.16	5.45	4.25	6.25	<b>7.26</b>	<b>3.16</b>	<b>4.27</b>
	K3DM <sub>FR</sub>	4666	4.13	4.27	3.33	3.24	2.60	3.51	3.91	4.61	5.26	5.56	4.32	6.43	7.48	3.26	4.42
Standard Deviation of Localization Error (mm)																	
	Images	Ex(L)	Ex(R)	En(L)	En(R)	N	Prn	Ac(L)	Ac(R)	Ch(L)	Ch(R)	Ls	Li	Pg	Sn	Mean	
All	Cruesot et al. [27]	4339	5.02	3.79	4.43	3.49	5.22	4.61	3.45	3.11	<b>2.95</b>	5.35	5.17	<b>3.95</b>	8.36	10.37	4.95
	Sukno et al. [64]	4339	4.01	3.86	<b>2.15</b>	<b>2.33</b>	<b>2.27</b>	<b>2.56</b>	2.37	<b>2.42</b>	5.06	4.75	<b>3.51</b>	6.86	7.16	<b>2.37</b>	3.69
	BFM [40]	4666	<b>2.84</b>	<b>2.97</b>	3.30	3.60	2.58	3.44	2.63	2.86	4.23	4.22	3.93	6.76	6.98	3.09	3.82
	<b>K3DM<sub>BO</sub></b>	4666	<b>2.69</b>	<b>2.82</b>	<b>2.06</b>	<b>2.23</b>	<b>1.63</b>	<b>1.61</b>	<b>1.45</b>	<b>1.59</b>	<b>3.13</b>	<b>3.03</b>	<b>2.99</b>	<b>4.23</b>	<b>4.28</b>	<b>2.23</b>	<b>2.57</b>
	K3DM <sub>BU</sub>	4666	2.87	3.01	2.98	3.17	2.42	3.04	<b>2.36</b>	2.56	4.01	<b>3.99</b>	3.75	6.13	<b>6.20</b>	2.88	<b>3.53</b>
	K3DM <sub>FR</sub>	4666	2.93	3.07	3.10	3.30	2.52	3.26	2.54	2.72	4.10	4.07	3.80	6.31	6.39	2.97	3.65

TABLE 4  
Comparison of Landmark Localization Results (Mean  $\pm$  SD) with the State-of-the-Art on BU3DFE Dataset

Author	Images	Ex(L)	En(L)	N	Ex(R)	En(R)	Prn	Ac	Ch	Ls	Li	Mean
Nair et al. [8]	2350	-	12.1	-	-	11.9	8.8	-	-	-	-	10.9
Segundo et al. [37]	2500	-	6.3 $\pm$ 4.8	-	-	6.3 $\pm$ 5.0	1.9 $\pm$ 1.1	6.6 $\pm$ 3.4	-	-	-	4.4 $\pm$ 3.5
Salazar et al. [35]	350	9.6 $\pm$ 6.1	6.8 $\pm$ 4.5	-	8.5 $\pm$ 5.8	6.1 $\pm$ 4.2	5.9 $\pm$ 2.7	6.8 $\pm$ 3.2	-	-	-	5.9 $\pm$ 4.3
Gilani et al. [28]	2500	4.4 $\pm$ 2.7	4.8 $\pm$ 2.6	4.5 $\pm$ 2.7	4.4 $\pm$ 2.7	3.3 $\pm$ 2.7	2.9 $\pm$ 2.0	4.3 $\pm$ 2.7	5.7 $\pm$ 3.7	4.2 $\pm$ 2.7	6.9 $\pm$ 6.3	3.7 $\pm$ 3.1
K3DM <sub>BU</sub>	2500	3.8 $\pm$ 2.2	2.2 $\pm$ 1.5	2.9 $\pm$ 2.1	3.3 $\pm$ 2.2	2.4 $\pm$ 1.6	2.5 $\pm$ 1.7	2.3 $\pm$ 1.6	4.6 $\pm$ 3.3	3.6 $\pm$ 2.3	6.4 $\pm$ 6.1	2.8 $\pm$ 2.5
K3DM <sub>FR</sub>	2500	4.0 $\pm$ 2.4	2.8 $\pm$ 1.6	4.3 $\pm$ 2.6	3.6 $\pm$ 2.4	2.7 $\pm$ 1.7	2.6 $\pm$ 1.8	2.9 $\pm$ 1.8	5.4 $\pm$ 3.5	3.8 $\pm$ 2.4	7.0 $\pm$ 6.6	3.1 $\pm$ 2.7

DB45R, DB60L and DB60R. This is a very challenging dataset due to large yaw rotations, noisy scans and self occlusions. The dense correspondence model is created from 200 neutral expression scans of FRGCv2. Eight landmarks including the two inner and outer eye corners, nose tip, mouth corners and chin tip are annotated on the mean face of K3DM<sub>FR</sub>. The mean and SD of landmark localization error for all 8 points is compared with the state-of-the-art in Table 5.

