Pattern Recognition 69 (2017) 238–250

Contents lists available at ScienceDirect

Pattern Recognition

journal homepage: www.elsevier.com/locate/patcog

Deep, dense and accurate 3D face correspondence for generating population specific deformable models

Syed Zulqarnain Gilani[a,*], Ajmal Mian[a], Peter Eastwood[b]

[a] School of Computer Science and Software Engineering, University of Western Australia, Australia
[b] Centre for Sleep Science, School of Anatomy, Physiology and Human Biology, University of Western Australia, Australia

ARTICLE INFO

Article history:
Received 20 October 2016
Revised 13 March 2017
Accepted 12 April 2017
Available online 21 April 2017

Keywords:
Dense 3D face correspondence
3D face morphing
Keypoint detection
Shape descriptor
Face recognition
Landmark identification
Deep learning

ABSTRACT

We present a multilinear algorithm to automatically establish dense point-to-point correspondence over an arbitrarily large number of population specific 3D faces across identities, facial expressions and poses. The algorithm is initialized with a subset of anthropometric landmarks detected by our proposed Deep Landmark Identification Network which is trained on synthetic images. The landmarks are used to segment the 3D face into Voronoi regions by evolving geodesic level set curves. Exploiting the intrinsic features of these regions, we extract discriminative keypoints on the facial manifold to elastically match the regions across faces for establishing dense correspondence. Finally, we generate a Region based 3D Deformable Model which is fitted to unseen faces to transfer the correspondences. We evaluate our algorithm on the tasks of facial landmark detection and recognition using two benchmark datasets. Comparison with thirteen state-of-the-art techniques shows the efficacy of our algorithm.

1. Introduction

Dense 3D shape correspondence is the process of establishing a mapping between a large number points on one surface to topologically similar points on other surfaces. It has many applications in statistical shape analysis, computer graphics and shape phenotyping. In the context of 3D face analysis, dense correspondence has been used for landmark detection [1], face recognition [2–5], facial morphometric measurements such as asymmetry for syndrome diagnosis [6,7], facial changes after maxillofacial surgery [8], non-rigid shape registration [9–11], statistical shape modelling [12–14], shape interpolation [15] and deformation analysis [16]. Despite the growing applications, dense 3D face correspondence remains challenging because human faces have non-linear surface dissimilarities due to variations in age, ethnicity and expressions. Moreover, the unavailability of ground-truth dense correspondence prohibits the direct use of learning based algorithms for this task and also makes subsequent evaluation of the results difficult.

A typical pipeline of establishing correspondence between faces starts with finding keypoints on 3D facial surfaces and then matching the descriptors of these keypoints [17]. The established correspondence can either be sparse [18–20] or dense [1,5] based on the target application. Once correspondences have been established on N number of faces, a statistical model can be generated and fitted on a target query face to transfer the correspondence. Manual annotation of a sparse set of landmarks on the 3D faces can be an alternative to establishing the initial correspondence.

Many techniques [1,5,18–25] have been proposed in the literature for establishing correspondence between an arbitrary number of faces. However, these techniques have one or more of the following limitations: (1) They are landmark specific and hence the number of corresponding points are sparse. Such techniques can detect only a few landmarks (~14) on the 3D human face [18,19]. (2) They depend on the manual annotation of a few landmarks for initialization [25] or fail if a specific number of landmarks are not automatically detected [3]. (3) Fully automatic methods generally have large correspondence errors and are restricted to faces with neutral expression [1]. (4) Some methods use texture to aid in correspondence [5,26]. However, a potential pitfall of texture based dense correspondence is that the facial texture is not always consistent with the underlying 3D facial morphology e.g. the shape and location of eyebrows. (5) Most techniques are computationally expensive [1,27].

In this context, we propose a fully automatic multilinear algorithm that can establish dense correspondence ($\geq$ 13, 500 points) over a large number of 3D human faces with varying identities and expressions. We first train a deep Convolutional Neural Network (CNN) for landmark identification. CNNs have been

* Corresponding author.
E-mail address: zulqarnain.gilani@uwa.edu.au (S.Z. Gilani).

http://dx.doi.org/10.1016/j.patcog.2017.04.013
0031-3203/

extensively used for 2D texture images where the training data with ground truth labels is abundantly available. However, in case of 3D faces there is a dearth of training data that contains significant variation in facial shape, ethnicity and expressions. Our Deep Landmark Identification Network (DLIN) is trained on synthetic 3D images generated from a commercial software (*FaceGen*™) and is able to detect 11 biologically significant [28] facial landmarks with high accuracy and efficiency. Next, we divide the 3D face into five Voronoi regions around a subset of these landmarks using geodesically evolved level set curves [29]. A sparse set of discriminative keypoints are detected within each region and used to elastically align the corresponding region shapes of two given faces. Dense correspondence is then achieved through nearest neighbour matches between the region vertices of all training faces. Finally, a 3D deformable model is constructed from the densely corresponding faces and the correspondence information is transferred to unseen 3D faces by fitting the deformable model in an iterative optimization.

Our novel contributions are as follows. Firstly, we propose a Deep Landmark Identification Network (DLIN) architecture that is trained on synthetic 3D data to efficiently detect biologically significant facial landmarks. Secondly, we propose a region based algorithm to efficiently establish dense correspondence between 3D faces where the identities and facial expressions vary simultaneously. Since we use the mean region shape to propagate the dense correspondence to the entire dataset of say N faces, the correspondence error may become large because the curvatures and discriminative keypoints start diminishing on the mean regions. Moreover, a model based representation of 3D faces can yield better results more efficiently [26] not only for correspondence transfer but for other applications like face recognition. These considerations have motivated us to propose a Region based 3D Deformable Model (R3DM). Our algorithm is capable of transferring correspondence to an unseen query face under expressions and pose variations with high accuracy and efficiency. Unlike the sparse landmark detection methods, our model can detect a very large number of salient landmarks on 3D faces.

Since direct comparison of dense correspondence algorithms is not possible due to the un-availability of ground truth [30], researchers often resort to applications such as landmark localization or face recognition for comparison. More accurate correspondence is likely to give higher face recognition and landmark localization accuracies. For these reasons, we have performed extensive experiments on these two tasks using the FRGCv2 [31] and Bosphorus [32] 3D face datasets. Comparisons with twelve benchmark algorithms show that our proposed technique outperforms all others in terms of accuracy and computational efficiency.

2. Related work

2.1. 3D face correspondence

Sun and Abidi [23,24] detected keypoints for surface matching by projecting geodesic contours around a 3D facial point onto their tangential plane and called them Finger Prints of the 3D point. This method is highly sensitive to surface noise and sampling density [33] of the underlying geometry [34]. The approach, with minor modifications, was employed by Salazar et al. [20] to establish point correspondence on 3D faces in the BU3DFE [35] database.

For landmark detection, Creusot et al. [18] proposed an algorithm that uses a machine learning approach to detect 14 corresponding biologically significant [28] landmarks on 3D faces. An LDA classifier was trained on a set of 200 faces and a linear model of 14 landmarks. They exploited a number of local descriptors and used a binary classification approach to detect the landmarks. Their method works well for neutral expression faces of the FRGCv2 [31] and Bosphorus [32] databases. A method to detect landmarks under large pose variations was proposed by Perakis et al. [19]. These authors used a statistical facial landmark model for the frontal face and another two models for the profile faces. Keypoints were 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 the three landmark models. Eight corresponding points were detected in the FRGCv2 and UND Ear databases using this method. Later, the authors proposed an improved version [22] for fusing features from 2D and 3D data to detect these landmarks.

Blanz and Vetter [25] manually annotated seven facial landmarks on 100 male and female faces each to initialize their dense correspondence algorithm. They proposed a dense correspondence algorithm based on optical flow on the texture and the 3D cylindrical coordinates of the facial points given their initial spatial alignment. The dense correspondence was used to construct a linear deformable 3D face model which was later used for face recognition [5,26]. However, they tested their algorithm on only 300 out of the 4,007 faces in the FRGCv2 database [26].

Passalis et al. [21] 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 [28] on 3D face scans and then segmenting it into different annotated areas. Later, Kakadiaris et al. [36] 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. [3] further improved the AFM by incorporating facial symmetry to perform pose invariant face recognition. However, the algorithm depends on the detection of at least five facial landmarks on a side pose scan.

Recently, Gilani et al. [1] presented a shape based dense correspondence algorithm for landmark detection. Their 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 the seed points of given faces to those of a reference face. The accuracy of landmark localisation depends on the number of initial seed points and does not improve further since the algorithm already employs a coarse to fine search. Dense correspondences are established between neutral expression scans only.

The literature also contains a few application specific methods for generating sparse correspondence on 3D faces using keypoints [27,37–41]. These keypoints are repeatable on the same identity and hence aid in face recognition and other face analysis tasks. For instance, Li et al. [27] proposed two principal curvature-based 3D keypoint detectors. Pose-invariant features are extracted using a 3D local coordinate system. Three keypoint descriptors are designed and their features are fused to perform face recognition on the Bosphorus dataset [32].The literature also points to the use of biologically inspired features for textured 2D face identification. Song et al. [42] proposed a general framework for extracting biologically inspired features, whereas [43] utilize these features to perform 2D face recognition.

