Nasal Patches and Curves for Expression-Robust 3D Face Recognition

Mehryar Emambakhsh and Adrian Evans

Abstract—The potential of the nasal region for expression robust 3D face recognition is thoroughly investigated by a novel five-step algorithm. First, the nose tip location is coarsely detected and the face is segmented, aligned and the nasal region cropped. Then, a very accurate and consistent nasal landmarking algorithm detects seven keypoints on the nasal region. In the third step, a feature extraction algorithm based on the surface normals of Gabor-wavelet filtered depth maps is utilised and, then, a set of spherical patches and curves are localised over the nasal region to provide the feature descriptors. The last step applies a genetic algorithm-based feature selector to detect the most stable patches and curves over different facial expressions. The algorithm provides the highest reported nasal region-based recognition ranks on the FRGC, Bosphorus and BU-3DFE datasets. The results are comparable with, and in many cases better than, many state-of-the-art 3D face recognition algorithms, which use the whole facial domain. The proposed method does not rely on sophisticated alignment or denoising steps, is very robust when only one sample per subject is used in the gallery, and does not require a training step for the landmarking algorithm.

Index Terms—Face recognition, facial landmarking, nose region, feature selection, Gabor wavelets, surface normals

✦

1 INTRODUCTION

W HILE much previous research on expression invariant 3D face recognition has focused on modelling expressions and detecting expression insensitive facial parts, there have been relatively few studies evaluating the potential of the nasal region for addressing this issue. Despite this, the nose has a number of salient features that make it suitable for expression robust recognition. It can be easily detected, due to its discriminant curvature and convexity [1], is difficult to hide without attracting suspicion [2], [3], is relatively stable over various facial expressions ([1], [4], [5], [6], [7], [8], [9]) and is rarely affected by unintentional occlusions caused by hair and scarves. Although it has been reported that the 2D image of the nose has too few discriminant features to be used as a reliable region for human identification [10], its 3D surface has much undiscovered potential. This paper further investigates the 3D nasal region for human identity authentication and verification purposes and presents a novel algorithm that provides very high discriminant strength, comparable with recent 3D face recognition algorithms, which use the whole facial domain.

The proposed approach is based on a very consistent and accurate landmarking algorithm, which overcomes the issue of robust segmentation of the nasal region. The algorithm first finds an approximate location of the nose tip and then finely tunes its location, while accurately determining the position of the nasal root and detecting the symmetry plane of the face. Next, the locations of three sets of landmarks are found: subnasale, eye corners and nasal alar groove. These landmarks are utilised on feature maps created by applying multi-resolution Gabor wavelets to the surface normals of the depth map. Two types of feature descriptors are used: spherical patches and nasal curves. Feature selection is then performed using a heuristic genetic algorithm (GA) and, finally, the expression-robust feature descriptors are applied to the well-known and widely used 3D Face Recognition Grand Challenge (FRGC) [11], Bosphorus [12] and Binghamton University 3D Facial Expression (BU-3DFE) [13] datasets.

Results show the algorithm's high potential to recognise nasal regions, and hence faces, over different expressions, with very few gallery samples per subject. The highest rank-one recognition rates (R_1RR) achieved are: 1) a R_1RR of 97.9 percent and equal error rate (EER) of 2.4 percent for FRGC v2.0 and receiver operator characteristic (ROC) III experiments, respectively; 2) a R_1RR of 98.45 and 98.5 percent for FRGC's neutral versus neutral and neutral versus non-neutral samples, respectively; 3) a R_1RR of 96.2 percent when one gallery sample per subject is used for the FRGC dataset (482 gallery samples (subjects) versus 4,330 probe samples); 4) a R_1RR of 95.35 percent for the Bosphorus dataset when 2,797 scans of 105 subjects are used as probes and the set of 105 neutral scans (one per subject) is used as the galley.

The remainder of the paper is organized as follows. After the literature review provided in Section 2, the alignment and nasal region cropping steps, followed by the nasal region landmarking, are detailed in Section 3. The feature extraction algorithm is described in Sections 4 and 5 explains the feature descriptors used. The feature selection algorithm is detailed in Section 6 and experimental results, including a thorough comparison with previous work, is provided in Section 7. Finally, conclusions are given in Section 8.

- M. Emambakhsh is with the Institute of Sensors, Signals and Systems at Heriot-Watt University, Edinburgh, United Kingdom.
  E-mail: m.emambakhsh@hw.ac.uk.
- A. Evans is with the Department of Electronic and Electrical Engineering, University of Bath, Bath, United Kingdom.
  E-mail: A.N.Evans@bath.ac.uk.

1.1 Scientific Contribution and Comparison with Previous Work

The major contribution of this paper is a novel surface normal-based recognition algorithm that provides a thorough evaluation of the recognition potential of the 3D nasal region. The results achieved are not only better than previous 3D nose recognition algorithms but also higher than many recognition algorithms that employ the whole face. The algorithm employs a novel, training-free, highly consistent and accurate landmarking algorithm for the nasal region and a robust feature space, based on the response of Gabor wavelets to surface normal vectors, is also introduced. To localise the expression robust regions on the nose a heuristic GA feature selection is applied to two different geometrical feature descriptors. Because of the smoothing effects of the Gabor wavelets, there is no need for sophisticated denoising algorithms. Indeed, only simple median filtering is required for the surface normals, even with noisy datasets such as the FRGC Spring 2003 folder. An additional advantage of the proposed approach is that a fast Principal Component Analysis (PCA)-based self-dependent method can be employed for facial pose correction. This eliminates the need for sophisticated pose correction algorithms or reference faces for fine tuning the alignment.

The proposed approach significantly extends our previous work [4] in which the nasal landmarking and recognition was performed on the depth map. This paper increases the number of landmarks and their detection accuracy and presents new feature extraction and selection algorithms. The work is inspired by recent algorithms on utilising facial normal vectors in 3D [14] and regional normal vectors [15]. To compare the new algorithms with previous approaches which used similar methodologies, the application of normals, computed over the nasal surface, is used for identification as well as the verification scenario. By using multiresolution Gabor wavelets the ability of the algorithm to handle more noisy samples is enhanced, providing higher R_1RR than the approach of Li et al. [15], which excluded the noisy FRGC Spring 2003 samples. This work also extends the application of facial curves, introduced as feature descriptors by Berretti et al. ([16] and [17]), to nasal spherical patches, producing a R_1RR increase of $> 2\%$, and showing a higher class separability for the spherical patches than for curves for 3D face recognition.

2 RECENT LITERATURE REVIEW

Robustness against the deformations caused by facial expressions has been a popular research topic in 3D face recognition. The face is a non-rigid object and therefore 3D matching techniques for rigid objects, such as the iterative closest point (ICP) algorithm [18], can become trapped in local minima and fail to provide accurate matching scores.

An empirical approach to deal with the variations caused by expressions is to capture a range of facial expressions for each subject and store them in the gallery [19]. Then, the facial biometric features of each test subject can be compared with all the stored expressions and a decision made on the identity of the subject. This method has numerous disadvantages: capturing a range of facial expressions for each subject is not always straightforward and requires a high storage capacity per subject. In addition, facial expressions will not

necessarily remain constant and may differ between the test and gallery captures [19].

One approach to overcome this problem is to use computer graphics algorithms to artificially create different expressions for each facial capture. In [20], expressions are learned using PCA eigenvectors and then used to regenerate the expressions on the probe samples. Although this approach does not require multiple samples per subject in the gallery, it is still vulnerable to the number of training samples used to model the facial expressions. Also, a universal definition of facial expression for all subjects still remains to be found [19] and the need to classify the expression types prior to face recognition increases the computational complexity.

Another approach is to employ region-based methods, in which the least variant parts of the face over different expressions are detected using facial segmentation [1], [6], [21], [22] or extracted using their expression invariant capabilities [5], [14], [15], [23]. Spreeuwers proposes a multiple regional approach based on a PCA-Linear Discriminant Analysis (LDA) feature extraction method [21]. In regional recognition, scores from a combination of different masks on the nose, cheek, forehead, chin and mouth are fused to finalise the decision making.

Alyüz et al. use a regional registration algorithm in conjunction with LDA classifiers, giving an expression robust 3D face recognition approach [6]. They also demonstrate that the nasal region has a high discriminatory power. A focus on integrating multiple regions is provided by Queirolo et al. [22] in which four regions (the upper face image, the whole face and two nasal regions) are segmented and stored for the gallery sessions before matching is performed using a novel matching criterion, called the surface interpenetration measure, and simulated annealing.
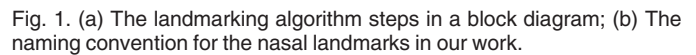
Using facial curves is another popular approach to 3D face recognition that can be categorised as a subset of regional algorithms. Drira et al. use the intersections of planes with the facial surface to define a set of radial curves which pass through the nose tip, and then perform a quality assessment in order to handle missing data and occlusions [24]. Another curve-based algorithm is proposed by Berretti et al. [17]. First, keypoints are detected on the facial surface and then the least variant curves on the face are selected using a statistical model and matched with those in the gallery. As an extension to curves, isogeodesic stripes centralised on the nose tip are used in an expression invariant 3D face recognition method that employs a novel descriptor, termed the 3D weighted walkthroughs, to quantify the differences between corresponding stripes [16]. In another curve-based approach, Drira et al. find geodesic curves on the nasal region for a subset of the FRGC dataset [7].

To overcome the sensitivity of holistic face recognition algorithms to expression variations, Mian et al. propose a landmark-based method, in conjunction with a localised feature descriptor that incorporates the 2D texture and 3D point clouds. In an alternative approach, Wang et al. apply shape difference boosting to the Bosphorus dataset to learn the expressions and identify those facial regions which remain constant over different expressions [23]. Instead of using depth or the point coordinates for 3D registration, Mohammadzade and Hatzinakos use the surface normals

of the points in conjunction with a Fisher's discriminant paradigm [14]. This approach selects the normals which maximise the concentration of within-class scatter while simultaneously maximising the between-class distribution. Recently, Li et al. proposed local normals histograms, captured from multiple rectangular regions on the face, to set up an expression-robust feature space and use a novel sparse classifier to perform the matching [15].