## 6.2 Face Recognition

Imaging of faces is considered to be one of the most important biometrics because it can be done passively and is highly distinctive between individuals. 3D face recognition has addressed many shortcomings of its counterpart in the 2D domain [67]. We consider this application apt to test the quality of the presented algorithms. Note that our main aim is to evaluate our proposed correspondence and model fitting algorithms as opposed to presenting a face recognition system *per se*.

**FRGCv2 Dataset.** We follow the FRGCv2 protocols [24] of face recognition and include only one scan of each individual (466 in total) in the gallery. To demonstrate the effectiveness of our K3DM Augmentation algorithm, we construct a dense correspondence model on the first available neutral scans of the first 200 identities of FRGCv2 dataset. The constructed model is then augmented with first available neutral scans of the remaining 266 identities of the database using Algorithm 2. The probe set consists of the remaining (3,541) scans of all identities. Note that there is only one scan per identity for 56 individuals in the dataset. All of these identities appear only in the gallery. The complete dataset was further classified into “neutral” and “non-neutral” expression subclasses following the protocol outlined in [44] to evaluate the effects of expressions on deformable model fitting and face recognition.

We employ a holistic and region based approach to model fitting and face recognition. It is well known that the generalization of a model can be increased by dividing faces into independent subregions that are morphed

independently [29]. This technique has been used extensively for face recognition [9], [67] and recently for matching offsprings to their parents [68]. We also use this approach and perform face recognition by morphing the complete face as well as the eyes and nose regions. We define these regions on the mean face which is sufficient to transfer the information to all the faces in the dense correspondence model.

The full K3DM<sub>FR</sub> and the eyes and nose models are separately morphed and fitted to each query face in the probe to obtain model parameters ( $\alpha$ -Step 12 in Algorithm 1). Next, the parameters from the whole face and the regions are concatenated to form the feature vector for face recognition. We then perform feature selection using the GEFS algorithm [69], [70] on the training data set of FRGCv2 containing 953 facial scans. Note that these scans are not used in testing the face recognition algorithm. The selected features of each query face are matched with those of the gallery faces in the model. The query face is assigned the identity of the gallery face with which it has the smallest distance  $d_f = \cos^{-1} \frac{\tilde{\alpha}_M^T \tilde{\alpha}_Q}{\|\tilde{\alpha}_M\|_2 \|\tilde{\alpha}_Q\|_2}$ , where  $\tilde{\alpha}_M$  are the selected features of each face in K3DM and  $\tilde{\alpha}_Q$  are the selected features of the query face.

Fig. 14 shows the process of model fitting in PCA space. The dense correspondence model is iteratively fitted on the query face, which in the figure is an extreme expression scan of the first identity. The model fitting starts from the mean face and in each iteration the fitted query model traverses closer to its gallery face in the PCA space. Face recognition is performed when the fitting residual error  $\epsilon_f$  is less than  $10^{-5}$ . Figs. 13a and b show the resulting CMC and ROC curves. Rank-1 identification rate for neutral probes is 99.85 percent while 100 percent accuracy is achieved at Rank-8. In the more difficult scenario of neutral versus non-neutral, the Rank-1 identification rate is 96.3 percent. A similar trend is observed in the verification rates at 0.1 percent FAR. Table 6

TABLE 5  
Comparative Landmark Localisation Results (mm) on UND Side Pose Scans

Database	DB45L	DB45R	DB60L	DB60R
Yaw Est [26]	-45° $\pm$ 9°	44° $\pm$ 8°	-59° $\pm$ 8°	57° $\pm$ 7°
# Scans	118	118	87	87
Passalis et al. [18]	6.02 $\pm$ 2.45	5.83 $\pm$ 2.49	6.08 $\pm$ 2.53	5.87 $\pm$ 2.4
Perakis et al. [26]	4.75 $\pm$ 1.91	5.03 $\pm$ 1.92	5.30 $\pm$ 2.49	4.95 $\pm$ 1.80
K3DM <sub>FR</sub>	4.04 $\pm$ 1.77	4.31 $\pm$ 1.90	4.36 $\pm$ 2.25	4.24 $\pm$ 1.28

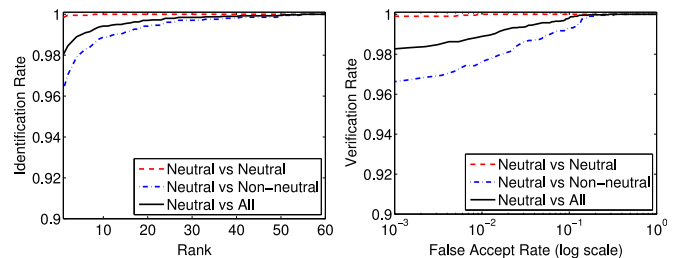


Fig. 13. (Left) ROC curves for identification and (Right) verification tasks on FRGCv2 database using our dense correspondence model and fitting algorithms.