2.2. Deep learning

After becoming the tool of choice for image classification [44] and object detection [45], CNNs have recently been used for semantic segmentation [46,47] and boundary prediction [48] in RGB images. Long et al. [49] analysed the specific local correspondence by investigating the ability of intermediate features learnt in

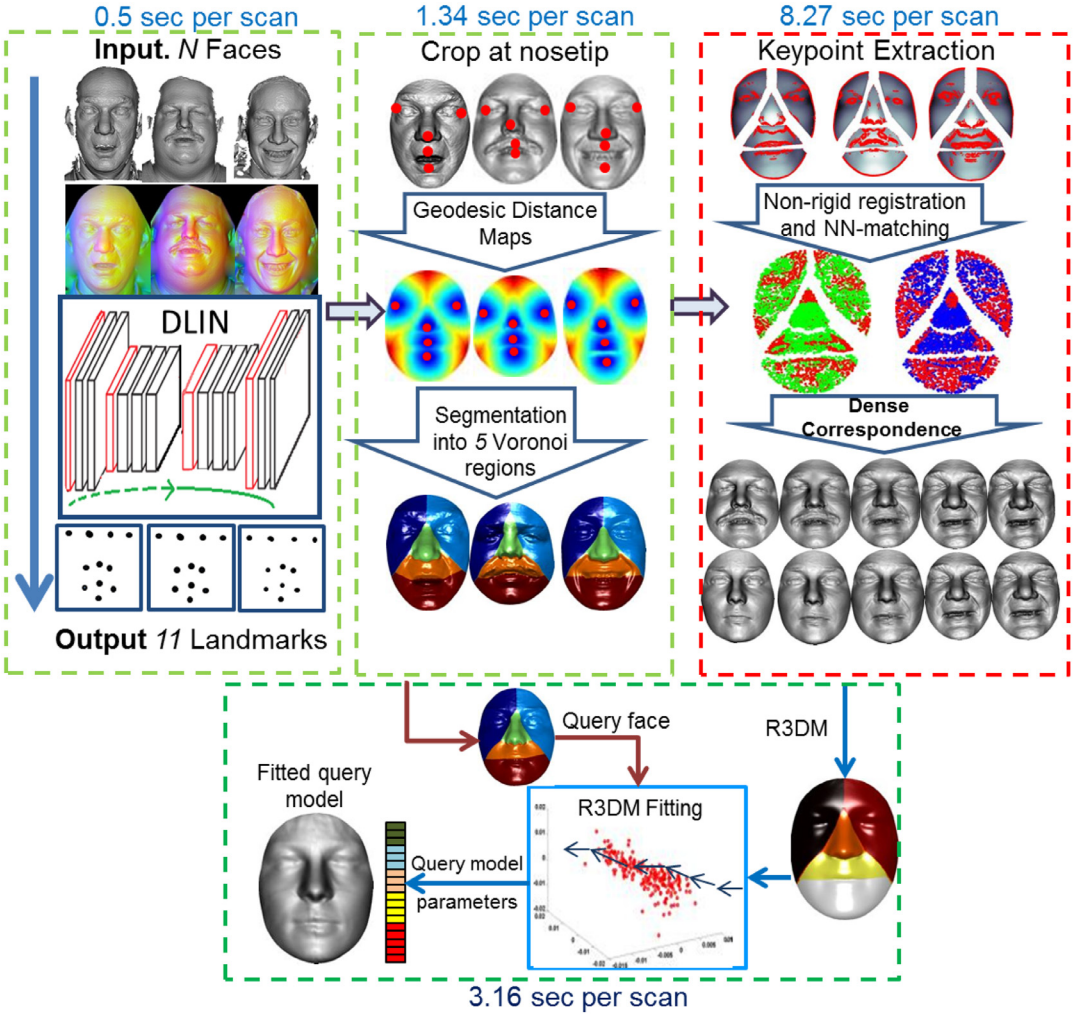


Fig. 1. Block diagram of the proposed dense 3D face correspondence algorithm. The red box contains steps that are performed offline. Note that both online (green boxes) and offline processes are fully automatic and the online process is very efficient given the complexity of the dense correspondence task. The timings are without using a GPU. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

CNNs. The authors compared the deep features with SIFT features and found that the former are more useful for extracting local visual information. Later the authors [50] adapted and extended a deep classification architecture to learn from whole image inputs and whole image ground truths for semantic segmentation. They trained a Fully Convolutional Network (FCN) end-to-end, pixel-to-pixel and showed that their results on semantic segmentation outperformed the state-of-the-art. To the best of our knowledge, CNNs have not been used for landmark detection in 3D faces using a point-to-point or pixel-to-pixel identification architecture.

3. Proposed algorithm

Fig. 1 shows the block diagram of the proposed dense correspondence algorithm and the following sections give details of each component.

3.1. The Deep Landmark Identification Network

3.1.1. Generating synthetic training data

To train the Deep Landmark Identification Network (DLIN), we synthetically generate realistic 3D faces using a commercial software (*FaceGen*^{TM}) by varying the facial shape and facial expressions. Since the faces are generated from a model, the ground truth locations of landmarks are already known. The training data cover a huge space in terms of facial shape variations due to age, ethnicity and expressions. We additionally induce pose variations in the training data to cater for pose variations in test images. Each 3D training face is rendered from five viewing angles, that is, frontal, $\pm$ 15° in pith and $\pm$ 15° in roll. Next, their depth images are generated by fitting a surface of the form $z(x, y)$ to each 3D pointcloud using the *gridfit* algorithm [51]. We also calculate the Cartesian surface normals (n_x, n_y, n_z) of each vertex in the pointcloud and convert them to spherical coordinates (n_θ, n_ϕ, n_r) where θ is the azimuth, ϕ is the elevation and r is the radius of the normal. A surface similar to the one used for depth images is fitted on the former two components of the normal. The depth, azimuth and elevation images are used as the three channels, instead of the usual RGB channels, as input images to the DLIN. The process is depicted in Fig. 2. Since the correspondence in the synthetic images is known *a priori*, we are able to automatically label 11 landmarks on each face as shown in Fig. 3. These are biologically significant landmarks [28] which define the high curvature points of the face and encode variation in expression. Ground truth landmark locations are similarly projected on a 2D surface, dilated with a *disk* shaped structure of size 10 $\times$ 10 and converted into a binary 2D image. The process is shown in Fig. 2. Note that each input image

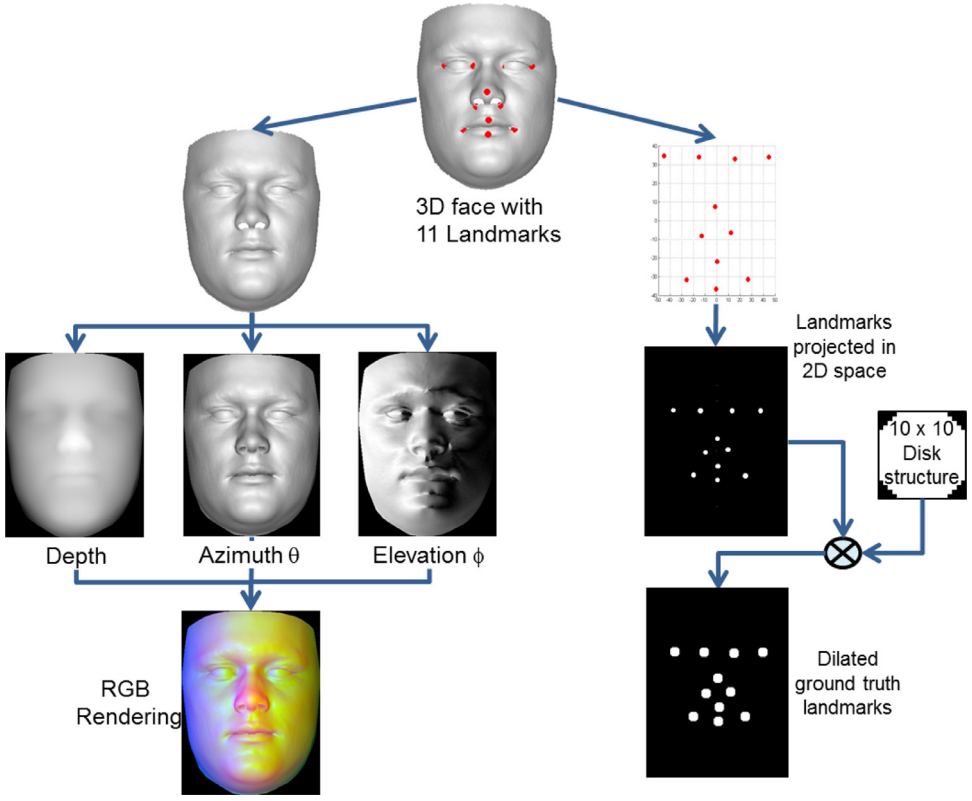
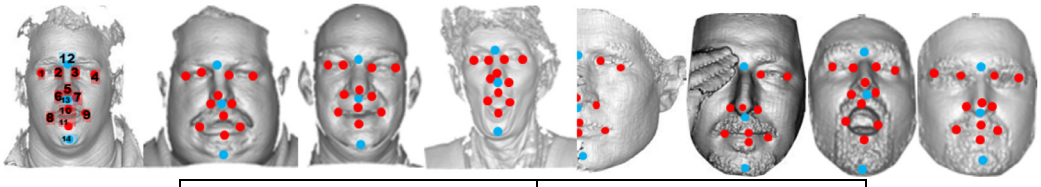


Fig. 2. The process of generating training data for the DLIN. Left shows the process of preparing the input image while preparation of ground truth landmarks is shown on the right. The process is repeated for all input images to the DLIN.



1. Outer Eye Corner (Right)-OEC(R)	8. Mouth Corner (Right) - MC(R)
2. Inner Eye Corner (Right)-IEC(R)	9. Mouth Corner (Left) - MC(L)
3. Inner Eye Corner (Left) IEC(L)	10. Upper Lip Centre (ULC)
4. Outer Eye Corner (Left)-OEC(L)	11. Lower Lip Centre (LLC)
5. Nose Tip (NT)	12. Nasal Root
6. Nose Corner (Right) NC(R)	13. Nasal Base
7. Nose Corner (Left) NC(L)	14. Chin Tip

Fig. 3. Sample images from the FRGCv2 (first four) and Bosphorus databases (last four) along with the landmarks used in this paper. We detect the 11 red coloured landmarks with DLIN and the three blue ones by fitting the R3DM. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

may have a different dimension (width and height) but the image and its corresponding binary image are of the same size.