Despite the robustness of these algorithms against facial expressions, they often rely on accurate and consistent facial segmentation, which is not a straightforward task in 3D. To address this issue, some researches have focused on the nasal region, which shows high consistency over different expressions. For example, in one of the first investigations on 3D nose recognition, Chang et al. initially segment the face into different non-overlapping regions, using the curvature information [1]. Then, three overlapping nasal regions are detected and stored in the gallery. The same regions are segmented in the probe images and matched using the ICP algorithm. Wang et al. propose the use of local shape difference boosting for 3D face recognition and also apply the boosting algorithm to different nasal regions [5]. The regions are cropped using the intersection of spheres of radius r, centred on the nose tip, with the face surface. When the value of r was increased, the recognition ranks reached a maximum and then plateaued. A combination of the nasal region, forehead and eyes are used for a 2D/3D face recognition by Mian et al. [9]. A modified ICP algorithm is used for matching, in conjunction with a pattern rejector based on spherical face representation (SFR) and shift-invariant feature transform (SIFT) [25], producing high recognition ranks on the FRGC dataset, in particular for the neutral probes. Dibeklioğlu et al. used the Dijkstra algorithm to segment the nose and evaluated the performance using a subset of the Bosphorus dataset [8].

3 PREPROCESSING AND NASAL REGION LANDMARKING

The algorithm explained in [4] is used to crop the face. Next, median filtering with a 2.5×2.5 mm^2 mask size is applied twice on the cropped face. The image is then resampled to a uniform grid with 0.5 mm/pixel horizontal and vertical resolutions using Delaunay triangulation and aligned using the iterative PCA algorithm [9]. The aligned face is then intersected with three cylinders to crop the nasal region, according to [4]. The depth map of the cropped nasal region is again median filtered with a 2.5×2.5 mm^2 mask to further smooth its surface and decrease the spike noise effects. The block diagram in Fig. 1a shows how the landmarks in Fig. 1b are detected.

3.1 Local Minima Detector, Nose Tip Re-Localisation, Nasal Root and Subnasale Detection

First, an initial position of the nasal root ($\mathbf{L1^0}$) is detected by [4]. Then, the location of the nose tip ($\mathbf{L4^0}$), found in Section 3, is more finely tuned. Various planes, passing through $\mathbf{L4^0}$ with normals $\cos{(\theta_i)}\hat{a}_x + \sin{(\theta_i)}\hat{a}_y$ are intersected with the nose surface, where θ_i is the angle of the ith plane with the y-axis, and $\hat{a}_x = [1, 0]$ and $\hat{a}_y = [0, 1]$ are the

Fig. 1. (a) The landmarking algorithm steps in a block diagram; (b) The naming convention for the nasal landmarks in our work.

unit vectors along the x and y axes, respectively. This process results in several curves on the nasal region, shown in Fig. 3a.

The proposed landmarking algorithm relies on a minima detector, which finds a set of minima on rotated versions of the curves and then maps them to the original curve. The rotation is required because some of the original curves are strictly decreasing functions that do not have an actual minimum. Assuming $\mathbf{P} = [\mathbf{X_f}, \mathbf{Y_f}]$ is a $K \times 2$ matrix representing points of a curve, instead of directly differentiating $\mathbf{P}$ to find the minima on the curves, as proposed by Segundo et al. [26], the curves are first rotated in the z-axis (roll direction) by an angle α around a given point on the curve. This operation is given by,

$$\begin{cases} \mathbf{P}_r = \mathbf{P} \times \begin{bmatrix} \cos\alpha & \sin\alpha \\ -\sin\alpha & \cos\alpha \end{bmatrix} = [\mathbf{X_f^\alpha}, \mathbf{Y_f^\alpha}], \\ \mathbf{MIN} = V_{n,\alpha}(\mathbf{P}_r) \end{cases} \quad (1)$$

where $\mathbf{P}_r$ is the rotated version of $\mathbf{P}$ and the function $V_{n,\alpha}(.)$ finds the location of the n smallest local minima on $\mathbf{P}_r$ and then remaps them to the original curve $\mathbf{P}$ using the rotation angle α. The output $\mathbf{MIN}$ is an $n \times 2$ matrix, containing the locations of the n local minima. $V_{n,\alpha}(.)$ computes the first order differentiation (first order difference in discrete space) of $\mathbf{P_r}$, which is then given to the signum function to detect its sign changes. This finds the locations of all the local minima in $\mathbf{P_r}$, which are then sorted based on their value in ascending order and the n lowest are selected and rotated back to the original curve using α. Fig. 2 shows an example of this procedure for $n = 1$ (the global extremum).
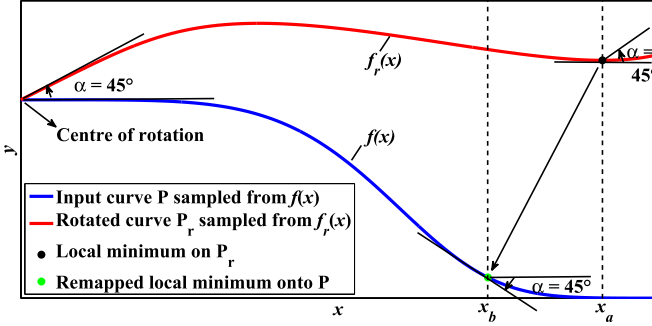
Fig. 2. The blue curve is strictly decreasing without any minima, while its 45 degree rotation (red curve) has a distinctive minimum. The interpretation of this procedure in the continuous space is $\alpha = \arctan\left(\left|\frac{df(x)}{dx}\big|_{x=x_b}\right|\right)$ while $\frac{df_r(x)}{dx}\big|_{x=x_a} = 0$.

The value allocated to α should be small enough in order to preserve the single-valued functionality of $\mathbf{X}_f^\alpha \rightarrow \mathbf{Y}_f^\alpha$, i.e., each projection of any point on $\mathbf{P_r}$ to the horizontal axis should correspond to only one point on the vertical axis. Based on the type of landmark to be extracted, the value of α is chosen using trial and error. The minima detector operator of (1) is applied to each curve in Fig. 3a as follows,

$$\mathbf{SMIN}_m = V_{1,\beta}(\mathbf{S}_{\gamma_m}^m) \qquad (2)$$

in which, $\mathbf{S}_{\gamma_m}^m$ represents the mth curve, which is rotated by γ_m around $\mathbf{L4^0}$. $\mathbf{SMIN}_m$ represents the location of the global minimum for each curve, which itself, constitutes a curve whose global maximum gives $\mathbf{L1^0}$, the initial location of the nasal root. Fig. 3 shows the set of curves $\mathbf{S}_\gamma^m$ (in blue), their minima (in red) and the maximum of the minima (green).

The nasal root and tip locations ($\mathbf{L1^0}$ and $\mathbf{L4^0}$) may be slightly inaccurate due to the depth variations caused by the noise and facial expressions. In order to improve the accuracy of their locations, for the points situated on a 5×5 mm^2 area (shown in Fig. 3b) around the nasal root and saddle, the following angular deviation is calculated,

$$\theta_z^k = \arccos \frac{\left|[L4_x^j - L1_x^i, L4_y^j - L1_y^i] \cdot \hat{a}_y\right|}{\left|[L4_x^j - L1_x^i, L4_y^j - L1_y^i]\right|}, \qquad (3)$$

in which, $[L1_x^i, L1_y^i]$ and $[L4_x^j, L4_y^j]$ are the $[x,y]$ projections of the two pairs of points $\mathbf{L1}^i$ and $\mathbf{L4}^j$ ($i = 1, \ldots, I, j = 1, \ldots, J$ and $k = 1, \ldots, J \times I$), which are selected from the overall I and J points on the region of interest (RoI) from the nasal root and tip regions, respectively, see Fig. 3b. θ_z^k is used to rotate the nose region in the roll direction and around $\mathbf{L4}^j$. Then the image is divided into the left and right halves. Assuming the rotated nasal region is translated so that the nose tip $\mathbf{L4}^j$ is at the origin, for the y-axis indices within the strip $\mathbf{S}_y^k$ shown in Fig. 3c (computed using θ_z^k), the objective function E^k is calculated by

$$E^k(\mathbf{L1}^i, \mathbf{L4}^j) = \max_{y \in \mathbf{S}_y^k}\left(\sum_x |\mathbf{Z}_L^k - \mathbf{Z}_R^k|\right) \qquad (4)$$

in which $\mathbf{Z}_L^k$ and $\mathbf{Z}_R^k$ are the depth maps of the flipped and cropped left and right sides of $\mathbf{L4}^j$ and $\mathbf{L1}^i$, respectively.
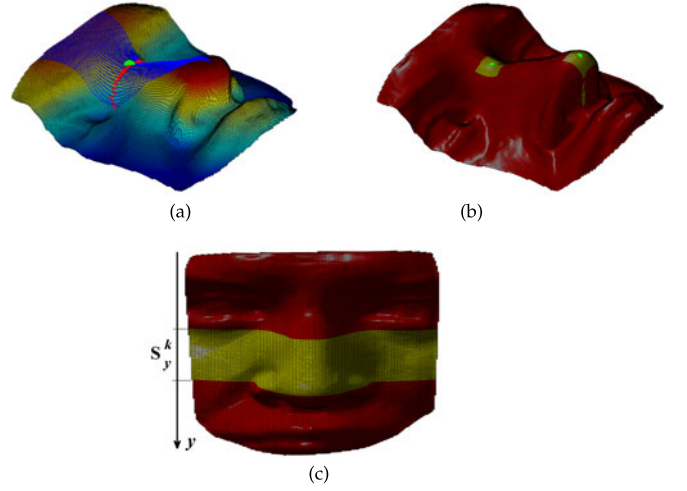

Fig. 3. (a) Localisation procedure for the nasal root, $\mathbf{L1^0}$, shown in green: the blue curves and red points represent planes intersections and their minima, respectively. (b) The 5×5 mm^2 RoIs around $\mathbf{L1}^i$ and $\mathbf{L4}^j$. (c) The horizontal strip $\mathbf{S}_y^k$ used in (4).