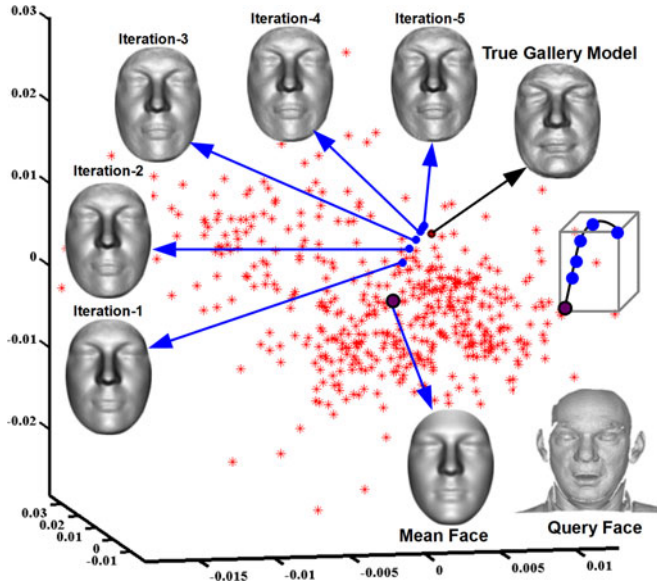


Fig. 14. Iterative model fitting. The 466 FRGCv2 identities are shown as red stars in the first three PC space. The model is morphed iteratively into the query face until the residual error is negligible. Notice how the fitting process takes the query face through a non-linear path (inset image) and removes the extreme facial expression to generate its equivalent neutral expression model.

compares our algorithm with the state-of-the-art. In most cases, our results are better than the state-of-the-art depicting the high quality of the dense correspondence model.

**Bosphorus Dataset.** Experiments are performed on the more versatile Bosphorus dataset to demonstrate the expression, pose and occlusion invariant face recognition capabilities of our proposed model.  $K3DM_{BO}$  is formed from the first available neutral scan of each identity in the dataset and a holistic approach to face recognition is adopted. We follow the model fitting and parameter matching technique as mentioned for FRGCv2. Comparative results are given in Table 7. Our proposed technique significantly outperforms the state-of-the-art in pose invariant face recognition, while at the same time it handles expressions and occlusions.

**UND Ear Dataset.** We perform face recognition experiments on this dataset to demonstrate the ability of  $K3DM$  to handle pose variations and self occlusions. The dataset is divided into three subsets following the protocol set by [18].  $UND00LR$  contains 466 subjects of FRGCv2 in the gallery.

TABLE 6  
Comparison of 3D Face Recognition Results with the State-of-the-Art in Terms of Rank-1 Identification Rate (I-Rate) and Verification Rate (V-Rate) at 0.1 Percent FAR

Author	Neutral		Non-neutral		All	
	I-Rate	V-Rate	I-Rate	V-Rate	I-Rate	V-Rate
Mian et al. [44]	<b>99.4%</b>	<b>99.9%</b>	92.1%	96.6%	96.1%	<b>98.6%</b>
Kakadiaris et al. [43]	-	99.0%	-	95.6%	97.0%	97.3%
Al-Osaimi et al. [71]	97.6%	98.4%	95.2%	<b>97.8%</b>	96.5%	98.1%
Queirolo et al. [72]	-	99.5%	-	94.8%	<b>98.4%</b>	96.6%
Drira et al. [73]	99.2%	-	<b>96.8%</b>	-	97.7%	97.1%
Smeets et al. [45]	-	-	-	-	89.6%	79.0%
Li et al. [47]	-	-	-	-	96.3%	-
<b><math>K3DM_{FR}</math></b>	<b>99.9%</b>	<b>99.9%</b>	<b>96.9%</b>	<b>96.6%</b>	<b>98.5%</b>	<b>98.7%</b>

Two 45 degree side scans each (left and right) for 39 subjects and two 60 degree side scans each (left and right) for 32 subjects make the probe set. These subjects are common between FRGCv2 and UND Ear databases.  $UND45LR$  is composed of 45 degree side scans from 118 subjects. The  $K3DM_{FR}$  made from 200 scans is fitted on the left side scan to get the gallery parameters and then fitted to the right side scan to get the probe parameters. A similar protocol is followed for  $UND60LR$  which contains 60 degree side scans from 87 subjects. Comparative Rank-1 recognition results are given in Table 8. Note that while Smeets et al. [45] report > 98 percent face recognition results on the side pose scans of this dataset, their performance on pose variation in the Bosphorus dataset is significantly low at 84.2 percent.