3.1.2. Training the Deep Landmark Identification Network (DLIN)

Our proposed architecture is motivated by the Fully Convolutional Network (FCN) of Long et al. [50] with changes to better suit 3D depth data instead of RGB. The FCN is based on the VGG architecture [50]. The FCN and VGG networks are both designed for 2D images/textures. While texture can change abruptly in images, 3D surfaces are generally smooth and this is especially the case for 3D facial surfaces. Hence a smaller network with fewer parameters is sufficient to learn the variations. Learning the parameters of a smaller network is fast and needs less training data. A smaller network executes faster when deployed. Finally, the FCN was designed for 21 classes, while we intend to learn DLIN for two classes only. For these reasons, we reduce the number of convolutional layers from 13 to 5 and change the final upsampling to 4x. Since, these changes are significant, the parameters of the proposed DLIN are learned from scratch.

The input to the DLIN comprises of the depth, azimuth and elevation channels. Notice from Fig. 2 that the landmarks we selected are very conspicuous in these channels which backs up our argument that fewer convolution layers are sufficient for their delineation. Experiments show that reducing eight convolution layers

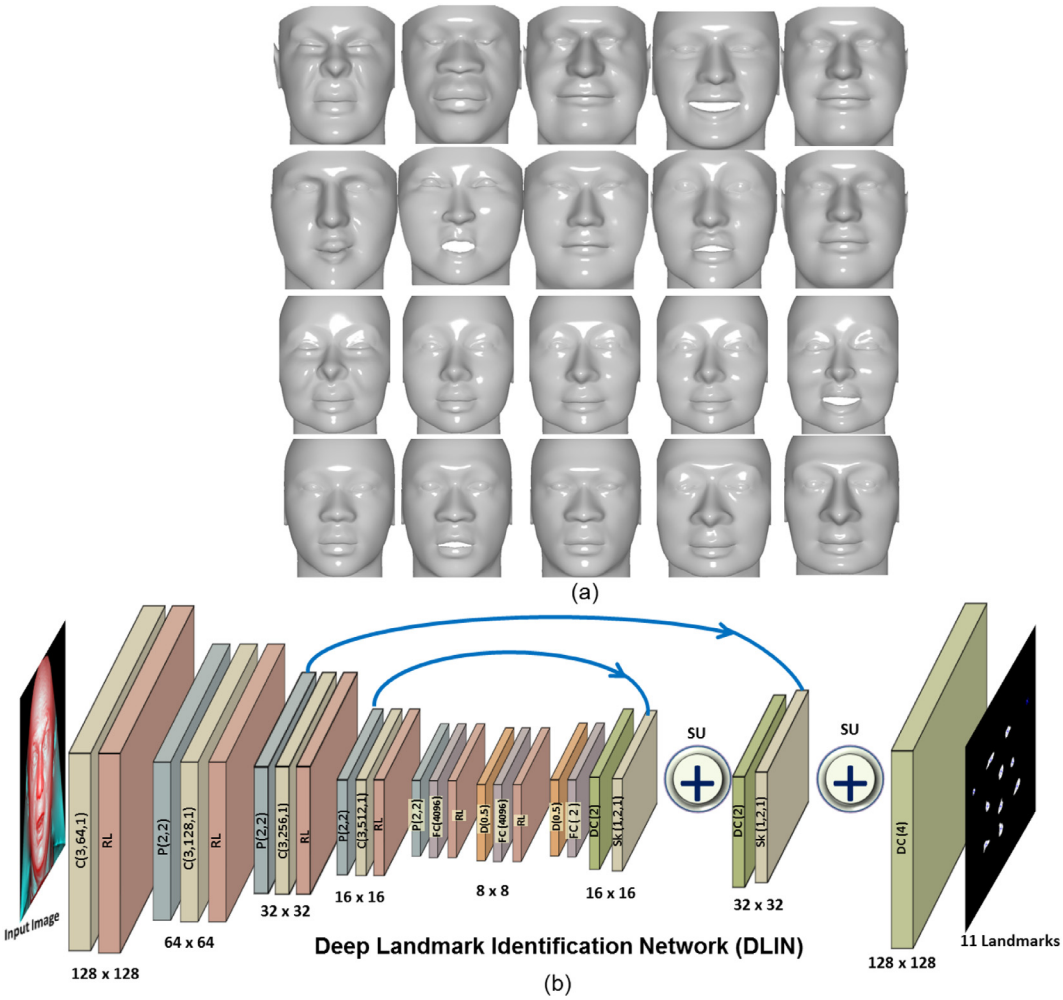


Fig. 4. (a) Sample images used to train the DLIN. Notice the variations in different modalities of the training data. (b) Network architecture of DLIN.

reduces the learning time by approximately three folds and the landmark detection time by four folds. Moreover, DLIN slightly improves the accuracy over the fine-tuned FCN. Further details are given in Section 5.1 .

We also replace the 21 class prediction layer with two classes for detecting the landmarks. At this stage we are only detecting the landmarks and not labelling them individually. Exact classification into the 11 landmarks is done outside the network based on their relative positions. We adopt this binary classification strategy rather thank training the network to classify and label the landmarks into one of the 11 classes because except for the nope tip, the remaining landmarks are symmetric on either side of the face and somewhat ambiguous. Assigning a different class label to each landmark causes errors in detection. This was confirmed experimentally. Moreover, once all the landmarks are detected with a single label, it is trivial to assign them to the 11 landmark classes based on facial anthropometry [28,52]. Therefore, we use two classes in this paper. The names and locations of the 11 landmarks is shown in Fig. 3. Let $C(k, n, s)$ denote a convolutional layer with kernel size $k \times k$, n filters and stride s, RL denote a rectified linear unit, $P(k, s)$ denote a max pooling layer with kernel size $k \times k$ and stride s, $FC(n)$ denote a fully connected layer with n filters and $D(r)$ denote a dropout layer with drop our ratio r. Let $Sk(k, n, s)$ denote the skip layer where the parameters have a similar meaning as in the convolutional layer, SU denote the sum layer and $DC(u)$ denote the deconvolution layer where u is the upsampling ratio. Figure. 4 shows the architecture of our proposed DLIN which is enumerated as follows:

$C(3, 64, 1) \rightarrow RL \rightarrow P(2, 2) \rightarrow C(3, 128, 1) \rightarrow RL \rightarrow P(2, 2) \rightarrow C(3, 256, 1) \rightarrow RL \rightarrow P(2, 2) \rightarrow C(3, 512, 1) \rightarrow RL \rightarrow P(2, 2) \rightarrow FC(4096) \rightarrow RL \rightarrow D(0.5) \rightarrow FC(4096) \rightarrow RL \rightarrow D(0.5) \rightarrow FC(2) \rightarrow DC(2) \rightarrow Sk(1, 2, 1) \rightarrow SU \rightarrow DC(2) \rightarrow Sk(1, 2, 1) \rightarrow SU \rightarrow DC(4)$

Our training data consists of 30,000 synthetic 3D faces which are composed of 125 images from 240 male and female identities. Each identity has 25 images in varying shapes where the variation is with respect to age, masculinity/femininity, weight, height and facial expressions of surprise, happiness, fear and disgust. Each of the 25 images is then rendered in 5 different poses. It has been shown [1,53,54] that these expressions contribute to significant shape variation of the lower face. We use the facial images in depth, azimuth and elevation format of 200 identities for training and those of the remaining 40 identities for validation for learning the Deep Landmark Identification Network. The new network is trained from scratch by zero initializing the model parameters. We use a momentum of 0.9, a weight decay of 0.0005 which are same as in the FCN [50]. We train the network for 200 epochs using Matconvenet [55]. Some training data samples are depicted in Fig. 4.

3.1.3. Landmark identification using DLIN

Let $\mathbf{F_j} = [x_i, y_i, z_i]^T$ $(j = 1, \ldots, N$ and $i = 1, \ldots, P_j)$ be a real arbitrary 3D face scan. We obtain the depth, azimuth and elevation image of this face by fitting a surface to $\mathbf{F_j}$ as described in

Fig. 5. Results of landmark detection using the proposed DLIN on sample scans from the FRGCv2 (left four columns) and the Bosphorus (right four columns) datasets.

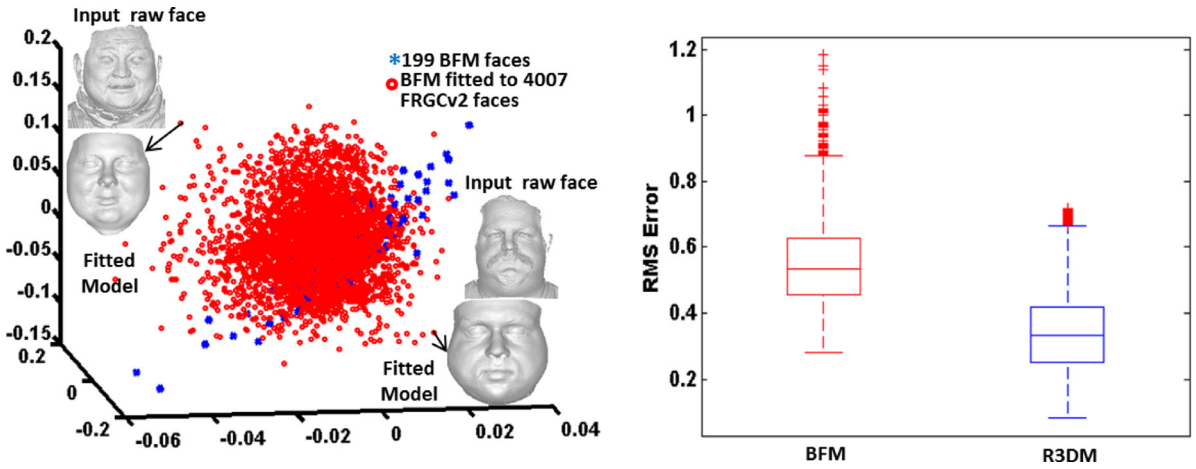


Fig. 6. (Right) BFM [56] and FRGCv2 [31] faces depicted in the space of their first three PCs. Notice that BFM is not able to generalize over some real faces. (Left) the minimum RMS error of each face of FRGCv2 to the BFM and our proposed R3DM in the PCA space.