The two points $\mathbf{L1}^{opt}$ and $\mathbf{L4}^{opt}$ that minimise E^k have the most similar values of $\mathbf{Z}_L^{opt}$ and $\mathbf{Z}_R^{opt}$ and their projections onto the x axis ($L1_x^{opt}$ and $L4_x^{opt}$) correspond to the x values of the accurate nose tip and root locations such that,

$$\theta_z^{opt} = \arg\min_{\theta_z^k} E^k(\mathbf{L1}^i, \mathbf{L4}^j). \qquad (5)$$

This is an example of a "min-max" optimisation, which finds the best worst case for the optimum [27]. A plane passing through $\mathbf{L1}^{opt}$ and $\mathbf{L4}^{opt}$ is then intersected with the nose surface, with normal vector $[\cos\theta_z^{opt}, \sin\theta_z^{opt}]$, see Fig. 4b. The locations of the maximum and minimum of the resulting curve are the positions of $L1_y^{opt}$ and $L4_y^{opt}$. This procedure is illustrated in Fig. 4. $\mathbf{L4} = [L4_x^{opt}, L4_y^{opt}, L4_z^{opt}]$ and
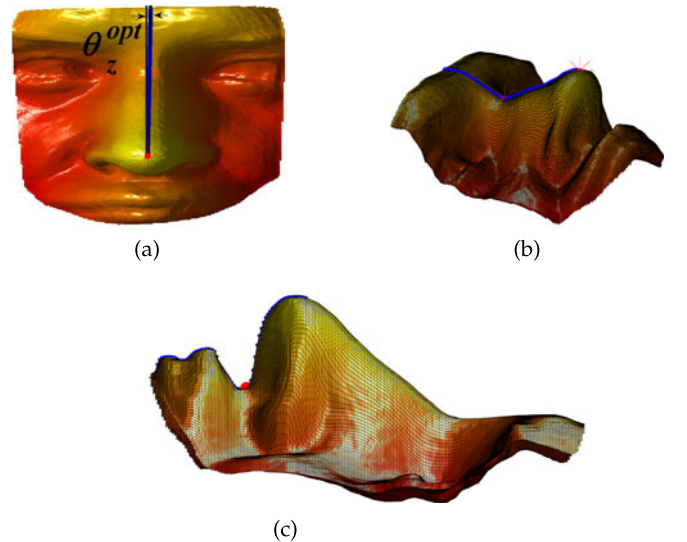

Fig. 4. Nasal root, tip and subnasale detection: (a) Updating the nasal region using θ_z^{opt}. (b) The maximum and minimum of a curve connecting the optimum $[L4_x^{opt}, L4_y^{opt}]$ and $[L1_x^{opt}, L1_y^{opt}]$ (blue curve) are used as $\mathbf{L4}$ and $\mathbf{L1}$, respectively (red points); (c) Blue points: symmetry plane intersection; Red point: the lowest minimum, detected as subnasale.
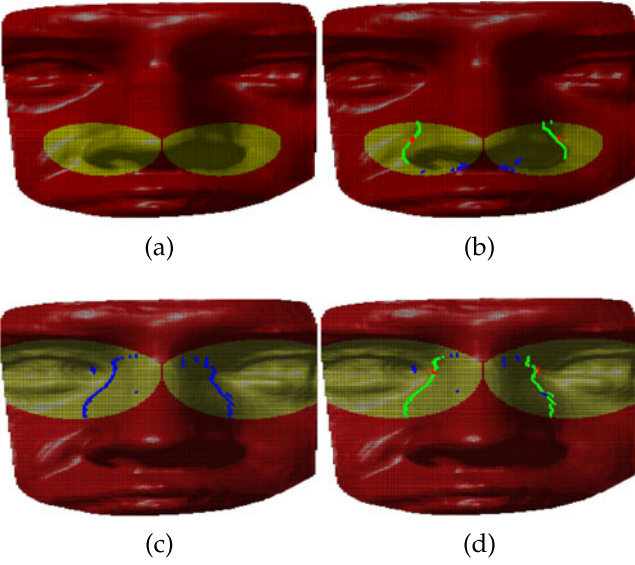
Fig. 5. (a) and (b) RoIs for detection of the nasal alar groove and eye corners landmarks. (b) and (d) show green and blue points as the inliers and outliers, while the red points are the selected locations for **L3**, **L6**, **L2** and **L7**, respectively.

$\mathbf{L1} = [L1_x^{opt}, L1_y^{opt}, L1_z^{opt}]$ give the final locations of the nose tip and nasal root, respectively.

The points on the same curve, which are located below the nasal tip **L4** (shown in Fig. 4c) are then rotated around **L4** by an angle ϕ. The location of the lowest minimum of the resulting curve ($\mathbf{S}_\phi$) provides the subnasale **L5** after applying (1), i.e., $\mathbf{L5} = V_{1,\phi}(\mathbf{S}_\phi)$. Finally, θ_z^{opt} is used to update the nose region and correct the pose by applying a roll directional rotation around **L4**.

3.2 Nose Alar Groove and Eye Corners Localisation

The location of the nose tip (**L4**) is moved to the origin and an RoI defined to detect the nasal alar grooves (Fig. 5a) by,

$$ r = \begin{cases} r_0 \cos^{a_1}(\theta) & 0 \leq \theta < \pi \\ r_0 \cos^{a_2}(\theta) & \pi \leq \theta < 2\pi \end{cases}, \tag{6} $$

where r_0, a_1 and a_2 are scalar constants determining the length and directivity of the lobes in the RoI. These are chosen to be able to crop the nasal alar region, while avoiding redundant parts (in subsequent experiments, r_0 = 30 mm, $a_1 = 4$ and $a_2 = 0.75$). r and θ are the distance from the nose tip and angular rotation from the horizontal axis passing the nose tip location, respectively. Similarly, the RoI used to detect the eye corners (**L2** and **L7**) is depicted in Fig. 5c and is found using,

$$ r' = \begin{cases} r'_0 \cos^{a_3}(\theta') & 0 \leq \theta' < \pi \\ r'_0 \cos^{a_4}(\theta') & \pi \leq \theta' < 2\pi \end{cases}, \tag{7} $$

in which $r'_0 = 45$ mm, $a_3 = 4$ and $a_4 = 0.75$. The polar coordinate system is characterised by r' and θ', which are the distance from the nasal root (**L1**) and the angular rotation from the horizontal axis passing **L1**, respectively.

Planes parallel with the xz-plane are then intersected with each row of the RoIs. For the jth and kth intersections over the nasal alar groove and eye corner RoIs, curves

$\mathbf{Q}^j$ and $\mathbf{Q}'^k$ are found, respectively, and (1) is used to find three minima for the nasal alar groove,

$$ \begin{cases} \mathbf{RMIN}_j = V_{3,\zeta}(\mathbf{Q}^j{}_\zeta) \\ \mathbf{LMIN}_j = V_{3,\eta}(\mathbf{Q}^j{}_\eta) \end{cases} \tag{8} $$

and the eye corners,

$$ \begin{cases} \mathbf{RMIN}'_k = V_{3,\zeta}(\mathbf{Q}'^k{}_\zeta) \\ \mathbf{LMIN}'_k = V_{3,\eta}(\mathbf{Q}'^k{}_\eta) \end{cases}, \tag{9} $$

where $\{\mathbf{RMIN}_j, \mathbf{LMIN}_j\}$ and $\{\mathbf{RMIN}'_k, \mathbf{LMIN}'_k\}$ are matrices with three rows, in which each row has the location of the lowest minimum found from the jth and kth rows of the RoIs, for the right and left sides of **L4** and **L1**, respectively. Then, for each row, $\mathbf{RMIN}_j$ and $\mathbf{LMIN}_j$ are compared and the pairs with the most similar euclidean distances to the nose tip **L4** selected. Using a similar approach, the distance of $\mathbf{RMIN}'_k$ and $\mathbf{LMIN}'_k$ to **L1** for each row is computed and those pairs with the most similar distances kept. Figs. 5b and 5c illustrate these processes.

The points found as candidates for the nasal alar groove and eye corners might contain some outliers. This is because of the imaging noise and deformations on the face due to the facial expressions. To remove the outliers, an iterative approach is used. First, the 3D euclidean distances between the points on each consecutive row are computed. Then the standard deviation (σ) of the resulting vector is used to reject the points whose σ is higher than a given threshold (in mm). This process continues until the number of inliers remains unchanged. Compared to the outlier removal method of [4], which uses K-means clustering as a criterion to localise the outliers, this approach is deterministic and, unlike K-means, is not vulnerable to empty clusters. The outlier removal algorithm results in the green points labelled as the inliers in Figs. 5b and 5d. The left and right pairs, which have the closest value of y to that of the nose tip are selected as **L3** and **L6**. Also, the points amongst the inliers in Fig. 5d, with the smallest depth values, are detected and the pair with the most similar distance to **L4** are selected as the eye corners (**L2** and **L7**). The eye corners and nasal alar groove landmarks are the red points in Figs. 5b and 5d, respectively.