**Cross Domain Face Recognition.** To compare  $K3DM$  with the state-of-the-art Basel Face Model (BFM) [40] we perform cross domain face recognition experiments on FRGCv2 and Bosphorus datasets as they include all the challenges of expressions, occlusions and pose variation. For FRGCv2 we use  $K3DM_{BO}$  (created from 105 neutral scans of Bosphorus database) and  $K3DM_{BU}$  (created from 100 neutral and 100 angry level-1 scans) while for Bosphorus dataset we use  $K3DM_{FR}$  and  $K3DM_{BU}$ . All three models are fitted to each scan in FRGCv2 and Bosphorus datasets. The model parameters of the first neutral scan of each identity in each database are used as gallery features. Table 9 details the Rank-1 recognition results from this experiment which show that  $K3DM$  outperforms the BFM.

TABLE 7  
Comparison of Rank-1 Recognition Results (in Percentage) with the State-of-the-Art on Bosphorus Dataset

Author	Expressions			Poses					Occlusions					All 4,543
	AU	Expr	All	YR < 90	YR90	PR	CR	All	Eye	Mouth	Glasses	Hair	All	
	2,150	647	2,797	525	210	419	211	1,365	105	105	104	67	381	
Alyz et al. [74]	-	-	-	-	-	-	-	-	93.6	93.6	97.8	89.6	93.6	-
Colombo et al. [75]	-	-	-	-	-	-	-	-	91.1	74.7	94.2	90.4	87.6	-
Drira et al. [73]	-	-	-	-	-	-	-	-	97.1	78.0	94.2	81.0	87.0	-
Berretti et al. [46]	-	-	95.7	81.6	45.7	98.3	93.4	88.6	-	-	-	-	93.2	93.4
Smeets et al. [45]	-	-	97.7	-	24.3	-	-	84.2	-	-	-	-	-	93.7
Li et al. [47]	<b>99.2</b>	<b>96.6</b>	<b>98.8</b>	<b>84.1</b>	<b>47.1</b>	<b>99.5</b>	<b>99.1</b>	<b>91.1</b>	<b>100.0</b>	<b>100.0</b>	<b>100.0</b>	<b>95.5</b>	<b>99.2</b>	<b>96.6</b>
<b><math>K3DM_{BO}</math></b>	<b>99.0</b>	<b>96.7</b>	<b>98.5</b>	<b>99.8</b>	<b>95.2</b>	<b>100.0</b>	<b>99.1</b>	<b>99.0</b>	<b>99.0</b>	<b>96.1</b>	<b>100.0</b>	<b>97.3</b>	<b>98.1</b>	<b>98.6</b>

AU=Action Units; YR=Yaw Rotation; PR= Pitch Rotation; CR= Cross Rotation.

TABLE 8  
Comparative of Rank-1 Recognition Results on  
Partial Faces of UND Side Pose Scans

Database	UND00LR	UND45LR	UND60LR
# Scans	608	236	174
Passalis et al. [18]	76.8%	86.4%	81.6%
Smeets et al. [45]	-	98.3%	100.0%
K3DM <sub>FR</sub>	86.0%	95.8%	98.6%

TABLE 9  
Comparison of Rank-1 Recognition on FRGCv2 and  
Bosphorus Datasets Using Cross Domain Models

FRGCv2					
Method	Neutral	Expressions	Poses	Occlusions	All
BFM [40]	87.7%	65.6%	-	-	76.4%
K3DM <sub>BU</sub>	92.7%	69.0%	-	-	80.5%
K3DM <sub>BO</sub>	92.1%	62.9%	-	-	77.1%
Bosphorus					
BFM [40]	-	81.1%	86.1%	86.6%	82.7%
K3DM <sub>FR</sub>	-	85.6%	86.5%	89.3%	85.8%
K3DM <sub>BU</sub>	-	90.3%	92.8%	90.7%	90.7%

## 7 DISCUSSION AND CONCLUSIONS

We have proposed an algorithm that simultaneously establishes dense correspondences between a large number of 3D faces. Based on the dense correspondences, a deformable face model was constructed. We also proposed morphable model fitting and updating algorithms that are useful for landmark identification and face recognition. Thorough experiments were performed on synthetic and real 3D faces. Comparison with existing state-of-the-art shows that our algorithm consistently achieves better or comparable performance on both the tasks on all datasets. It is interesting to note that while the face recognition algorithm proposed by Li et al. [47] performs well on Bosphorus database, it does not fare that well on FRGCv2. Similar trend can be observed in case of Smeets et al. [45] for face recognition and Sukno et al. [64] for landmark localization. To the best of our knowledge this is the first paper that has reported consistent comparable results on a variety of application on four public datasets.

Although the dense correspondence model assumes frontal and neutral pose scans, for landmark localization and face recognition it demonstrates robustness to occlusion as well as pose and expression variation during the fitting process. Hence the three proposed algorithms present a unified solution to a variety of applications under expression, occlusion and pose variation. The model can handle pose variation up to  $\pm 90^\circ$ .

With regards to the computational complexity, it may be noted that the model building process has to be done offline. The algorithm iterates over geodesic patches between vertices for each image. Building a dense correspondence model on 105 identities of Bosphorus database in approximately 30 iterations took over 48 hours on a Core-i7 machine with 8 GB RAM using MATLAB. However, the model fitting process on an unseen face takes less than seven seconds.