Section. 3.1.1 and pass it through the learned DLIN. The output is a binary mask of landmark locations. Subsequently we apply some basic morphological operations to this binary image and utilize the veridical Cartesian coordinates of the DLIN input scan to convert the landmark mask to 3D point coordinates and denote them by $\mathcal{L}_j = [x_k, y_k, z_k]^T$, where $k = 1, \ldots, 11$. This enables us to report the localisation error in metric units (mm) and facilitates comparison with the state-of-the-art. Using the nose tip as the centre we crop a sphere of 90 mm radius to discard non-facial regions. Fig. 3 defines the location of the 11 landmarks. Fig. 5 shows the detected landmarks on sample faces from the FRGCv2 and Bosphorus datasets.

3.2. Region based dense correspondence

3.2.1. Motivation

Our proposed algorithm uses a subset of landmarks detected by the DLIN as seed points for subsequent region based dense correspondence. Before going into the details of each algorithmic component we present the motivation for using a deep CNN for seed point detection. The literature contains some algorithms (See [57,58]) that use pre-computed face models for real time facial performance capture or facial animation. These pre-computed face models can also be used to detect a sparse set of landmarks to initialize our proposed R3DM. However, this strategy suffers from two main drawbacks. Firstly, The pre-computed face models span a limited face space and cannot generalize to extreme cases of real life data. We fitted the publicly available Basel Face Model (BFM) [56], a variant of Blanz and Vetter's [25] morphable model to the 4,007 faces of the FRGCv2 [31] dataset. Some of the faces in the dataset do not fall in the space covered by the BFM faces. We demonstrate this graphically in Fig. 6 (Left), by depicting the BFM and FRGCv2 faces in the space of their first three Principal Components(PCs). Notice that BFM is unable to generalize to some faces from FRGCv2 and would thus fail in detecting accurate landmarks. We also show in Fig. 6 (Right) the minimum RMS error of each face to the model in the PCA space and compare BFM with our proposed R3DM. It is evident that the data driven model (R3DM) is able to generalize better than a pre-computed model. The empirical validation is provided in our results (Section 5.1). Accuracy of landmark detection and correspondence is of paramount importance in medical applications where the dense correspondence must be very precise and often population specific to detect subtle shape changes for syndrome delineation (e.g. Autism [59–61], Schizophrenia [7]). Therefore, initializing the R3DM algorithm with points detected by a pre-computed model will induce further error in establishing dense correspondence. Secondly, the pre-computed face models require seed points for initial registration. On the other hand, we propose an algorithm for "making" these models from real scans. There is a need for an automatic algorithm that can establish dense correspondence between new faces without using pre-computed models to avoid bias.

3.2.2. Dense correspondence

The input to our dense correspondence algorithm is a set of N 3D faces denoted by $\mathbf{F_j}$ and their corresponding fiducial landmarks $\mathcal{L}_j$. We select a subset of five biologically significant landmarks which include the outer eye corners, the nosetip and the upper and lower lip centres. The selection of the last two landmarks is intuitive since the expressions that mainly involve discontinuity in the mouth region (e.g. open smile, surprise) separate the two lips. Exploiting the upper and lower lip centres aids in better correspondences across facial expressions.

Starting from each of the five landmarks, we evolve level set curves to find the geodesic distance between the landmarks and all the vertices of the face. At any given point i on the 3D face $\mathbf{F_j}$,

the level set interface [62] is given by $\Phi(i) = 0$ and the level set equation by,

$$\Phi_t + \frac{\nabla|\Phi|}{\mathcal{P}} = 0 \tag{1}$$

where Φ_t is the position of the level set interface at time t [62], $\mathcal{P}$ is some metric over the face manifold and $\mathcal{F} = 1/\mathcal{P}$ is the speed function of the level set interface. The level set curve propagates following the evolution equation $\frac{1}{\mathcal{P}_i}\vec{n_i}$, where $\vec{n_i}$ is the exterior unit vector normal to the curve at point i. The distance function between any two points, $\Gamma_i \stackrel{def}{=} d(i, j)$ (where $d(i, j)$ is the surface distance between points i and j [29]) satisfies the Eikonal equation,

$$\|\nabla\Gamma_i\| = \frac{1}{\mathcal{P}_i} \tag{2}$$

When $\mathcal{P}_i = 1$, $\|\nabla\Gamma_i\|$ denotes the shortest surface distance between the two points (i, j) and is called the geodesic distance. We solve the LHS of (2) by making use of the Fast Marching [29] algorithm on an orthogonal grid adopting an upwind finite difference scheme. Using an efficient implementation [63] of the solution, we find the geodesic distance map $\mathbf{D}(k, i)$ where $k = 1, \ldots, 5$ and $i = 1, \ldots, P_j$. $\mathbf{D}(k, i)$ is the geodesic distance given in Eq. (2) between the five landmarks and each vertex on the 3D face. Next, the 3D face is segmented into five Voronoi regions pertaining to the five landmarks. By definition, a vertex i on a 3D manifold belongs to the Voronoi region $\mathcal{V}_k$ if the geodesic distance $\mathbf{D}(k, i) < \mathbf{D}(l, i)$, where $k \neq l$. This process is repeated for all N 3D faces resulting in Voronoi regions $\mathcal{V}_k^j = [x_i, y_i, z_i]^T$, where $i = 1, \ldots, p$ and $k = 1, \ldots, 5$.

Next, we proceed by detecting keypoints inside the Voronoi regions $\mathcal{V}_k^j$ to find discriminative points for aligning these regions across the N input faces. We follow the procedure outlined by Mian et al. [37] and crop a small surface of radius ρ around each point p in $\mathcal{V}_k^j$. However, rather than using, a large radius in order to capture identity specific features [64], we use a small value of $\rho = 5mm$, to encode surface specific discriminative features across identities. We perform Principle Component Analysis (PCA) of this surface and find the ratio of the first two principal components. A point p is accepted as a keypoint if the ratio is more than 1.5, a threshold that was empirically determined.

To establish dense correspondence across N faces, a random reference face is first selected. We align the keypoints of Voronoi Region $\mathcal{V}_k^2$ of an arbitrary face $\mathbf{F_2}$ in the collection of N faces with the keypoints of the reference face Voronoi region $\mathcal{V}_k^1$. The alignment is non-rigid [65,66] and matches the shapes of the regions from the two faces. Furthermore, this alignment process is fast and efficient as it is performed on only a sparse set of discriminative keypoints. Next the registration information is used to align the two Voronoi regions and dense correspondence is established between them through nearest neighbour match using the k-d tree data structure [67]. Once the correspondences are established between the two Voronoi regions, their mean shape is used as the reference and the region $\mathcal{V}_k^3$ from a third face is aligned with this shape following the same procedure outlined above. This process is repeated for all k Voronoi regions across N faces to establish dense correspondence. Note that segmenting the face into regions and establishing region based dense correspondence introduces inaccurate correspondences at the boundaries of the Voronoi regions. To mitigate this problem, we perform a single iteration of global registration and matching between the reference and the target face and update the correspondences at the boundaries only. The output of this step is a set of densely corresponding facial regions $\mathbb{V}_k^j$ which together form corresponding 3D faces $\mathbb{F}_j^c = [x_i, y_i, z_i]^T$, where $j = 1, \ldots, N$ and $i = 1, \ldots, P$.

3.2.3. Deformable model fitting

Our dense correspondence algorithm outlined above is efficient and capable of establishing vertex level mapping between N faces. The correspondences can then be used in a variety of applications, such as detecting a large number of fiducial landmarks [1]. However, some applications like face recognition can be performed more efficiently by creating a 3D deformable model from the set of N example faces and then matching the model coefficients [5,26]. For this purpose we propose a region based 3D deformable model (R3DM). Propagating dense correspondence to a large number of query faces through deformable model fitting is faster than establishing region based correspondence. However, the accuracy of the propagated correspondences depends on the quality of the model which depends directly on the quality of the initial dense correspondence. Thus, the R3DM also serves as a way of validating the quality of our dense correspondence algorithm.

Let the R3DM of region k be denoted by $\mathbf{\Theta}_k = [\mathbf{v}_k^1, \mathbf{v}_k^2, \ldots, \mathbf{v}_k^N,]$, where $\mathbf{v}_k^j$ is the vectorised form of densely corresponding facial regions $\mathbb{V}_k^j$, i.e.,

$\mathbf{v}_k^j = [x_1, \ldots, x_i, y_1, \ldots, y_i, z_1, \ldots, z_i]^T$ and $i = 1, \ldots, p$. The row mean $\boldsymbol{\mu}_k = \frac{1}{N}\sum_{n=1}^{N}\mathbf{v}_k^n$ of the R3DM is subtracted from the $\mathbf{\Theta_k}$ to obtain a zero mean R3DM $\overline{\mathbf{\Theta}}_k$. Next, we model the R3DM by a multivariate Gaussian distribution and obtain it's eigenvalue decomposition,

$$\mathbf{USV}^T = \overline{\mathbf{\Theta}}_k \tag{3}$$

where $\mathbf{US}$ are the principal components (PCs), and the columns of V are their corresponding loadings. The mean of each facial region is given by $\overline{\mathbb{V}}_k^j$.