4 FEATURE EXTRACTION

The proposed feature space is based on surface normals. For an aligned depth map of the nasal region, represented by its point clouds as $\mathbf{N} = [\mathbf{N_x}, \mathbf{N_y}, \mathbf{N_z}]$ the normals are $\mathbf{n} = [\mathbf{n_x}, \mathbf{n_y}, \mathbf{n_z}] = \nabla\mathbf{N}$, where $\mathbf{n_x} \circ \mathbf{n_x} + \mathbf{n_y} \circ \mathbf{n_y} + \mathbf{n_z} \circ \mathbf{n_z} = \mathbf{1}$ ($\circ$ and $\mathbf{1}$ represent the Hadamard product operator and matrix of ones, respectively). In order to reduce the sensitivity of the normal vectors to noise and enable the extraction of multi-resolution directional region-based information from the nasal region, instead of calculating the normal vectors directly from the nose surface, they are derived from the Gabor wavelet [28] filtered depth map. The algorithm proposed by Manjunath and Ma is used to minimise the wavelets overlap and redundancy in the filtered images [29].
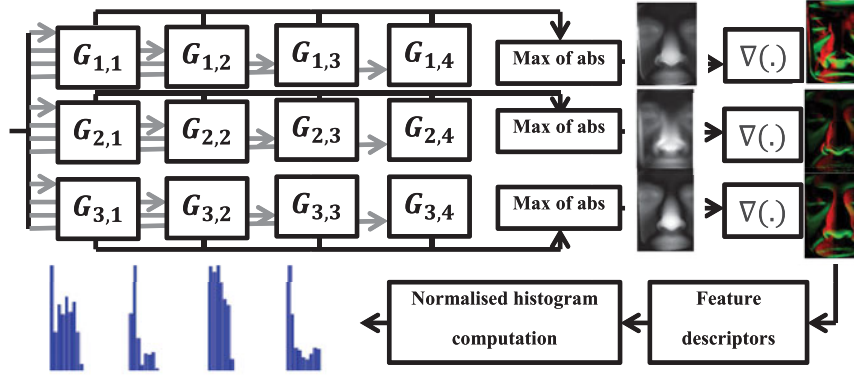
Fig. 6. The overall feature space creation procedure: 1) the wavelets are applied in different orientations and scales; 2) normals are computed on the maximum of absolute values of the filtered images per scale; 3) feature descriptors are applied; 4) normalised histograms are concatenated for all descriptors.

The discrete Fourier transform of the resampled Gabor wavelet $\mathbf{G}_{s,o}$ for the sth scale and oth orientation level ($s = \{1, 2, \ldots, s_m\}$ and $o = \{1, 2, \ldots, o_m\}$) is computed and its zero frequency component is set to zero. The Hadamard product of the resulting $\mathbf{G}_{s,o}^f$ and the Fourier transform of $\mathbf{N_z}$ is then calculated and the absolute value of its inverse Fourier transform is computed for each scale and orientation, i.e., $\mathbf{N}_{\mathbf{z}s,o}^{\mathbf{f}} = \left| \mathcal{F}^{-1}\left\{ \mathcal{F}\{\mathbf{N_z}\} \circ \mathbf{G}_{s,o}^f \right\} \right|$. The maximum of all the corresponding elements of the filtered images is computed over all orientations for each scale s: $\{\mathbf{NGm}_s | \forall i, j, o : \mathbf{NGm}_s(i, j) \geq \mathbf{N}_{\mathbf{z}s,o}^{\mathbf{f}}(i, j)\}$. In other words, $\mathbf{NGm}_s = \max_o(\mathbf{N}_{\mathbf{z}s,o}^{\mathbf{f}})$, where $\max(.)$ computes the maximum of the corresponding elements along orientations o. Finally, the normal vectors of the resulting per scale maximal map $\mathbf{NGm}_s$ is calculated using the aligned nose coordinate maps $\mathbf{N_x}$ and $\mathbf{N_y}$,

$$\begin{cases} \mathbf{n}_s = \nabla\big[\mathbf{N_x}, \mathbf{N_y}, \mathbf{NGm}_s\big] \\ s = 1, 2, \ldots, s_m \end{cases}, \qquad (10)$$

where $\mathbf{n}_s = [\mathbf{n}_{\mathbf{x}s}, \mathbf{n}_{\mathbf{y}s}, \mathbf{n}_{\mathbf{z}s}]$ is a block matrix containing the normal vectors for the sth scale level.

5   LOCALISED FEATURE DESCRIPTORS USING SPHERICAL PATCHES AND CURVES

The feature descriptors are used to define a part of the nasal region, containing a set of normal vectors from the Gabor wavelets filters. Histograms of the resulting feature vectors for the $\mathbf{X}$, $\mathbf{Y}$ and $\mathbf{Z}$ maps are concatenated to create the feature space. This procedure is illustrated in Fig. 6 for $s_m = 3$ and $o_m = 4$. The feature descriptors are used to reduce the dimensionality of the feature space, decrease the redundancy and enable the use of probabilistic feature selection to lower the sensitivity to facial expressions while maintaining the most discriminative parts.

The basic landmarks previously identified, see Fig. 1b, are used to create the new keypoints shown in Fig. 7a. These new landmarks are easily obtained by dividing the horizontal and vertical lines that connect the landmarks. A sphere centralised on each point is then intersected with the nasal surface and its inner parts are cropped. Then, the histogram of the normals of Gabor-wavelet filtered depth images are

computed, based on the procedure explained in Section 4. The intersection process is depicted in Fig. 7b. A set of spheres of identical radii (in this case 7 mm) are intersected with the nose surface. These spherical feature descriptors provide the capability to evaluate the potential of overlapping spherical regions on the nasal surface, when used as feature vectors.

Alternatively, using different pairs of landmarks, a set of orthogonal planes to the nasal region can be found. Intersecting the planes with the nose surface results in a set of curves on the nasal region. For example, the normal vector of a plane passing through two nasal landmarks $\mathbf{A_1}$ and $\mathbf{A_2}$, and orthogonal to the xy plane can be defined by $\hat{p}_{\mathbf{A_1 A_2}} = \hat{a}_z \times \frac{[\mathbf{A_1} - \mathbf{A_2}]}{\sqrt{[\mathbf{A_1} - \mathbf{A_2}]}}$, where $\hat{a}_z = [0, 0, 1]$ is the unit vector along the z-axis. When $\mathbf{A_1}$ and $\mathbf{A_2}$ are selected from the set of landmarks shown in Fig. 7c, they can be used to create the set of curves shown in Fig. 7d, which provide the feature descriptors. For each curve, the concatenated histograms of the x, y and z components of the normal vectors from the Gabor wavelet filters outputs are computed, giving the feature vector.

6   FEATURE SELECTION USING GA

The feature selection step selects those subsets of feature vectors extracted from the curves and spherical patches that
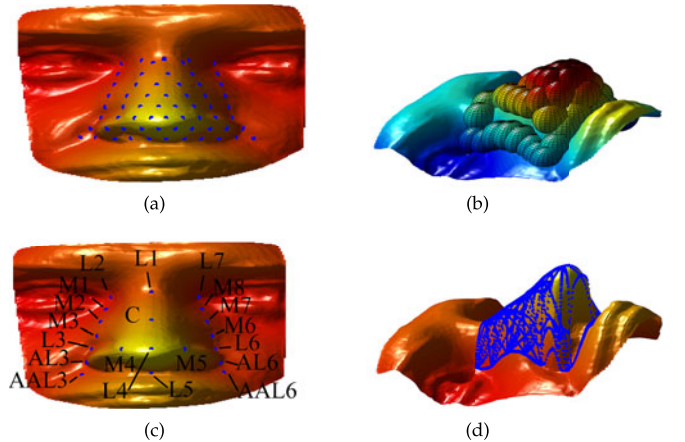

Fig. 7. (a) Grid of landmarks used for the spherical patches in (b). The nasal curves (d) are found using the combination of new landmarks, illustrated in (c).

are more robust against facial expressions. For a given feature descriptor and n different Gabor wavelets scales $\{s_1, s_2, \ldots, s_n\}$, the feature vector is computed by,

$$\begin{cases} \mathbf{F} = \left[\mathbf{F}_{s_1}, \mathbf{F}_{s_2}, \ldots, \mathbf{F}_{s_n} \right] \\ \mathbf{F}_{s_k} = \left[\mathbf{Fx}_{s_k}, \mathbf{Fy}_{s_k}, \mathbf{Fz}_{s_k} \right], \end{cases} \quad (11)$$

where $\mathbf{Fx}_{s_k}$, $\mathbf{Fy}_{s_k}$ and $\mathbf{Fz}_{s_k}$ are the features of the s_k^{th} scale, for the x, y and z surface normal components, respectively. For K feature descriptors, each feature set of the normal maps is represented by the concatenation of K different histograms, of length h_l from the feature descriptors, giving

$$\begin{cases} \mathbf{Fx}_{s_k} = \left[\mathbf{Hx}_{1,s_k}, \mathbf{Hx}_{2,s_k}, \ldots, \mathbf{Hx}_{K,s_k} \right] \\ \mathbf{Fy}_{s_k} = \left[\mathbf{Hy}_{1,s_k}, \mathbf{Hy}_{2,s_k}, \ldots, \mathbf{Hy}_{K,s_k} \right] \\ \mathbf{Fz}_{s_k} = \left[\mathbf{Hz}_{1,s_k}, \mathbf{Hz}_{2,s_k}, \ldots, \mathbf{Hz}_{K,s_k} \right]. \end{cases} \quad (12)$$

In (12), $\mathbf{Hx}_{i,s_k}$, $\mathbf{Hy}_{i,s_k}$ and $\mathbf{Hz}_{i,s_k}$ are the normalised histograms computed using the ith feature descriptor ($i = 1, \ldots, K$) for the s_k^{th} scale ($k = 1, \ldots, n$) on the normal map $\mathbf{n}_{s_k}$, which is computed using (10).