## ACKNOWLEDGMENTS

This research was supported by ARC Discovery Grants DP110102399 and DP160101458. I. Reid gratefully acknowledges the financial support of the Australian Research Council through grants CE140100016 and FL130100102.

## REFERENCES

- [1] O. Van Kaick, H. Zhang, G. Hamarneh, and D. Cohen-Or, "A survey on shape correspondence," *Comput. Graph. Forum*, vol. 30, no. 6, pp. 1681–1707, 2011.
- [2] V. Jain, H. Zhang, and O. van Kaick, "Non-rigid spectral correspondence of triangle meshes," *Int. J. Shape Model.*, vol. 13 no. 1, pp. 101–124, 2007.
- [3] W. Chang and M. Zwicker, "Automatic registration for articulated shapes," *Comput. Graph. Forum*, vol. 27, no. 5, pp. 1459–1468, 2008.
- [4] H. Zhang, A. Sheffer, D. Cohen, Q. Zhou, O. van Kaick, and A. Tagliasacchi, "Deformation-driven shape correspondence," *Comput. Graph. Forum*, vol. 27, no. 5, pp. 1431–1439, 2008.
- [5] S. Z. Gilani, K. Rooney, F. Shafait, M. Walters, and A. Mian, "Geometric facial gender scoring: Objectivity of perception," *PLoS One*, vol. 9, no. 6, 2014, Art. no. e99483.
- [6] P. Hammond, "The use of 3D face shape modelling in dysmorphology," *Archives Disease Childhood*, vol. 92, no. 12, pp. 1120–1126, 2007.
- [7] L. Farkas, "Anthropometry of the head and face in clinical practice," in *Anthropometry of the Head and Face*, 2nd ed. Ann Arbor, MI, USA: Univ. Michigan, 1994, pp. 71–111.
- [8] P. Nair and A. Cavallaro, "3-D face detection, landmark localization, and registration using a point distribution model," *IEEE Trans. Multimedia*, vol. 11, no. 4, pp. 611–623, Jun. 2009.
- [9] V. Blanz and T. Vetter, "Face recognition based on fitting a 3D morphable model," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 25, no. 9, pp. 1063–1074, Sep. 2003.
- [10] S. Z. Gilani, A. Mian, and P. Eastwood, "Deep, dense and accurate 3D face correspondence for generating population specific deformable models," *Pattern Recognit.*, vol. 69, pp. 238–250, 2017.
- [11] 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 *Proc. Eur. Conf. Comput. Vis.*, 2002, pp. 3–20.
- [12] T. Heimann and H.-P. Meinzer, "Statistical shape models for 3D medical image segmentation: A review," *Med. Image Anal.*, vol. 13, no. 4, pp. 543–563, 2009.
- [13] M. Alexa, "Recent advances in mesh morphing," *Comput. Graph. Forum*, vol. 21, no. 2, pp. 173–198, 2002.
- [14] D. Aiger, N. Mitra, and D. Cohen, "4-points congruent sets for robust pairwise surface registration," *ACM Trans. Graph.*, vol. 27, no. 3, 2008, Art. no. 85.
- [15] B. Brown and S. Rusinkiewicz, "Global non-rigid alignment of 3-D scans," *ACM Trans. Graph.*, vol. 26, no. 3, 2007, Art. no. 21.
- [16] H. Mirzaalian, G. Hamarneh, and T. Lee, "A graph-based approach to skin mole matching incorporating template-normalized coordinates," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2009, pp. 2152–2159.
- [17] T. Funkhouser and P. Shilane, "Partial matching of 3D shapes with priority-driven search," in *Proc. Eurographics Symp. Geometry Process.*, 2006, vol. 256, pp. 131–142.
- [18] G. Passalis, P. Perakis, T. Theoharis, and I. A. Kakadiaris, "Using facial symmetry to handle pose variations in real-world 3D face recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 10, pp. 1938–1951, Oct. 2011.
- [19] U. Prabhu, J. Heo, and M. Savvides, "Unconstrained pose-invariant face recognition using 3D generic elastic models," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 10, pp. 1952–1961, Oct. 2011.
- [20] V. Kraevoy and A. Sheffer, "Cross-parameterization and compatible remeshing of 3D models," *ACM Trans. Graph.*, vol. 23, pp. 861–869, 2004.
- [21] H. Hochheiser, et al., "The facebase consortium: A comprehensive program to facilitate craniofacial research," *Develop. Biology*, vol. 355, no. 2, pp. 175–182, 2011.
- [22] A. J. Whitehouse, et al., "Prenatal testosterone exposure is related to sexually dimorphic facial morphology in adulthood," *Proc. Roy. Soc. B*, vol. 282, no. 1816, 2015, Art. no. 7.