A query face $\mathbf{Q}$ of unknown and unseen identity goes through the process of landmark identification through DLIN and segmentation into five facial regions as described in the previous sections. Centring both the R3DM and the query face at their nosetips achieves initial face alignment. Next, the statistical model of each region given in (3) is deformed to synthesize a random subset of 3D points in its respective query face region in a two step optimization process. The vectorised query model is obtained by $\mathbf{m}_k^q = \mathbf{U}\boldsymbol{\alpha}_k + \boldsymbol{\mu}_k$, where $\boldsymbol{\alpha}_k$ denotes the parameters required to vary the shape of the model $\mathbf{m}_k^q$. A rigid transformation of the form $\mathbf{R}\mathcal{V}_k^q + \mathbf{t}$ aligns the query face region to its model region. Here, $\mathbf{R}$ is the rotation matrix and the required translation is denoted by $\mathbf{t}$. The two step process is repeated iteratively. Empirically, we found that the fitting error converges in $n = 20$ iterations to obtain an optimal representation for region k of the query model. Finally, the parameters of all regions are concatenated to be used as features representing the face.

4. Experimental setup

4.1. Datasets

The synthetic dataset is used only for training the DLIN, whereas rest of the experiments are conducted on real 3D face datasets, including the benchmark FRGCv2 [31] and Bosphorus [32] datasets. Both datasets are rich in facial expression variations whereas the latter additionally includes pose variations and occlusions. FRGCv2 comprises of 4,007 scans from 466 identities of different ethnicities and age groups. The scans are mostly frontal with minor ($\pm$ 10°) pose variations. The facial expressions range from neutral to extreme but are not labelled as such. Manual landmark annotations provided by Szeptycki et al. [68] and Creusot et al. [18] are used as ground truth for comparison. The Bosphorus dataset contains 4,666 3D faces from 105 subjects with considerable variation in ethnicity and age. The dataset is structured into

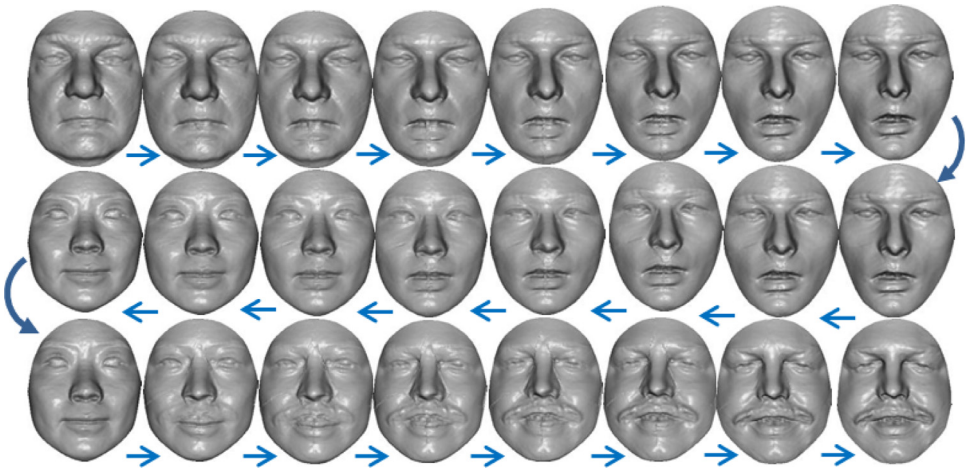


Fig. 7. Subjective evaluation of our dense correspondence algorithm by morphing four identities of FRGCv2. Notice the seamless morphs across expressions.

Table 1
Comparison of detection timing (seconds) and mean landmark localization error (mm) on 4,007 scans of the FRGCv2 dataset. The landmarks are defined in Fig. 3.

	Time /	Landmark localisation error (mm)								
Method	Image	EC	NR	NT	NC	MC	ULC	CT	NB	Mean
FCN [50]	2 s	2.9	3.1	2.4	3.5	3.0	3.4	3.4	3.3	3.1
DLIN	0.5 s	2.8	3.1	2.4	3.4	2.7	3.2	3.4	3.2	3.0

Action Units (AU) including the generic expressions of happy, sad, surprise, fear, disgust, anger and neutral. Poses vary within a range of $\pm$ 90° in both yaw and pitch along with some cross pose variations in yaw and pitch simultaneously. The dataset also contains scans with four different types of occlusions. Ground truth landmark locations are provided with the dataset. Sample images from the two datasets along with details of the landmarks used in this paper are shown in Fig. 3.

4.2. Evaluation criteria

There is no known direct objective evaluation criterion for dense shape correspondence due to the unavailability of the ground truth [30]. Subjective evaluations can be carried out by visually inspecting the morph between densely corresponding faces. A seamless and smooth morph depicts higher quality of shape correspondence. Fig. 7 allows subjective evaluation of our proposed algorithm. The seamless morphs across identities and expressions are possible only when the underlying dense correspondence is of high quality [69,70]. For indirect objective evaluation, applications whose results are correlated with the dense correspondence accuracy are used. We have chosen two applications that are widely used in the literature for our experiments; facial landmark identification and face recognition. We compare our results with twelve state-of-the-art algorithms.

5. Results and analysis

5.1. Facial landmark detection

5.1.1. Comparison between DLIN and FCN

To evaluate the improvement of DLIN over the FCN [50], we follow the same training protocol as detailed in Section 3.1.2. More specifically, the DLIN is learned from scratch while FCN is fine-tuned on the same training data. We test both the networks on 4,007 scans of FRGCv2 and compare the landmark localisation error as well the detection time per image. Results in Table 1 show that the landmark detection speed of DLIN is four times faster than FCN whereas the accuracy is either equal to or slightly better than FCN for all landmarks.

5.1.2. FRGCv2 Data

The proposed DLIN provides 11 biologically significant [28] landmarks which can be compared with algorithms that are designed to detect only these landmarks. Fig. 8 shows the results of our proposed DLIN on the 4,007 scans of FRGCv2 and compares them with the state-of-the-art. It is clear that DLIN detects the landmarks with high accuracy and outperforms it competitors. In some cases the reduction in error is more than 50%. Overall the improvement in landmarking accuracy is more than 35%.

To obtain a large number of landmarks efficiently and to have a fair comparison with model based algorithms [1] we use our deformable model fitting algorithm. In this case, we annotate 14 landmarks on the mean face of R3DM (FGRCv2 and Bosphorus) and back project them on the individual faces. Landmark localisation error is calculated as the 3D Euclidean distance between the ground truth and the detected landmarks.

To ensure unbiased results (landmark detection on unseen faces), we keep the training and test identities mutually exclusive. We construct two region based dense correspondence models of 300 scans each from the FRGCv2 dataset. The first R3DM includes the first available neutral scans of identities 1–200 combined with the first available extreme expression scans of identities 1–100. This model is then fitted on the faces of identities 201–466 (total 1,961 scans) to transfer correspondences. The second R3DM is constructed from the first available neutral scans of identities 201–400 and the first available extreme expression scans of identities 201–300. This model is then used to transfer correspondences and detect landmarks in identities 1–200 (total 2,046 scans). Thus we report landmark detection results on unseen faces in all 4,007 scans of FRGCv2.

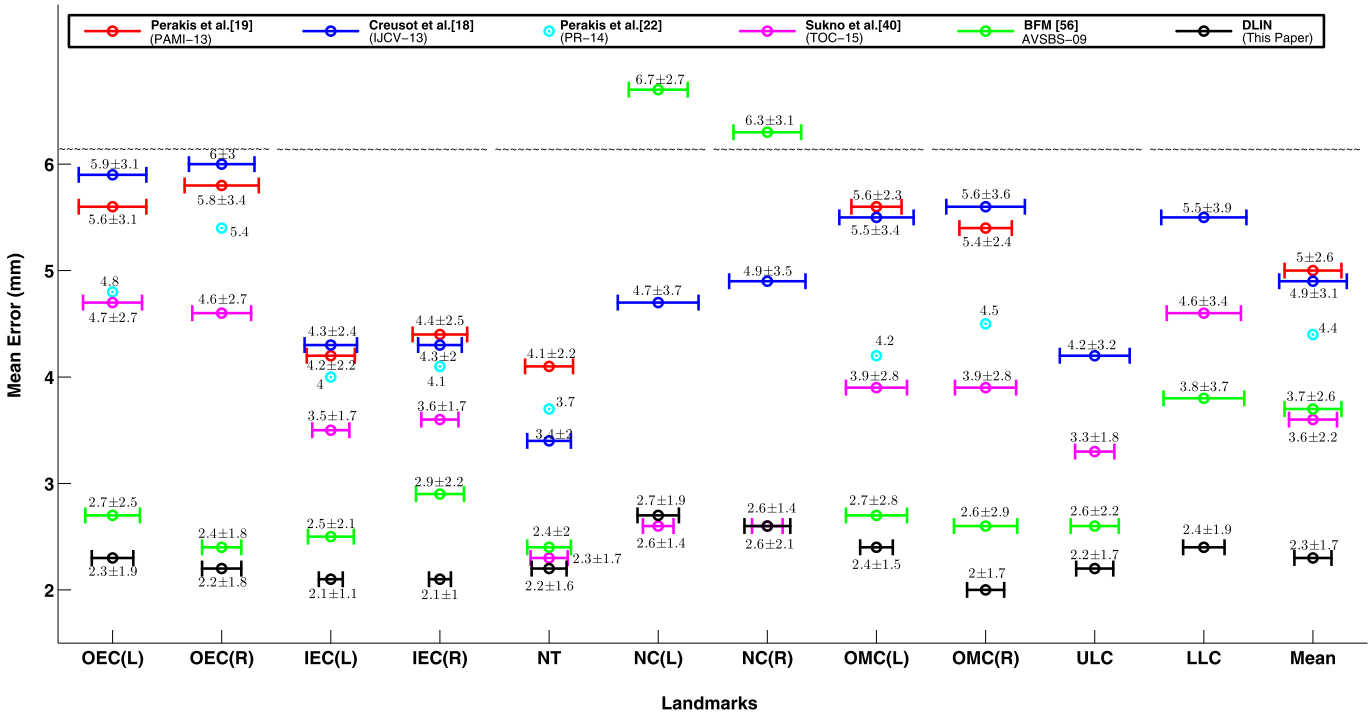


Fig. 8. Comparison of mean± SD (mm) of landmark localisation error on 11 landmarks in FRGCv2. It is clear that the DLIN outperforms the state-of-the-art. Note that [19,22] have not reported results for all landmarks shown in the figure, while [22] has not reported the SD of error. It takes only 0.5 s to detect the 11 landmarks through DLIN. The landmarks are defined in Fig. 3.