Here the aim is to find a binary vector to be used as a switch to select the most robust and remove the vulnerable feature descriptors to facial expressions. Using a $1 \times K$ binary vector $\mathbf{Bn}$, the vector $\mathbf{B}_{s_k}$, whose length is equal to the length of $\mathbf{Fx}_{s_k}$, $\mathbf{Fy}_{s_k}$ or $\mathbf{Fz}_{s_k}$ can be computed for the s_k^{th} scale by,

$$\begin{cases} \mathbf{B}_{s_k} = \left[\mathbf{B}_{1,s_k}, \mathbf{B}_{2,s_k}, \ldots, \mathbf{B}_{K,s_k} \right] \\ \mathbf{B}_{i,s_k} = \begin{cases} [0, 0, \ldots, 0] \text{ if } \mathbf{Bn}(i) = 0 \\ [1, 1, \ldots, 1] \text{ if } \mathbf{Bn}(i) = 1 \end{cases}. \end{cases} \quad (13)$$

The elements of $\mathbf{B}_{i,s_k}$ ($i = 1, \ldots, K$) are set to zero or one, depending on the value of the ith element of $\mathbf{Bn}$. Finally, $\mathbf{B}_{s_k}$ is concatenated over all scales to create a binary vector $\mathbf{B}$, whose length is equal to the feature space dimensionality,

$$\mathbf{B} = \left[\underbrace{\overbrace{[\mathbf{B}_{s_1}, \mathbf{B}_{s_1}, \mathbf{B}_{s_1}]}^{\text{for all normals in scale 1}}, \ldots, \overbrace{[\mathbf{B}_{s_n}, \mathbf{B}_{s_n}, \mathbf{B}_{s_n}]}^{\text{for all normals in scale n}}}_{\text{length}=s_n \times 3 \times K \times h_l} \right]. \quad (14)$$

The value of each element of $\mathbf{B}$ can be altered using the nucleus binary vector $\mathbf{Bn}$. A curve or patch is selected or omitted based on the value of $\mathbf{Bn}$ elements. If $\mathbf{Bn}(i) = 1$, then the ith curve or patch is selected, otherwise it is omitted. By grouping the neutral samples for the gallery and the non-neutral samples for the test phase, and varying $\mathbf{Bn}$, the most expression robust curves and patches can be selected. As shown in (13) and (14) when $\mathbf{B}_{s_1} = \mathbf{B}_{s_2} = \ldots = \mathbf{B}_{s_n}$, all curves and patches are selected or removed, simultaneously, for all scales. The resulting low dimensional samples are matched with those in the gallery using the Mahalanobis cosine distance,

$$\mathbf{D}_{g,p} = -\left(\frac{\mathbf{X}_g}{\sqrt{|\mathbf{X}_g \Sigma^{-1} \mathbf{X}_g{}^||}} \right) \Sigma^{-1} \left(\frac{\mathbf{X}_p}{\sqrt{|\mathbf{X}_p \Sigma^{-1} \mathbf{X}_p{}^||}} \right). \quad (15)$$

The Kernel Fisher's analysis (KFA) algorithm with polynomial kernel is applied to the feature space to project the features to a lower dimensional space using a supervised approach. If S_g and S_p are the number of gallery and probe samples, respectively, and d_p is the dimension of the projected subspace (in all subsequent experiments, $d_p = S_g$), $\mathbf{X}_\mathbf{g}$ and $\mathbf{X}_\mathbf{p}$ will be matrices of dimensions $S_g \times d$ and $S_p \times d$, respectively. Σ is the $d_p \times d_p$ covariance matrix computed over $\mathbf{X}_\mathbf{g}$, and $\mathbf{D}_{\mathbf{g,p}}$ is a $S_g \times S_p$ distance matrix containing the matching errors. To maximise the probability of assigning the test samples $\mathbf{X}_\mathbf{t}$ to their corresponding classes (subjects), when compared with the gallery samples $\mathbf{X}_\mathbf{g}$, $\mathbf{Bn}$ can be varied and its optimum found by,

$$\mathbf{Bn}_{opt} = \underset{\mathbf{Bn}}{\arg\max}\{R_1\} \quad (16)$$

in which R_1 is the average probability that the label corresponding to the smallest matching error, found by (15), is the same as the label of the probe sample. In other words, R_1 is the rank one recognition rate which is maximised as $\mathbf{Bn}$ is changed. The excellent capability of GA in high dimensional binary parameter space [22], [30] make it well suited for this non-convex optimisation problem. The GA used in this work is a modified Non-dominated Sorting Genetic Algorithm-II (NSGA-II) [31] which, in comparison with NSGA [32], is an elitism-based approach, relies on an improved sorting algorithm, has lower computational complexity and does not require sharing parameter assignment [31]. The modified NSGA-II incorporates elitism over the individuals that increase the diversity of the population in addition to those with better fitness output. The parameter assignments for the GA are explained in Section 7.3.

7 EXPERIMENTAL RESULTS

7.1 3D Datasets and Experiments

Three datasets are used to evaluate the proposed recognition algorithms. The first one is the FRGC dataset, which is widely recognised as the largest 3D face dataset, with 557 subjects. The captures in the dataset were obtained using the Minolta laser sensors over three different sets of sessions: Spring 2003, Fall 2003 and Spring 2004 [11]. The samples in the Spring 2003 folders are known as the v1.0 [15], while the collection in the other two folders constitute v2.0. FRGC v1.0 and v2.0 have 267 subjects (838 samples) and 466 subjects (4,007 samples), respectively. To evaluate the algorithm on this dataset, three sets of experiments are defined. For the first set, FRGC v2.0 is divided into a 466 samples gallery and 3,541 probe samples, an arrangement that has been extensively used in the literature [9], [15], [16], [17], [20], [23], [33], [34]. The second experiment is known is FRGC's ROC III on Exp III [11]. This is a verification scenario, which uses the between season 3D data samples. For this experiment, usually the equal error rate or 0.1 percent false accept rate (FAR) is reported. The third evaluation using FRGC is termed expression versus expression. FRGC consists of samples with neutral and non-neutral facial expressions, and using different sets of facial expressions for the probe samples enables neutral gallery versus neutral probe and neutral gallery versus non-neutral probe evaluations. The purpose of this experiment is to quantitatively evaluate

TABLE 1
Landmarking Consistency Error in mm

Dataset	Nasal root (L1)	Left eye corner (L2)	Left alar groove (L3)	Subnasale (L5)	Right alar groove (L6)	Right eye corner (L7)
Bosphorus	1.06 ± 0.58	1.76 ± 1.03	1.06 ± 0.62	1.11 ± 0.38	1.19 ± 0.60	2.12 ± 1.14
FRGC	2.04 ± 1.09	2.95 ± 1.61	1.29 ± 0.82	1.86 ± 0.85	1.22 ± 0.62	2.91 ± 1.53

how the performance of a face recognition system with a neutral gallery changes when the probe samples are replaced by non-neutral samples.

The other two datasets used are the Bosphorus 3D face dataset [12] and the BU-3DFE dataset [13], which contain captures of six prototypic expressions (anger, disgust, fear, happiness, sadness and surprise) in addition to neutral. BU-3DFE contains 100 subjects (56 female and 44 male) with age range 18 to 70 years. It contains four different levels of intensity of each facial expression and only one neutral sample per subject, making it one of the most challenging benchmarks for face recognition. For the Bosphorus database, one neutral sample per subject is used in the gallery (105 samples) and the remaining 2,797 samples as probes [15], [35]. For both the Bosphorus and BU-3DFE a specific expression robustness evaluation uses the neutral expression as the gallery and the captures for each expression in turn as probes [35], [36].

7.2 Landmarking Consistency and Accuracy

The face recognition rates reported in the following sections provide indirect evidence of the landmarking algorithm's consistency. However, to provide an independent assessment, the within-class similarity of the landmarking is investigated. In this evaluation, all subjects in the datasets are translated so that the nose tip is located at the origin. Then, the mean of the positions of each landmark for each subject's different facial expressions is computed. This process is performed for all subjects in the dataset and the averages and standard deviations are calculated.

For an ideal landmarking algorithm, the average for each landmark would be zero. However, due to the noise in the data and image acquisition errors, in practice the averages are non-zero. Although FRGC contains 557 subjects, some only have one sample and these are discarded, as it is impossible to compute the mean distance for such subjects.

Table 1 shows the results for the FRGC and Bosphorus datasets, when the rotation angles in (2) for the nasal root and $\mathbf{S}_\phi$ for the subnasale detection are $\max_m(\gamma_m) = \beta = \phi =$

Fig. 8. Precision curves for the proposed landmarking algorithm computed using Bosphorus.

$\pi/3$ and the angles in (8) and (9) for the eye corners and nasal alar groove landmarks are $\eta = -\zeta = \pi/4$. The errors are higher for the FRGC dataset as its samples are noisier, especially those in the Spring 2003 folder. The most consistent pair of landmarks on both datasets are the two nasal alar (L3 and L6), while the errors are slightly higher for the eye corners (L2 and L7). The location of subnasale (L5) is more consistently detected for the samples of the Bosphorus dataset. This is mainly due to the 3D reconstruction noise for the higher frequency regions, as is in subnasale, for the FRGC samples. The other important factor for a landmarking algorithm is its accuracy and, in order to evaluate this, the locations of the landmarks are compared with the ground truth provided by the Bosphorus dataset. First, the PCA alignment matrix and θ_z^{opt} are applied to the ground truth landmarks to remap the points to the aligned faces domain. The precision curve [38] is then computed for the landmarks $\{L2, L3, L4, L6, L7\}$. As the location of nasal root (L1) has not been assigned by the dataset providers, it is excluded from the accuracy evaluations. The curves shown in Fig. 8 are found over the action units samples (2,150 observations) and samples with neutral and non-neutral expressions in the Bosphorus dataset (653 observations), constituting 2,803 samples. For the range of > 10 mm, all the landmarks are detected with $\geq 95\%$ accuracy. A comparison with the landmarking results reported by Creusot et al. [37] is shown in Table 2. Although the results in [37] are provided for different types of facial expressions, considering the number of samples and the recognition rates, they can be computed for all three action units, neutral and non-neutral samples in the Bosphorus dataset. For the nasal alar and tip, the proposed algorithm has higher accuracy, in particular the accuracy for the nose tip is improved by 2 percent. However, for the eye corners, the algorithm in [37] performs 2 percent better. Although the proposed landmarking algorithm is not as robust as the approach of [37] in the cases of partial and self-occlusions, it has the advantage of not requiring a training step.