- [23] S. Z. Gilani, et al., "Sexually dimorphic facial features vary according to level of autistic-like traits in the general population," *J. Neurodevelopmental Disorders*, vol. 7, no. 1, 2015, Art. no. 14.
- [24] P. Phillips, et al., "Overview of the face recognition grand challenge," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, 2005, pp. 947–954.
- [25] A. Savran, et al., "Bosphorus database for 3D face analysis," in *Proc. Eur. Workshop Biometrics Identity Manage.*, 2008, pp. 47–56.
- [26] P. Perakis, G. Passalis, T. Theoharis, and I. A. Kakadiaris, "3D facial landmark detection under large yaw and expression variations," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 7, pp. 1552–1564, Jul. 2013.
- [27] C. Creusot, N. Pears, and J. Austin, "A machine-learning approach to keypoint detection and landmarking on 3D meshes," *Int. J. Comput. Vis.*, vol. 102, no. 1–3, pp. 146–179, 2013.
- [28] S. Z. Gilani, F. Shafait, and A. Mian, "Shape-based automatic detection of a large number of 3D facial landmarks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2015, pp. 4639–4648.
- [29] V. Blanz and T. Vetter, "A morphable model for the synthesis of 3D faces," in *Proc. ACM Conf. Comput. Graph. Interactive Techn.*, 1999, pp. 187–194.
- [30] Y. Sun, J. Paik, A. Koschan, D. Page, and M. Abidi, "Point fingerprint: A new 3D object representation scheme," *IEEE Trans. Syst. Man Cybern. Part B: Cybern.*, vol. 33, no. 4, pp. 712–717, Aug. 2003.
- [31] L. Yin, X. Wei, Y. Sun, J. Wang, and M. J. Rosato, "A 3D facial expression database for facial behavior research," in *Proc. 7th Int. Conf. Automatic Face Gesture Recognit.*, 2006, pp. 211–216.
- [32] Y. Sun and M. A. Abidi, "Surface matching by 3D point's fingerprint," in *Proc. 8th IEEE Int. Conf. Comput. Vis.*, 2001, pp. 263–269.
- [33] S. Wang, Y. Wang, M. Jin, X. D. Gu, and D. Samaras, "Conformal geometry and its applications on 3D shape matching, recognition, and stitching," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 7, pp. 1209–1220, Jul. 2007.
- [34] J. Novatnack and K. Nishino, "Scale-dependent/invariant local 3D shape descriptors for fully automatic registration of multiple sets of range images," in *Proc. 10th Eur. Conf. Comput. Vis.*, 2008, pp. 440–453.
- [35] A. Salazar, S. Wuhner, C. Shu, and F. Prieto, "Fully automatic expression-invariant face correspondence," *Mach. Vis. Appl.*, vol. 25, no. 4, pp. 859–879, 2014.
- [36] X. Lu and A. K. Jain, "Automatic feature extraction for multiview 3D face recognition," in *Proc. 7th Int. Conf. Automatic Face Gesture Recognit.*, 2006, pp. 585–590.
- [37] M. Segundo, L. Silva, P. Bellon, and C. C. Queirolo, "Automatic face segmentation and facial landmark detection in range images," *IEEE Trans. Syst. Man Cybern. Part B: Cybern.*, vol. 40, no. 5, pp. 1319–1330, Oct. 2010.
- [38] P. Perakis, T. Theoharis, and I. A. Kakadiaris, "Feature fusion for facial landmark detection," *Pattern Recognit.*, vol. 47, no. 9, pp. 2783–2793, 2014.
- [39] V. Blanz, K. Scherbaum, and H.-P. Seidel, "Fitting a morphable model to 3D scans of faces," in *Proc. 11th IEEE Int. Conf. Comput. Vis.*, 2007, pp. 1–8.
- [40] P. Paysan, R. Knothe, B. Amberg, S. Romdhani, and T. Vetter, "A 3D face model for pose and illumination invariant face recognition," in *Proc. Int. Conf. Adv. Video Signal Based Surveillance*, 2009, pp. 296–301.
- [41] B. Amberg, S. Romdhani, and T. Vetter, "Optimal step nonrigid ICP algorithms for surface registration," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2007, pp. 1–8.
- [42] G. Passalis, I. Kakadiaris, T. Theoharis, G. Toderici, and N. Murruza, "Evaluation of 3D face recognition in the presence of facial expressions: An annotated deformable model approach," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. Workshops*, 2005, pp. 171–171.
- [43] I. A. Kakadiaris, et al., "Three-dimensional face recognition in the presence of facial expressions: An annotated deformable model approach," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 4, pp. 640–649, Apr. 2007.
- [44] A. Mian, M. Bennamoun, and R. Owens, "Keypoint detection and local feature matching for textured 3D face recognition," *Int. J. Comput. Vis.*, vol. 79, no. 1, pp. 1–12, 2008.
- [45] D. Smeets, J. Keustermans, D. Vandermeulen, and P. Suetens, "MeshSIFT: Local surface features for 3D face recognition under expression variations and partial data," *Comput. Vis. Image Understanding*, vol. 117, no. 2, pp. 158–169, 2013.
- [46] S. Berretti, N. Werghi, A. Del Bimbo, and P. Pala, "Matching 3D face scans using interest points and local histogram descriptors," *Comput. Graph.*, vol. 37, no. 5, pp. 509–525, 2013.
- [47] H. Li, D. Huang, J.-M. Morvan, Y. Wang, and L. Chen, "Towards 3D face recognition in the real: A registration-free approach using fine-grained matching of 3D keypoint descriptors," *Int. J. Comput. Vis.*, vol. 113, no. 2, pp. 128–142, 2014.
- [48] F. L. Bookstein, "Principal warps: Thin-plate splines and the decomposition of deformations," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 11, no. 6, pp. 567–585, Jun. 1989.
- [49] D. Rueckert, L. Sonoda, C. Hayes, D. L. G. Hill, M. O. Leach, and D. J. Hawkes, "Nonrigid registration using free-form deformations: Application to breast MR images," *IEEE Trans. Med. Imag.*, vol. 18, no. 8, pp. 712–721, Aug. 1999.
- [50] D.-J. Kroon, "Finite iterative closest point," *MATLAB Central File Exchange*, 2009, <https://au.mathworks.com/matlabcentral/fileexchange/24301-finite-iterative-closest-point>
- [51] Y. Guo, F. Sohel, M. Bennamoun, M. Lu, and J. Wan, "Rotational projection statistics for 3D local surface description and object recognition," *Int. J. Comput. Vis.*, vol. 105, no. 1, pp. 63–86, 2013.
- [52] F. Tombari, S. Salti, and L. Di Stefano, "Unique signatures of histograms for local surface description," in *Proc. 11th Eur. Conf. Comput. Vis.*, 2010, pp. 356–369.
- [53] R. C. Gonzalez and R. E. Woods, *Digital Image Processing*, 2nd ed. Upper Saddle River, NJ USA: Prentice Hall, 2002.
- [54] A. Mian, M. Bennamoun, and R. Owens, "On the repeatability and quality of keypoints for local feature-based 3D object retrieval from cluttered scenes," *Int. J. Comput. Vis.*, vol. 89, no. 2/3, pp. 348–361, 2010.
- [55] B. C. Munsell, A. Teml, and S. Wang, "Fast multiple shape correspondence by pre-organizing shape instances," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2009, pp. 840–847.
- [56] J. A. Sethian, "Evolution, implementation, and application of level set and fast marching methods for advancing fronts," *J. Comput. Physics*, vol. 169, no. 2, pp. 503–555, 2001.
- [57] G. Peyré, "The numerical tours of signal processing-advanced computational signal and image processing," *IEEE Comput. Sci. Eng.*, vol. 13, no. 4, pp. 94–97, Jul./Aug. 2011.
- [58] J. L. Bentley, "Multidimensional binary search trees used for associative searching," *Commun. ACM*, vol. 18, no. 9, 1975, Art. no. 509.
- [59] A. Todorov, S. Baron, and N. Oosterhof, "Evaluating face trustworthiness: A model based approach," *Social Cognitive Affect. Neuroscience*, vol. 3, no. 2, pp. 119–127, 2008.
- [60] N. N. Oosterhof and A. Todorov, "The functional basis of face evaluation," *Proc. Nat. Academy Sci. United States America*, vol. 105, no. 32, pp. 11087–11092, 2008.
- [61] P. Yan and K. Bowyer, "Empirical evaluation of advanced ear biometrics," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. Workshops*, 2005, pp. 41–41.
- [62] P. Yan and K. W. Bowyer, "An automatic 3D ear recognition system," in *Proc. 3rd Int. Symp. 3D Data Process. Vis. Transmiss.*, 2006, vol. 6, pp. 326–333.
- [63] B. C. Munsell, P. Dalal, and S. Wang, "Evaluating shape correspondence for statistical shape analysis: A benchmark study," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 30, no. 11, pp. 2023–2039, Nov. 2008.
- [64] F. M. Sukno, J. L. Waddington, and P. F. Whelan, "3-D facial landmark localization with asymmetry patterns and shape regression from incomplete local features," *IEEE Trans. Cybern.*, vol. 45, no. 9, pp. 1717–1730, Sep. 2015.
- [65] S. Z. Gilani, F. Shafait, and A. Mian, "Biologically significant facial landmarks: How significant are they for gender classification?" in *Proc. Int. Conf. Digit. Image Comput.: Techn. Appl.*, 2013, pp. 1–8.
- [66] P. Szeptycki, M. Ardabilian, and L. Chen, "A coarse-to-fine curvature analysis-based rotation invariant 3D face landmarking," in *Proc. IEEE 3rd Int. Conf. Biometrics: Theory Appl. Syst.*, 2009, pp. 1–6.
- [67] A. Mian, M. Bennamoun, and R. Owens, "An efficient multimodal 2D-3D hybrid approach to automatic face recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 11, pp. 1927–1943, Nov. 2007.
- [68] A. Dehghan, E. Ortiz, R. Villegas, and M. Shah, "Who do I look like? Determining parent-offspring resemblance via gated autoencoders," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops*, 2014, pp. 171–171.
- [69] S. Z. Gilani and A. Mian, "Perceptual differences between men and women: A 3D facial morphometric perspective," in *Proc. 22nd Int. Conf. Pattern Recognit.*, 2014, pp. 2413–2418.