Table 2
Comparison of mean landmark localization error on 4,007 scans of FRGCv2 with Gilani et al. [1] (CVPR-15). The results are based on fitting a deformable model on the test dataset. Results of landmarks that occur in pairs have been averaged. The landmarks are defined in Fig. 3.

Author	OEC	IEC	NR	NT	NC	MC(L)	MC(R)	ULC	LLC	CT	NB	Mean
CVPR-15 [1]	4.1	2.9	3.6	2.7	4.3	5.3	4.4	3.3	4	4.2	4.1	3.9 ± 2.8
R3DM	2.9	2.7	3.1	2.4	3.4	2.8	2.6	2.9	3.6	3.4	3.2	2.9 ± 2.3

Table 2 compares the landmark localisation errors of our proposed R3DM with the latest algorithm proposed by Gilani et al. [1]. Once again the improvement in accuracy is more than 25%.

5.1.3. Bosphorus data

We use a similar strategy on the Bosphorus dataset to perform unbiased landmark detection on unseen faces. The first R3DM is constructed from identities 1–55 (containing 95 neutral and 55 happy expression scans) and fitted to the 2,095 test scans of the remaining identities. The second R3DM is constructed from identities 56–105 (containing 100 neutral and 50 happy expression scans) and fitted to the remaining 2,571 scans of the first 55 identities. Note that in both cases, the R3DM is constructed using only 150 near frontal pose scans and contains only the happy expression. However, our results show that our algorithm is still robust to facial expressions, pose variations and occlusions. (Table 3)

Table 2 reports our results separately for scans containing facial expressions, rotations and occlusions. The 11 landmarks detected by the DLIN are accurate within 2.95 mm and a spread of 1.83 mm, thereby improving the accuracy over the state-of-the-art by 30%. These results demonstrate the high accuracy of our proposed method as well as robustness to large pose and expression variations.

5.2. Face recognition

3D face recognition is a convenient application to test the quality of dense correspondence and R3DM fitting. Significant improvements can be achieved if the underlying dense correspondence is accurate. We define a challenging protocol to evaluate the generalization ability of our algorithm. For this purpose we build dense correspondence models that contain cross domain data. The first model ($R3DM_1$) is constructed from the first available neutral scan of the 466 identities of FRGCv2 and 100 expression scans of Bosphorus dataset. The second model ($R3DM_2$) is constructed from the first available neutral scans of the 105 identities of Bosphorus dataset and first 100 extreme expression scans of FRGCv2. Notice that the linear span of the example faces in both R3DMs contains neutral as well as non-neutral scans which on one hand helps in cross domain data analysis and on the other hand results in an enhancement of the face space and better generalisation to expressions.

We fit $R3DM_1$ to all 4,007 scans of FRGCv2 and use the model parameters of the first available neutral scans of the 466 identities as gallery (training data). We report Rank-1 recognition rate on the remaining 3,541 scans by minimising the cosine distance between the gallery and the probe model parameters. Fig. 9 compares the Rank-1 recognition rates of our proposed algorithm with the state-of-the-art on the FRGCv2 dataset. R3DM outperforms all others, especially in the neutral vs. non-neutral case.

Similarly, $R3DM_2$ is fitted to the 4,666 scans of the Bosphorus dataset. The model parameters of the first neutral scans of all 105 identities are used to form the gallery and Rank-1 recognition rates on the remaining 4,505 scans for the three types of variations in the dataset are reported in Table 4. R3DM performs better than

Table 3
Comparison of landmark localization results with the state-of-the-art on Bosphorus dataset. Results of landmarks that occur in pairs have been averaged and the last three were obtained through R3DM.Refer to Fig. 3 for landmark definitions.

Mean of localization error(mm)												
	Author	# Images	OEC	IEC	NT	NC	MC	LC	CT	NB	NR	Mean
Expression	Cruesot [18]	2803	5.15	4.64	4.47	4.15	6.03	6.51	8.83	15.23	6.33	6.27
	Sukno [40]	2803	5.06	2.85	2.33	3.02	6.08	5.27	7.58	2.81	2.22	4.25
	DLIN	2920	2.98	2.68	2.24	2.68	2.76	3.47	6.12	2.50	2.63	**2.80**
Rotation	Creusot [18]	1155	4.77	4.42	4.89	3.48	4.17	3.83	4.68	9.47	5.17	4.68
	Sukno [40]	1155	4.72	3.10	4.36	3.37	3.76	4.24	7.77	4.19	3.40	4.15
	DLIN	1365	3.39	2.84	2.63	3.11	3.01	3.79	6.47	3.93	2.81	**3.13**
Occlusion	Creusot [18]	381	6.79	5.30	4.72	5.10	4.86	4.56	5.44	11.05	7.78	5.87
	Sukno [40]	381	6.46	3.85	3.83	4.54	4.91	4.21	7.63	3.76	4.12	4.80
	DLIN	381	3.60	2.93	2.82	3.53	3.38	4.07	5.25	3.03	2.82	**3.39**
All	Creusot [18]	4339	5.14	4.61	4.60	4.05	5.44	5.59	7.35	13.20	6.10	5.78
	Sukno [40]	4339	5.09	3.01	3.00	3.25	5.36	4.90	7.63	3.26	2.70	4.27
	DLIN	4666	3.15	2.75	2.40	2.87	2.88	3.61	6.15	2.97	2.70	**2.95**
	Standard deviation of localization error (mm)											
	Author	# Images	OEC	IEC	NR	NT	NC	MC	LC	CT	NB	Mean
All	Creusot [18]	4339	3.94	2.24	2.56	2.40	4.91	5.19	7.16	2.37	2.27	3.69
	Sukno [40]	4339	2.91	3.45	3.44	2.75	4.23	5.35	6.98	3.09	2.58	3.82
	DLIN	4666	2.12	1.25	1.57	1.78	2.01	2.23	5.99	2.57	2.37	**1.83**

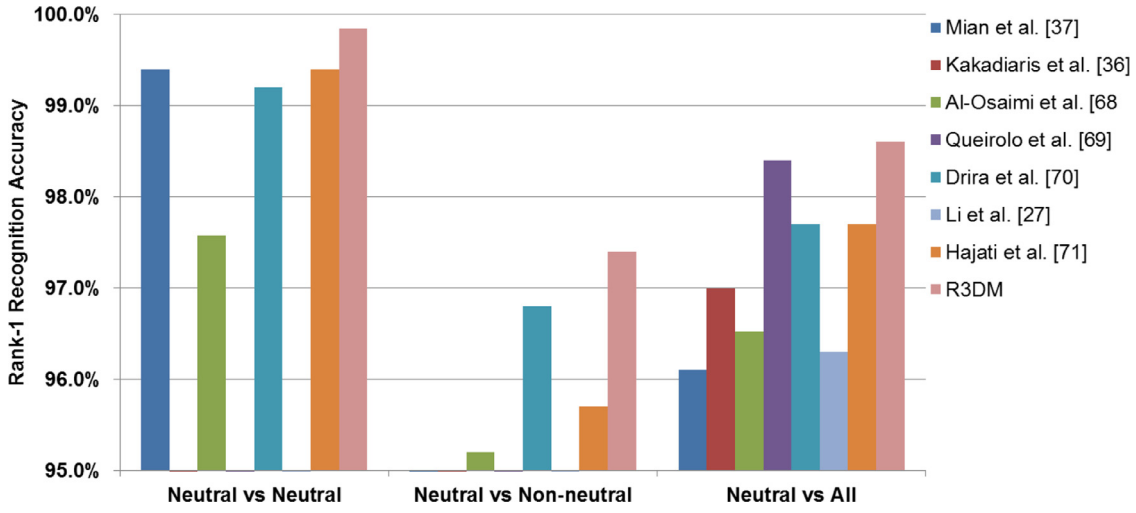


Fig. 9. Comparison of Rank-1 Recognition Rate for FRGCv2 on neutral and non-neutral expression scans. Missing bars show that authors have not reported the particular results.

Table 4
Comparison of Rank-1 recognition results (in %age) with the state-of-the-art on Bosphorus dataset. Values highlighted in bold correspond to the best recognition rate.