7.3 Robustness against Facial Expressions

This section explains the parameters used for the feature descriptors, Gabor wavelets, the KFA and the GA optimiser. The supplementary material, which can be found on the Computer Society Digital Library at http://doi. ieeecomputersociety.org/10.1109/TPAMI.2016.2565473, presents more extensive experiments that evaluate the effects of varying these parameters on the overall face recognition performance.

Feature descriptors and Gabor wavelets. The radius of the spherical patches is 11 mm, while the numbers of histogram bins are 21 and 15 for the spherical and nasal curves, respectively. The Gabor wavelets at each orientation and scale can be defined by the lower and higher frequency levels (Ω_l and
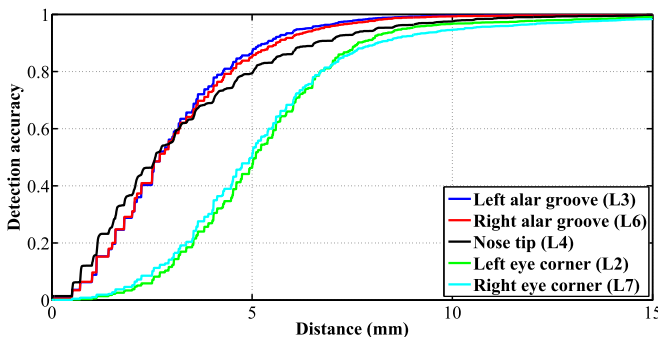
TABLE 2
Landmarking Precision Accuracy from the Ground Truth over the Bosphorus Dataset Samples

Algorithm	Threshold (mm)	L3	L6	L2	L7	Nose tip L4
Proposed method	< 10	99.55%	99.35%	96.69%	94.59%	97.52%
	< 12	99.62%	99.62%	97.73%	96.56%	99.04%
	< 15	99.66%	99.66%	99.04%	98.38%	99.66%
	< 20	99.69%	99.66%	99.59%	99.62%	99.79%
Creusot et al. [37]	< 10	97.96%	98.43%	98.82%	98.50%	95.47%
	< 12	99.18%	99.71%	99.65%	99.43%	98.15%
	< 15	99.82%	99.86%	99.93%	99.75%	98.97%
	< 20	99.90%	99.90%	99.93%	99.86%	99.33%

Ω_h) and the maximum number of orientation and scale levels (o_m and s_m). The parameter values used to obtain the results in the subsequent sections are: $\Omega_l = 0.05$, $\Omega_h = 0.7$, and $o_m = s_m = 4$. For the Gabor wavelet implementation the "Feature Extraction and Gabor Filtering" code is employed.[1] The supplementary material, available online, includes extensive results of the effects of the landmarking distribution for the spherical patches and Gabor wavelets parameters on the face recognition ranks.

KFA. The polynomial kernel $\left(\mathbf{X}_g^r \mathbf{X}_p^r + k_1 \right)^{\circ k_2}$ is applied to the $S_g \times d_r$ and $S_p \times d_r$ gallery and probe feature matrices $\mathbf{X}_g^r$ and $\mathbf{X}_p^r$, respectively, where $(.)^\circ$ is the Hadamard element-wise power operator. d_r is the input feature space dimensionality, which is reduced to d after KFA is utilised ($d = S_g - 1$ for all experiments in this paper). k_1 and k_2 are the scalar parameters of the kernel, set to 0 and 2.65, respectively. The KFA implementation is based on a modified version of the publicly available "Pretty Helpful Development" functions for face recognition toolbox (PHD toolbox).[2]

GA. The modified NSGA-II algorithm used to perform the global optimisation to maximise (16) stops when the variation in the stalled best fitness value is $< 10^{-4}$, which is achieved after approximately 70,000 iterations for the curves and spherical patches. The population size is 15 times the number of variables, while the Pareto and Cross-over and Migration fractions are 0.35, 0.8 and 0.2, respectively. The uniform creation function is used to initialise the population and the phenotype distance crowding measurement is utilised to compute the individuals' distance measure. The code is implemented using Matlab's "Global Optimization Toolbox."[3]

7.3.1 Feature Selection

The algorithm described in Section 6 is used to select the most discriminative patches and curves. For the feature selection stage, the neutral samples are employed for training and all the non-neutral samples for the test phase. The polynomial kernel is used for subspace projection. Then, the Mahalanobis cosine distance of (15) is applied for the matching step at each iteration of the GA. In order to quantitatively illustrate how a subset of selected features can boost the recognition ranks, as an example, Fig. 9 shows the face recognition ranks before and after applying the feature selection over the Bosphorus dataset. Over the first two

ranks, the average improvement for both curves and patches is $\approx 1\%$ (more extensive results for the feature selection and feature space parameters are provided in supplementary material, available online).

For all experiments presented in the following sections, different datasets are used for feature selection and testing, such that the test dataset is always completely unseen by the feature selector. Either the Bosphorus or the BU-3DFE dataset is used for feature selection, as these both have high variations in their expression types. The Bosphorus dataset is used for feature selection for all experiments in which FRGC or BU-3DFE are the test datasets, while BU-3DFE is used for feature selection with the Bosphorus dataset.

7.3.2 Variable Training Size for FRGC

To investigate the effects of using a different number of training samples per subject, the training size is increased from 1 to 7 for each subject in the FRGC dataset and the R_1RR computed for the selected patches and curves, see Table 3. In this experiment, all the folders in FRGC from different seasons are merged. Then, for each subject, the number of samples in the gallery is changed and the average R_1RR are reported after the samples are interchanged in the gallery and probe. The results show the high discrimination of the feature space for the spherical patches over the FRGC dataset. For example, when only one sample per subject is used in the gallery (482 gallery samples versus 4,330 probe samples to recognise 482 subjects), a R_1RR of 96.2 percent is achieved. To the best of our knowledge, this is the highest 3D nasal region recognition rank ever obtained from this dataset for a single training sample and comparable with many state-of-the-art 3D face recognition algorithms, which use the whole facial domain. Although the curves have lower R_1RR for one training sample per subject,

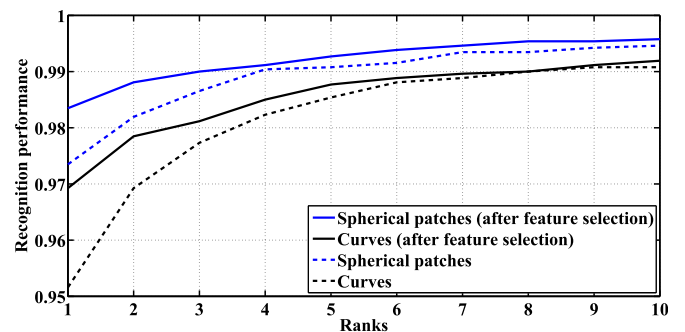Fig. 9. CMC curves before and after feature selection on Bosphorus: neutral gallery versus non-neutral probe.

1. http://old.vision.ece.ucsb.edu/texture/software/
2. luks.fe.uni-lj.si/sl/osebje/vitomir/face%5Ftools/PhDface/
3. http://mathworks.com/help/gads/gamultiobj.html

TABLE 3
R_1RR Performance for Varying the Training Size Per Subject, When All Samples
of the FRGC Dataset Are Merged from the Three Seasons

Feature descriptors	Number of gallery samples per subject (No. gallery samples/ No. probe samples)						
	1 (482/4330)	2 (880/3848)	3 (1206/3408)	4 (1432/3006)	5 (1610/2648)	6 (1752/2326)	7 (1757/2034)
Spherical patches	96.2%	98.9%	99.4%	99.6%	99.6%	99.7%	99.75%
Curves	91.6%	96.8%	98.1%	98.8%	98.9%	99.3%	99.4%

there is a big increase in their recognition performance when the samples per subject are increased. For instance, when two samples per subject are used in the gallery, the nasal curves R_1RR performance increases by $> 5\%$.

7.3.3 FRGC v2.0 and ROC III

The recognition performance for the FRGC v2.0 dataset and the ROC III experiment are widely used face recognition benchmarks. The dataset contains samples with different facial expressions for both the gallery and probe. The rank recognition performance shown in Fig. 10a increases with the rank to $> 99\%$ for the spherical patches. A similar trend exists for the cumulative matching characteristic (CMC) curves increasing with a high gradient for all ranks > 1.

The results of applying different patches and curves to this dataset are given in Table 4. The table also compares the performance of the proposed approach with recently proposed 3D face recognition techniques that employ the nasal region and also the whole face. For the FRGC v2.0 experiment, the spherical patches outperform the curves, with a R_1RR of 97.9 percent. For this experiment, although only the nasal region is used, the recognition rates of the proposed algorithm recognition rates are higher than or very close to the results of state-of-the-art approaches, in which the whole facial surface is utilised [5], [6], [9], [15], [17], [20], [24], [34]. The recognition performance of recent methods that only use the nasal region, for example [5], [6], [9], [21] is at least 3 percent worse, and in some cases up to 6 percent lower. Also, when the samples used in the probe set are changed from neutral to non-neutral, the proposed algorithm has the lowest variation in the R_1RR for both curves and spherical patches ($\approx 0.06\%$ for the spherical descriptors). It is interesting to note that there is even a slight increase in the recognition rates when non-neutral samples are used as probes (Table 4, last column), which shows how robustly the algorithm has learned the facial expressions. In comparison with the algorithm of Li et al. [15], which also uses the normal vectors as the basis for its feature space, there is $\approx 1.5\%$ improvement in the R_1RR and nearly a 4 percent increase in R_1RR when the non-neutral samples of FRGC are used as the probes.

ROC III curves are the FRGC's cross-seasonal verification scenario and these are plotted in Fig. 10b using a logarithmic scale. The EER and 0.1 percent FAR for the proposed approach are 2.4 and 93.5 percent, respectively. Although these results are higher than some previous 3D nose and face recognition methods (see the fifth column in Table 4), they are outperformed by a number of algorithms that use the whole face, which show a higher robustness for the verification scenario. For example, the algorithms of [20], [21], [24], [33] have lower identification performance than the proposed algorithm but a better verification performance. One conclusion from this may be that, for verification scenarios, the whole facial region might provide a higher confidence level when matched with a claimed identity.