- [70] S. Z. Gilani, F. Shafait, and A. Mian, "Gradient based efficient feature selection," in *Proc. IEEE Winter Conf. Appl. Comput. Vis.*, 2014, pp. 191–197.
- [71] F. Al-Osaimi, M. Bennamoun, and A. Mian, "An expression deformation approach to non-rigid 3D face recognition," *Int. J. Comput. Vis.*, vol. 81, no. 3, pp. 302–316, 2009.
- [72] C. Queirolo, L. Silva, O. Bellon, and M. Segundo, "3D face recognition using simulated annealing and the surface interpenetration measure," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 2, pp. 206–219, Feb. 2010.
- [73] H. Drira, B. Ben Amor, A. Srivastava, M. Daoudi, and R. Slama, "3D face recognition under expressions, occlusions, and pose variations," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 9, pp. 2270–2283, Sep. 2013.
- [74] N. Alyuz, B. Gokberk, and L. Akarun, "A 3D face recognition system for expression and occlusion invariance," in *Proc. IEEE 2nd Int. Conf. Biometrics: Theory Appl. Syst.*, 2008, pp. 1–7.
- [75] A. Colombo, C. Cusano, and R. Schettini, "Three-dimensional occlusion detection and restoration of partially occluded faces," *J. Math. Imag. Vis.*, vol. 40, no. 1, pp. 105–119, 2011.
- [76] J. D. Erico, "Surface fitting using gridfit," *MATLAB Central File Exchange*, 2008, <https://au.mathworks.com/matlabcentral/fileexchange/8998-surface-fitting-using-gridfit>
- [77] J. Booth, A. Roussos, S. Zafeiriou and D. Dunaway, "A 3D morphable model learnt from 10,000 faces," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 5543–5552.
- [78] H. Li, R. W. Sumner, and M. Pauly, "Global correspondence optimization for non-rigid registration of depth scans," *Comput. Graph. Forum*, vol. 27, no. 5, pp. 1421–1430, 2008.
- [79] H. Li, B. Adams, L. J. Guibas, and M. Pauly, "Robust single-view geometry and motion reconstruction," *ACM Trans. Graph.*, vol. 28, no. 5, 2009, Art. no. 175.
- [80] T. Bolkart and S. Wuhrer, "A robust multilinear model learning framework for 3D faces," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 4911–4919.
- [81] V. G. Kim, Y. Lipman, and T. Funkhouser, "Blended intrinsic maps," *ACM Trans. Graph.*, vol. 30, no. 4, 2011, Art. no. 79.