Author	Expressions			Poses					Occlusions					All
	AU	Expr	All	YR < 90	YR90	PR	CR	All	Eye	Mouth	Glasses	Hair	All	
# scans	2150	647	2797	525	210	419	211	1365	105	105	104	67	381	**4543**
Drira et al. [71]	–	–	–	–	–	–	–	–	97.1	78.0	94.2	81.0	87.0	–
Berretti et al. [39]	–	–	95.7	81.6	45.7	98.3	93.4	88.6	–	–	–	–	93.2	93.4
Hajati et al. [72]	–	–	–	–	–	–	–	–	–	–	–	–	–	95.2*
Li et al. [27]	99.2	96.6	98.8	84.1	47.1	99.5	**99.1**	91.1	**100.0**	**100.0**	**100.0**	95.5	**99.2**	96.6
R3DM	**99.3**	**97.9**	**99.0**	**94.8**	**86.2**	**100.0**	98.6	95.7	**100.0**	97.8	**100.0**	**97.1**	98.9	**98.1**

* The algorithm was not tested on profile and occluded scans [72]. AU=Action Units; YR=Yaw Rotation; PR= Pitch Rotation; CR= Cross Rotation.

the state-of-the-art especially for the case of expressions since it encodes expressions in addition to identity information. Significant improvement (from 47.1% to 86.2%) is achieved in the case of large pose variations which demonstrates the capability of R3DM to generalize over missing data. Our algorithm performs consistently well on both the FRGCv2 and the Bosphorus datasets.

We emphasize that each identity of the test dataset appears only once in its respective R3DM and that too in the neutral expression. Facial expressions are encoded in the R3DM in an unsupervised manner using unlabelled faces with non-neutral expressions from a different dataset. The R3DM neither includes identity specific facial expressions nor all of the facial expression types.

6. Efficiency and computational cost

DLIN was trained on an Intel Core i7 3.4 GHz machine, 32GB RAM with one Tesla K40C GPU and a solid state hard drive. Training was done in Matconvenet [55] for three days to complete 200 epochs. All modules of the proposed method were implemented in

MATLAB™ on an Intel Core i7 3.4 GHz machine with 8GB RAM. Dense correspondence takes 10.13 s per scan. This includes detection of 11 facial landmarks using DLIN in 0.5 s, segmenting the image into its Voronoi regions in 1.34 s and establishing dense correspondence on a 3D image with 5 regions after keypoint detection and NN matching in 8.27 s. Developing the dense correspondence model on 300 images over 13,394 vertices of FRGCv2 took 40 min. while it took 29 min. to do the same on 215 scans of the Bosphorus dataset over 13,975 vertices. Note that training the DLIN and establishing region based dense correspondences are performed offline. However, use of parallel computation on GPUs can considerably speed up the latter process as well.

The task for establishing dense correspondence on an unseen query face by fitting the R3DM and then performing face recognition takes only 3.16 s, which includes detection of 11 landmarks in 0.5 s, segmenting the image into Voronoi regions in 1.34 s, model fitting in 1.30 s and matching the probe with the gallery (size 466 in case of FRGCv2 and 105 in case of Bosphorus) in 5 ms (on average).

Empirical comparison of efficiency runs into a bottle neck due to unavailability of codes for dense correspondence. In the absence of codes from the original authors, it is difficult to perform a fair comparison with the state-of-the-art algorithms. Furthermore, comparison of timing using published results is not straightforward as they report timing for establishing correspondences between different number of vertices [1], on different number of 3D faces and using machines with different processors [27].

7. Conclusions

We presented an algorithm for dense correspondence over a large number of 3D faces across varying facial expressions and identities. We trained a Deep Landmark Identification Network (DLIN) using synthetic images to detect salient landmarks and used them to segment the 3D face into Voronoi regions by generating a geodesic distance map through level set curves. Keypoints in these regions were used to align similar regions across faces in a non-rigid manner and dense correspondence was established through NN search. We also proposed a Region based 3D Deformable Model (R3DM) to propagate the dense correspondences to large datasets efficiently. Experiments on benchmark datasets with challenging protocols show that our algorithm is faster and more accurate than existing state-of-the-art. Our algorithm is able to generate population specific accurate deformable 3D face models from scratch without relying on existing linear 3D face models such as the BFM. In the future, we intend to use our algorithm for phenotyping medical conditions such as Sleep Apnoea and Autistic Spectrum Disorder.

Acknowledgement

This research was supported by the Australian National Health and Medical Research Council (NHMRC) project grant number APP1109057. The authors thank NVIDIA for providing the GeForce GTX TITAN X GPU used in our experiments.

References

[1] S.Z. Gilani, F. Shafait, A. Mian, Shape-based automatic detection of a large number of 3D facial landmarks, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015, pp. 4639–4648.
[2] T. Funkhouser, P. Shilane, Partial matching of 3D shapes with priority-driven search, in: Eurographics Symposium on Geometry Processing, vol. 256, 2006, pp. 131–142.
[3] G. Passalis, P. Perakis, T. Theoharis, I.A. Kakadiaris, Using facial symmetry to handle pose variations in real-world 3D face recognition, IEEE TPAMI 33 (10) (2011) 1938–1951.
[4] U. Prabhu, J. Heo, M. Savvides, Unconstrained pose-invariant face recognition using 3D generic elastic models, IEEE TPAMI 33 (10) (2011) 1952–1961.
[5] V. Blanz, T. Vetter, Face recognition based on fitting a 3D morphable model, IEEE TPAMI 25 (9) (2003) 1063–1074.
[6] P. Hammond, et al., The use of 3D face shape modelling in dysmorphology, Arch. Dis. Child. 92 (12) (2007) 1120.
[7] P. Hammond, C. Forster-Gibson, A. Chudley, et al., Face–brain asymmetry in autism spectrum disorders, Mol. Psychiatry 13 (6) (2008) 614–623.
[8] A. Almukhtar, A. Ayoub, B. Khambay, J. McDonald, X. Ju, State-of-the-art three--dimensional analysis of soft tissue changes following le fort i maxillary advancement, Br. J. Oral Maxillofacial Surg. 54 (7) (2016) 812–817.
[9] D. Aiger, N. Mitra, D. Cohen, 4-points congruent sets for robust pairwise surface registration, ACM TOG 27 (3) (2008) 85.
[10] B. Brown, S. Rusinkiewicz, Global non-rigid alignment of 3-D scans, ACM TOG 26 (3) (2007) 21.
[11] W. Chang, M. Zwicker, Automatic registration for articulated shapes, Comput. Graphics Forum 27 (5) (2008) 1459–1468.
[12] R.H. Davies, C.J. Twining, T.F. Cootes, J.C. Waterton, C.J. Taylor, 3D statistical shape models using direct optimisation of description length, ECCV, Springer, 2002.
[13] R. Davies, C. Twining, C. Taylor, Statistical Models of Shape: Optimisation and Evaluation, Springer, 2008.
[14] T. Heimann, H.-P. Meinzer, Statistical shape models for 3D medical image segmentation: a review, Med. Image Anal. 13 (4) (2009) 543–563.
[15] M. Alexa, Recent advances in mesh morphing, Comput. Graphics Forum 21 (2) (2002) 173–198.
[16] H. Mirzaalian, G. Hamarneh, T. Lee, A graph-based approach to skin mole matching incorporating template-normalized coordinates, in: IEEE CVPR, 2009.
[17] O. Van Kaick, H. Zhang, G. Hamarneh, D. Cohen-Or, A survey on shape correspondence, in: Computer Graphics Forum, vol. 30, 2011, pp. 1681–1707.
[18] C. Creusot, N. Pears, J. Austin, A machine-learning approach to keypoint detection and landmarking on 3D meshes, IJCV 102 (1–3) (2013) 146–179.
[19] P. Perakis, G. Passalis, T. Theoharis, I.A. Kakadiaris, 3D facial landmark detection under large yaw and expression variations, IEEE TPAMI 35 (7) (2013) 1552–1564.
[20] A. Salazar, S. Wuhrer, C. Shu, F. Prieto, Fully automatic expression-invariant face correspondence, Mach. Vis. Appl. 25 (4) (2014) 859–879.
[21] G. Passalis, I. Kakadiaris, T. Theoharis, G. Toderici, N. Murtuza, Evaluation of 3d face recognition in the presence of facial expressions: an annotated deformable model approach, IEEE CVPR Workshops, 2005.
[22] P. Perakis, T. Theoharis, I.A. Kakadiaris, Feature fusion for facial landmark detection, Pattern Recognit. 47 (9) (2014) 2783–2793.
[23] Y. Sun, M.A. Abidi, Surface matching by 3D point's fingerprint, IEEE ICCV, 2001.
[24] Y. Sun, J. Paik, A. Koschan, D. Page, M. Abidi, Point fingerprint: a new 3D object representation scheme, IEEE Trans. Syst. Man Cybern. Part B 33 (4) (2003) 712–717.
[25] V. Blanz, T. Vetter, A morphable model for the synthesis of 3d faces, in: ACM Conference on Computer Graphics and Interactive Techniques, 1999.
[26] V. Blanz, K. Scherbaum, H.-P. Seidel, Fitting a morphable model to 3d scans of faces, IEEE ICCV, 2007.
[27] H. Li, D. Huang, J.-M. Morvan, Y. Wang, 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. 113 (2) (2014) 128–142.
[28] L. Farkas, Anthropometry of the head and face in clinical practice, in: Anthropometry of the head and face, second ed., 1994, pp. 71–111.
[29] G. Peyré, L. Cohen, Geodesic computations for fast and accurate surface remeshing and parameterization, in: Elliptic and Parabolic Problems, Springer, 2005, pp. 157–171.
[30] B.C. Munsell, P. Dalal, S. Wang, Evaluating shape correspondence for statistical shape analysis: a benchmark study, IEEE TPAMI 30 (11) (2008) 2023–2039.
[31] P. Phillips, P. Flynn, T. Scruggs, K. Bowyer, et al., Overview of the face recognition grand challenge, IEEE CVPR, 2005.
[32] A. Savran, N. Alyüz, H. Dibeklioğlu, O. Çeliktutan, B. Gökberk, B. Sankur, L. Akarun, Bosphorus database for 3D face analysis, in: Biometrics and Identity Management, Springer, 2008, pp. 47–56.
[33] S. Wang, Y. Wang, M. Jin, X.D. Gu, D. Samaras, Conformal geometry and its applications on 3D shape matching, recognition, and stitching, IEEE TPAMI 29 (7) (2007) 1209–1220.
[34] J. Novatnack, K. Nishino, Scale-dependent/invariant local 3D shape descriptors for fully automatic registration of multiple sets of range images, ECCV, Springer, 2008.
[35] L. Yin, X. Wei, et al., A 3D facial expression database for facial behavior research, in: Automatic Face and Gesture Recognition, 2006, pp. 211–216.
[36] I.A. Kakadiaris, G. Passalis, G. Toderici, M. Murtuza, Y. Lu, N. Karampatziakis, T. Theoharis, Three-dimensional face recognition in the presence of facial expressions: an annotated deformable model approach, IEEE TPAMI 29 (4) (2007) 640–649.
[37] A. Mian, M. Bennamoun, R. Owens, Keypoint detection and local feature matching for textured 3D face recognition, IJCV 79 (1) (2008) 1–12.
[38] D. Smeets, J. Keustermans, D. Vandermeulen, P. Suetens, meshsift: local surface features for 3D face recognition under expression variations and partial data, Comput. Vision Image Understanding 117 (2) (2013) 158–169.
[39] S. Berretti, N. Werghi, A. Del Bimbo, P. Pala, Matching 3D face scans using interest points and local histogram descriptors, Comput. Graph. 37 (5) (2013) 509–525.
[40] F.M. Sukno, J.L. Waddington, P.F. Whelan, 3-D facial landmark localization with asymmetry patterns and shape regression from incomplete local features, Trans. Cybern. 45 (9) (2015) 1717–1730.