7.3.4 Robustness against Different Facial Expressions

To more extensively evaluate the face recognition performance over different facial expressions the BU-3DFE and Bosphorus datasets are used, as they both contain samples with known expression types. For the BU-3DFE results the Bosphorus dataset is used for feature selection and vice versa.

For the first set of experiments, BU-3DFE is used for feature selection and the Bosphorus dataset for test. The gallery consists of 105 neutral samples, one for each subject, and all the remaining non-neutral and neutral samples are used as probes. A comparison with the previous results is provided in Table 5. While most comparison approaches have used the whole 3D face, the results of the nasal spherical patches is highly competitive when applied over the same dataset.

The second experiment is based on using various expressions for the probe samples. The Bosphorus dataset expressions include neutral, anger, disgust, fear, happiness, sadness, and surprise. To investigate the algorithm's performance in recognising probe samples with unseen facial
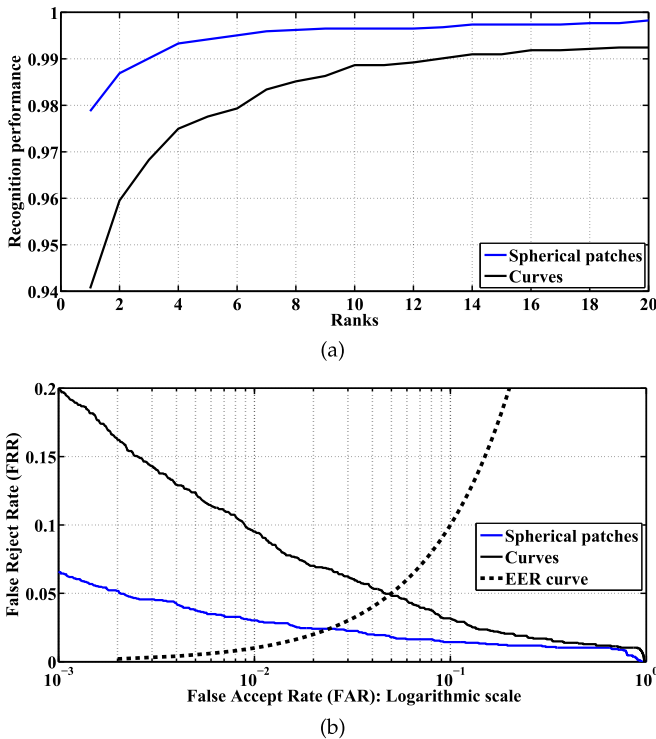

(a)

(b)

Fig. 10. (a) CMC curves for FRGC v2.0; (b) Between seasons verification results for FRGC: ROC III.

TABLE 4
Performance Comparison on the FRGC Dataset

Algorithm	Modality	Rank-one FRGC v2.0	EER ROC III	0.1% FAR ROC III	Neutral vs. Neutral	Neutral vs. Non-neutral
Spherical patches	3D Nose	97.9%	2.4%	93.5%	98.45% (R_1RR)	98.5% (R_1RR)
Curves		94.1%	4.9%	80.0%	95.8% (R_1RR)	97.5% (R_1RR)
Smeets et al. [34]	3D Face	89.6%	3.8%	77.2%	-	-
Osaimi et al. [20]	3D Face	96.5%	-	94.05%	98.35% (0.1% FAR)	97.8% (0.1% FAR)
Spreeuwers [21]	3D Face	99.0%	-	94.6%	-	-
	3D Nose	94.5%		83.7%		
Drira et al. (2013) [24]	3D Face	97.0%	-	97.1%	99.2% (R_1RR)	96.8% (R_1RR)
Alyüz et al. (2010) [6]	3D Face	97.5%	1.91%	85.6%	98.39% (R_1RR)	96.40% (R_1RR)
	3D Nose	91.81	-	-	-	-
Wang et al. (2010) [23]	3D Face	98.39%	-	98.04%	99.2% (0.1% FAR)	97.7% (0.1% FAR)
Wang et al. (2008) [5]	3D Nose	95% (44mm) 78% (24mm)	-	-	-	-
Drira et al. (2009) [7]	3D Face/Nose (125 gallery) (125 probe)	88% (Face) 77.5% (Nose)	-	-	-	-
Chang et al. [1]	3D Nose	-	Neutral 12% Non-neutral 23%	-	97.1% (R_1RR)	86.1% (R_1RR)
Emambakhsh et al. [4]	3D Nose	89.61%	Neutral 8% Non-neutral 18%	-	90.87% (R_1RR)	81.61% (R_1RR)
Li et al. (2014) [15]	3D Face	96.3%	-	-	98.0% (R_1RR)	94.2% (R_1RR)
Queirolo et al. [22]	3D Face	99.6%	-	96.6%	99.5% (R_1RR)	94.8% (R_1RR)
Berretti et al. (2013) [17]	3D Face	95.6%	-	-	97.3% (R_1RR)	92.8% (R_1RR)
Berretti et al. (2010) [16]	3D Face	94.15%	-	-	$\approx$97.3% (R_1RR)	$\approx$91.0% (R_1RR)
Mohammadzade and Hatzinakos (2013) [14]	3D Face	-	-	99.2%	-	-
Mian et al. (2008) [33]	3D Face	93.5%	-	Neutral 99.9% Non-neutral 92.7%	99%	86.7%
Mian et al. (2007) [9]	2D+3D Face	95.91%	-	99.3%	99.2%	95.37%
	2D+3D Nose	$\approx$92.2%	-	92.5%	$\approx$94.9%	$\approx$80.0%

expressions, for each subject one expression type is selected for the gallery samples and another for the probe. The results are shown in Table 6, where each column relates to a facial expression and the number of samples used in the probe set is given. The average recognition rank for the 76 subjects with more than one neutral sample per subject is reported for the neutral versus neutral experiment, as shown in the last column.

The spherical patches again outperform the nasal curves. The disgust expression deforms the noses more significantly than the other expressions and produces the lowest recognition ranks. On the other hand, the feature space is most invariant for the surprise expressions, as 100 percent of the

TABLE 5
Comparison of Some of the Previous Works on the 105 Subjects in the Bosphorus Dataset: the Numbers in Parentheses Show the No. Gallery Samples/ No. of Probe Samples

Algorithm	Modality and size	R_1RR
Spherical patches	3D Nose	95.35%
Curves	(105/2797)	86.1%
Li et al. (2014) [15]	3D Face (105/2797)	95.4%
Dibeklioğlu [8]	3D Nose (47/1527)	89.2%
	(47/423) [rotated]	62.6%
Li et al. (2011) [35]	3D Face (105/4561)	94.1%
Alyüz et al. (2008) [39]	3D Face (34/441)	95.9%
	(47/1508)	95.3%

samples are correctly recognised for this probe set. The algorithm is compared with that of Li et al. [35], in which similar evaluations are performed. The spherical patches show higher robustness, in particular for disgust ($\approx$ 12%), anger and fear ($\approx$ 6%), happy ($\approx$ 3%), sadness and surprise ($\approx$ 1.5%). The results reported in [35], however, are $\approx$ 1% better when a neutral expression is used for the probe samples.

For a similar experiment with the BU-3DFE dataset, the 100 neutral samples are used as the gallery and the remaining samples with different intensity levels of each expression used as probes. Table 7 shows the R_1RR for each expression type. As the expressions in the BU-3DFE dataset are significantly more intense than those in the Bosphorus dataset, the recognition rates are lower than those in Table 6. The spherical patches still provide the best R_1RR performance for five of the six expression types, with an average R_1RR of 88.9 percent, $\approx$ 4% above that of [36] which uses both the 2D and 3D information from the whole facial domain.

8 CONCLUSIONS

To address the problem of expression invariant face recognition, a novel algorithm is introduced, that utilises the 3D shape of nose. The algorithm is based on a highly consistent and accurate landmarking algorithm, a robust feature space, discriminative feature descriptors and feature selectors. The proposed method is applied over three well-known face datasets, FRGC, BU-3DFE and Bosphorus. The matching results show that the algorithm is very successful for both the

TABLE 6
$R_1 RR$ for Different Expression Types from the Bosphorus Dataset Used as Probes, with 105 Neutral Samples Used as the Gallery

Algorithm	Facial expression (No. of probe samples)						
	Happy (106)	Surprise (71)	Fear (70)	Sadness (66)	Anger (71)	Disgust (69)	Neutral (194)
Spherical patches	98.08%	100%	98.55%	96.92%	94.12%	88.24%	98.96%
Curves	85.85%	92.96%	87.14%	92.31%	84.06%	69.12%	96.88%
Li et al. [35]	95.28%	98.59%	92.86%	95.45%	88.73%	76.81%	100%

TABLE 7
$R_1 RR$ for Different Expression Types from the BU-3DFE Dataset Used as Probes, with 100 Neutral Samples Used as the Gallery

Algorithm	Facial expression (No. of probe samples)					
	Happy (400)	Surprise (400)	Fear (400)	Sadness (400)	Anger (400)	Disgust (400)
Spherical patches	88.5%	91.0%	89.8%	92.3%	90.1%	81.8%
Curves	81.8%	87.8%	85.3%	87.6%	83.2%	69.8%
Hajati et al. [36]	86.0%	84.0%	82.0%	85.0%	93.0%	79.0%

identification and verification scenarios, producing a $R_1 RR$ of 97.9 percent on FRGC v2.0, an EER of 2.4 percent on ROC III, and $R_1 RR$ of 98.45 and 98.5 percent for neutral and non-neutral probes, respectively. The proposed method does not rely on sophisticated preprocessing algorithms for its denoising and alignment. In addition, when there is only one sample per subject in the gallery, for all the merged folders of the FRGC dataset, a $R_1 RR$ of 96.2 percent is obtained. For the Bosphorus dataset a $R_1 RR$ of 95.35 percent is obtained when one neutral sample per subject is used for gallery and the remaining samples with various expression types as probes. The results of the proposed method reveal the high potential of the nasal region for 3D face recognition. The recognition ranks are not only significantly higher than previous nasal region-based algorithms, but also have a better performance than many 3D holistic and multi-modal approaches.