**Syed Zulqarnain Gilani** received the MS degree in EE from the National University of Sciences and Technology (NUST), Pakistan, in 2009 and secured the President's Gold Medal. He received the PhD degree from the University of Western Australia. His research interests include 3D facial morphometrics with applications to syndrome delineation and machine learning.



**Ajmal Mian** is an Associate Professor of Computer Science at The University of Western Australia. He has received several awards including the West Australian Early Career Scientist of the Year Award, the Vice-chancellors Mid-career Research Award and the Outstanding Young Investigator Award. He has received two prestigious fellowships and seven major grants from the Australian Research Council and the National Health and Medical Research Council with total funding of \$3.0 Million. His research interests include computer vision, machine learning, 3D face analysis, human action recognition and remote sensing.



Germany in 2008. His research interests include machine learning and pattern recognition.

**Faisal Shafait** currently working as an Associate Professor at School of Electrical Engineering and Computer Science (SEECs), NUST, Pakistan. He is also an Adjunct Senior Lecturer at the School of Computer Science at The University of Western Australia. Formerly, he was a Senior Researcher at the German Research Center for Artificial Intelligence (DFKI), Germany and a visiting researcher at Google, California. He received his Ph.D. in computer engineering from Kaiserslautern University of Technology, Germany in 2008. His research interests include machine learning and pattern recognition.



**Ian Reid** received the DPhil degree from the University of Oxford, in 1991. He is a professor of computer science with the University of Adelaide. He has been employed in the Robotics Research Group, conducting research in computer vision, including holding an EPSRC Advanced Research Fellowship (1997-2000), and he has been a University lecturer since 2000. In 2005, he was awarded the title of reader and in 2010 the title of professor.

► For more information on this or any other computing topic, please visit our Digital Library at [www.computer.org/publications/dlib](http://www.computer.org/publications/dlib).