[41] S.Z. Gilani, K. Rooney, F. Shafait, M. Walters, A. Mian, Geometric facial gender scoring: objectivity of perception, PLoS ONE 9 (6) (2014).
[42] D. Song, D. Tao, Biologically inspired feature manifold for scene classification, IEEE Trans. Image Process. 19 (1) (2010) 174–184.
[43] D. Song, D. Tao, Discrminative geometry preserving projections, in: 2009 16th IEEE International Conference on Image Processing (ICIP), IEEE, 2009, pp. 2457–2460.
[44] A. Krizhevsky, I. Sutskever, G.E. Hinton, Imagenet classification with deep convolutional neural networks, in: Advances in Neural Information Processing Systems, 2012, pp. 1097–1105.
[45] R. Girshick, J. Donahue, T. Darrell, J. Malik, Rich feature hierarchies for accurate object detection and semantic segmentation, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2014, pp. 580–587.
[46] P.H. Pinheiro, R. Collobert, Recurrent convolutional neural networks for scene labeling, ICML, 2014.
[47] C. Farabet, C. Couprie, L. Najman, Y. LeCun, Learning hierarchical features for scene labeling, Pattern Anal. Mach. Intell. IEEE Trans. 35 (8) (2013) 1915–1929.
[48] D. Ciresan, A. Giusti, L.M. Gambardella, J. Schmidhuber, Deep neural networks segment neuronal membranes in electron microscopy images, in: Advances in Neural Information Processing Systems, 2012, pp. 2843–2851.
[49] J.L. Long, N. Zhang, T. Darrell, Do convnets learn correspondence? in: Advances in Neural Information Processing Systems, 2014, pp. 1601–1609.
[50] J. Long, E. Shelhamer, T. Darrell, Fully convolutional networks for semantic segmentation, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015, pp. 3431–3440.
[51] J. DÉrico, Surface fitting using gridfit, MATLAB Central File Exchange, 2008.
[52] A. Mian, Robust realtime feature detection in raw 3d face images, in: Applications of Computer Vision (WACV), 2011 IEEE Workshop on, IEEE, 2011, pp. 220–226.
[53] S.Z. Gilani, A. Mian, Perceptual differences between men and women: a 3D facial morphometric perspective, IEEE 22nd International Conference on Pattern Recognition (ICPR), 2014.
[54] S.Z. Gilani, F. Shafait, A. Mian, Biologically significant facial landmarks: how significant are they for gender classification? IEEE International Conference on Digital Image Computing: Techniques and Applications (DICTA), 2013.
[55] A. Vedaldi, K. Lenc, Matconvnet – convolutional neural networks for matlab, 2015.
[56] P. Paysan, R. Knothe, B. Amberg, S. Romdhani, T. Vetter, A 3D face model for pose and illumination invariant face recognition, in: International Conference on Advanced Video and Signal Based Surveillance, IEEE, 2009, pp. 296–301.
[57] P.-L. Hsieh, C. Ma, J. Yu, H. Li, Unconstrained realtime facial performance capture, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015, pp. 1675–1683.
[58] S. Bouaziz, Y. Wang, M. Pauly, Online modeling for realtime facial animation, ACM Trans. Graph. 32 (4) (2013) 40.
[59] K. Aldridge, I.D. George, K.K. Cole, J.R. Austin, T.N. Takahashi, Y. Duan, J.H. Miles, Facial phenotypes in subgroups of prepubertal boys with autism spectrum disorders are correlated with clinical phenotypes, Mol. Autism 2 (1) (2011) 1.
[60] S.Z. Gilani, D.W. Tan, S.N. Russell-Smith, M.T. Maybery, A. Mian, P.R. Eastwood, F. Shafait, M. Goonewardene, A.J. Whitehouse, Sexually dimorphic facial features vary according to level of autistic-like traits in the general population, J. Neurodev. Disord. 7 (1) (2015) 1.
[61] A.J. Whitehouse, S.Z. Gilani, F. Shafait, A. Mian, D.W. Tan, M.T. Maybery, J.A. Keelan, R. Hart, D.J. Handelsman, M. Goonawardene, et al., Prenatal testosterone exposure is related to sexually dimorphic facial morphology in adulthood, Proc. R. Soc. B 282 (1816) (2015) 20151351.
[62] J.A. Sethian, Evolution, implementation, and application of level set and fast marching methods for advancing fronts, J. Comput. Phys. 169 (2) (2001) 503–555.
[63] G. Peyré, The numerical tours of signal processing-advanced computational signal and image processing, IEEE Comput. Sci. Eng. 13 (4) (2011) 94–97.
[64] A. Mian, M. Bennamoun, R. Owens, On the repeatability and quality of keypoints for local feature-based 3D object retrieval from cluttered scenes, IJCV 89 (2–3) (2010) 348–361.
[65] D. Rueckert, L. Sonoda, C. Hayes, et al., Nonrigid registration using free-form deformations: application to breast mr images, IEEE Trans. Med. Imaging 18 (8) (1999) 712–721.
[66] D.-J. Kroon, Finite iterative closest point, MATLAB Central File Exchange, 2009.
[67] J.L. Bentley, Multidimensional binary search trees used for associative searching, Commun. ACM 18 (9) (1975) 509–517.
[68] P. Szeptycki, M. Ardabilian, L. Chen, A coarse-to-fine curvature analysis-based rotation invariant 3D face landmarking, IEEE- Biometrics: Theory, Applications, and Systems, 2009.
[69] H. Zhang, A. Sheffer, D. Cohen, et al., Deformation-driven shape correspondence, in: Computer Graphics Forum, vol. 27, 2008, pp. 1431–1439.
[70] V. Kraevoy, A. Sheffer, Cross-parameterization and compatible remeshing of 3D models, in: ACM TOG, vol. 23, 2004, pp. 861–869.
[71] H. Drira, B. Ben Amor, A. Srivastava, M. Daoudi, R. Slama, 3D face recognition under expressions, occlusions, and pose variations, IEEE TPAMI 35 (9) (2013) 2270–2283.
[72] F. Hajati, A. Cheraghian, S. Gheisari, Y. Gao, A.S. Mian, Surface geodesic pattern for 3d deformable texture matching, Pattern Recognit. 62 (2017) 21–32.

Syed Zulqarnain Gilani received his B.Sc Engineering degree from National University of Sciences and Technology (NUST), Pakistan. He later did his MS in Electrical Engineering from the same university in 2009 and secured the Presidents Gold Medal. He served as an Assistant Professor in the same university. He recently completed his Ph.D. degree from the School of Computer Science at University of Western Australia. His research interests include 3D Morphometric Face Analysis, pattern recognition and machine learning.

Ajmal Mian completed his Ph.D. with distinction from The University of Western Australia in 2006 and received the Australasian Distinguished Doctoral Dissertation Award from Computing Research and Education Association of Australia in 2007. He received two prestigious fellowships; the Australian Post-Doctoral Fellowship in 2008 and the Australian Research Fellowship in 2011. He was named the West Australian Early Career Scientist of the Year 2012. He has secured seven national competitive research grants from the Australian Research Council and the National Health and Medical Research Council. He is currently with the School of Computer Science and Software Engineering, The University of Western Australia. His research interests include computer vision, pattern recognition, machine learning, multimodal biometrics, and hyperspectral image analysis.

Peter Eastwood completed his Ph.D. studies in respiratory muscle physiology at the University of Western Australian and Sir Charles Gardiner Hospital (SCGH) and his Postdoctoral studies in control of breathing during sleep as a NH&MRC CJ Martin Fellow at University of Wisconsin, Madison, Wisconsin, U.S.A. and SCGH, Nedlands, W.A. He received his undergraduate training in Physiology and Health Science at Pennsylvania State University and Lock Haven University, Pennsylvania, U.S.A. He is currently a Winthrop Professor in the School of Anatomy & Human Biology, University of Western Australia apart from being NH&MRC Senior Research Fellow, West Australian Sleep Disorders Research Institute, SCGH and an Adjunct Professor, School of Physiotherapy, Curtin University of Technology. Prof Eastwood is also the Editor-in-Chief of Respirology. His research interests include 3D imaging and computational biomechanics to assess sleep apnoea, the upper airway during wakefulness, sleep and general anaesthesia and understanding the relationship between upper airway (pharyngeal) function during anaesthesia and sleep.