There are several aspects of the algorithm which can be utilised in other applications. For example, the feature extraction step, which is based on histograms of Gabor wavelet normals, can be applied to other 3D object recognition methods. Also, the feature selection paradigm described here can be easily applied to other pattern recognition algorithms, to maximise the within-class and between-class similarity and dissimilarity, respectively, enabling the extraction of a lower dimensional and less redundant feature space. The application of the proposed landmarking algorithm can be investigated for performing facial alignment, low dimensional face recognition and pattern rejection. Finally, the application of the feature extraction step on the whole facial region, to make it robust against occlusions is an interesting area of future research.

REFERENCES

[1] K. Chang, K. Bowyer, and P. Flynn, "Multiple nose region matching for 3D face recognition under varying facial expression," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 10, pp. 1695–1700, Oct. 2006.
[2] M. Emambakhsh and A. Evans, "Self-dependent 3D face rotational alignment using the nose region," in *Proc. 4th IET Int. Conf. Imag. Crime Detection Prevention*, 2011, pp. 1–6.
[3] A. Moorhouse, A. Evans, G. Atkinson, J. Sun, and M. Smith, "The nose on your face may not be so plain: Using the nose as a biometric," in *Proc. 3rd IET Int. Conf. Crime Detection Prevention*, 2009, pp. 1–6.
[4] M. Emambakhsh, A. Evans, and M. Smith, "Using nasal curves matching for expression robust 3D nose recognition," in *Proc. 6th IEEE Int. Conf. Biometrics: Theory, Appl. Syst.*, 2013, pp. 1–6.
[5] Y. Wang, X. Tang, J. Liu, G. Pan, and R. Xiao, "3D face recognition by local shape difference boosting," in *Proc. Eur. Conf. Comput. Vis.*, 2008, vol. 5302, pp. 603–616.
[6] N. Alyüz, B. Gökberk, and L. Akarun, "Regional registration for expression resistant 3-D face recognition," *IEEE Trans. Inf. Forensics Security*, vol. 5, no. 3, pp. 425–440, Sep. 2010.
[7] H. Drira, B. Amor, M. Daoudi, and A. Srivastava, "Nasal region contribution in 3D face biometrics using shape analysis framework," in *Proc. 3rd Int. Conf. Advances Biometrics*, 2009, pp. 357–366.
[8] H. Dibeklioğlu, B. Gökberk, and L. Akarun, "Nasal region-based 3D face recognition under pose and expression variations," in *Proc. 3rd Int. Conf. Adv. Biometrics*, 2009, pp. 309–318.
[9] 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.
[10] W. Zhao, R. Chellappa, P. Phillips, and A. Rosenfeld, "Face recognition: A literature survey," *ACM Comput. Surveys*, vol. 35, pp. 399–458, 2003.
[11] P. Phillips, P. Flynn, T. Scruggs, K. Bowyer, J. Chang, K. Hoffman, J. Marques, J. Min, and W. Worek, "Overview of the face recognition grand challenge," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, 2005, pp. 947–954.
[12] A. Savran, N. Alyüz, H. Dibeklioğlu, O. Çeliktutan, B. Gökberk, B. Sankur, and L. Akarun, "Bosphorus database for 3D face analysis," in *Biometrics and Identity Management*. Berlin, Germany: Springer, 2008, vol. 5372, pp. 47–56.
[13] 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. Autom. Face Gesture Recog.*, 2007, pp. 211–216.
[14] H. Mohammadzade and D. Hatzinakos, "Iterative closest normal point for 3D face recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 2, pp. 381–397, Feb. 2013.
[15] H. Li, D. Huang, J.-M. Morvan, L. Chen, and Y. Wang, "Expression-robust 3D face recognition via weighted sparse representation of multi-scale and multi-component local normal patterns," *Neurocomputing*, vol. 133, pp. 179–193, 2014.
[16] S. Berretti, A. Del Bimbo, and P. Pala, "3D face recognition using isogeodesic stripes," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 12, pp. 2162–2177, Dec. 2010.
[17] S. Berretti, A. Del Bimbo, and P. Pala, "Sparse matching of salient facial curves for recognition of 3-D faces with missing parts," *IEEE Trans. Inf. Forensics Security*, vol. 8, no. 2, pp. 374–389, Feb. 2013.
[18] P. Besl and N. McKay, "A method for registration of 3-D shapes," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 14, no. 2, pp. 239–256, Feb. 1992.
[19] K. Bowyer, K. Chang, and P. Flynn, "A survey of approaches and challenges in 3D and multi-modal 3D+2D face recognition," *Comput. Vis. Image Understanding*, vol. 101, pp. 1–15, 2006.

[20] F. Al-Osaimi, M. Bennamoun, and A. Mian, "An expression defor-
mation approach to non-rigid 3D face recognition," *Int. J. Comput.
Vis.*, vol. 81, no. 3, pp. 302–316, 2009.

[21] L. Spreeuwers, "Fast and accurate 3D face recognition," *Int. J.
Comput. Vis.*, vol. 93, no. 3, pp. 389–414, 2011.

[22] C. Queirolo, L. Silva, O. Bellon, and M. Segundo, "3D face recogni-
tion using simulated annealing and the surface interpenetration
measure," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 2,
pp. 206–219, Feb. 2010.

[23] Y. Wang, J. Liu, and X. Tang, "Robust 3D face recognition by local
shape difference boosting," *IEEE Trans. Pattern Anal. Mach. Intell.*,
vol. 32, no. 10, pp. 1858–1870, Oct. 2010.

[24] H. Drira, B. B. Amor, A. Srivastava, M. Daoudi, and R. Slama, "3D
face recognition under expressions, occlusions, and pose var-
iations," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 9,
pp. 2270–2283, Sep. 2013.

[25] D. Lowe, "Distinctive image features from scale-invariant key-
points," *Int. J. Comput. Vis.*, vol. 60, no. 2, pp. 91–110, 2004.

[26] M. Segundo, L. Silva, O. Bellon, and C. Queirolo, "Automatic face
segmentation and facial landmark detection in range images,"
IEEE Trans. Syst., Man, Cybern., vol. 40, no. 5, pp. 1319–1330,
Oct. 2010.

[27] H. Aissi, C. Bazgan, and D. Vanderpooten, "Minmax and minmax
regret versions of combinatorial optimization problems: A
survey," *Eur. J. Oper. Res.*, vol. 197, no. 2, pp. 427–438, 2009.

[28] T. S. Lee, "Image representation using 2D Gabor wavelets," *IEEE
Trans. Pattern Anal. Mach. Intell.*, vol. 18, no. 10, pp. 959–971,
Oct. 1996.

[29] B. Manjunath and W. Ma, "Texture features for browsing and
retrieval of image data," *IEEE Trans. Pattern Anal. Mach. Intell.*,
vol. 18, no. 8, pp. 837–842, Aug. 1996.

[30] M. Emambakhsh, H. Ebrahimnezhad, and M. Sedaaghi,
"Integrated region-based segmentation using color components
and texture features with prior shape knowledge," *Int. J. Appl.
Math. Comput. Sci.*, vol. 20, no. 4, pp. 711–726, 2010.

[31] K. Deb, A. Pratap, S. Agarwal, and T. Meyarivan, "A fast and elit-
ist multiobjective genetic algorithm: NSGA-II," *IEEE Trans. Evol.
Comput.*, vol. 6, no. 2, pp. 182–197, Apr. 2002.

[32] N. Srinivas and K. Deb, "Muiltiobjective optimization using non-
dominated sorting in genetic algorithms," *Evol. Comput.*, vol. 2,
no. 3, pp. 221–248, 1994.

[33] 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.

[34] 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 Under-
standing*, vol. 117, no. 2, pp. 158–169, 2013.

[35] H. Li, D. Huang, P. Lemaire, J.-M. Morvan, and L. Chen,
"Expression robust 3D face recognition via mesh-based histo-
grams of multiple order surface differential quantities," in *Proc.
18th IEEE Int. Conf. Image Process.*, 2011, pp. 3053–3056.

[36] F. Hajati, A. Raie, and Y. Gao, "2.5D face recognition using patch
geodesic moments," *Pattern Recog.*, vol. 45, no. 3, pp. 969–982,
2012.

[37] 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, pp. 146–179, 2013.

[38] R. Raguram, C. Wu, J.-M. Frahm, and S. Lazebnik, "Modeling and
recognition of landmark image collections using iconic scene
graphs," *Int. J. Comput. Vis.*, vol. 95, no. 3, pp. 213–239, 2011.

[39] N. Alyüz, B. Gökberk, and L. Akarun, "A 3D face recognition sys-
tem for expression and occlusion invariance," in *Proc. 2nd IEEE
Int. Conf. Biometrics: Theory, Appl. Syst.*, 2008, pp. 1–7.

Mehryar Emambakhsh received the PhD
degree in electronic and electrical engineering
from the University of Bath in 2015, researching
the potential of the 3D shape of the nose for bio-
metric authentication. He has been a post-
doctoral research associate (PDRA) in big data
science at Aston University and is currently a
PDRA in RADAR and video data fusion for auton-
omous vehicles at Heriot-Watt University. His
research interests are in 3D object recognition
and data mining.

Adrian Evans graduated from Loughborough
University with a BEng in electronics and
manufacturing engineering in 1990 and recieved
the PhD degree in medical image processing
from the University of Southampton in 1994. After
a number of years as an academic at Massey
University, New Zealand, he joined the Depart-
ment of Electronic and Electrical Engineering at
the University of Bath in 1997, where he is cur-
rently the head of department. His research inter-
ests are in image and video processing and
analysis and he has more than 70 journal and refereed conference
papers in this area. Current research projects include mathematical mor-
phology, remote sensing, colour image processing, and biometrics.

▷ **For more information on this or any other computing topic,
please visit our Digital Library at** www.computer.org/publications/dlib.