Surface geodesic pattern for 3D deformable texture matching

Farshid Hajati a, b,*, Ali Cheraghian c, Soheila Gheisari d, Yongsheng Gao b, Ajmal S. Mian e

a Department of Electrical Engineering, Tafresh University, Tafresh 39518 79611, Iran
b School of Engineering, Griffith University, QLD 4111, Australia
c Department of Electrical Engineering, Amirkabir University of Technology, Tehran, Iran
d Centre for Quantum Computation and Intelligent Systems, Faculty of Engineering and Information Technology, University of Technology Sydney, NSW 2007, Australia
e School of Computer Science and Software Engineering, The University of Western Australia, WA 6009, Australia

A R T I C L E   I N F O

Article history:
Received 28 February 2016
Received in revised form 31 July 2016
Accepted 18 August 2016
Available online 21 August 2016

Keywords:
3D non-rigid object recognition
Geodesic derivatives
Deformable surface matching

A B S T R A C T

This paper presents a Surface Geodesic Pattern (SGP) representation for matching textured 3D deformable surfaces. SGP encodes the local variations of the surface texture derivatives to extract local information from distinctive textural relationships contained in a geodesic neighborhood. Thus, SGP derives its strength from the fusion of surface texture and shape information at the data level in a way that is invariant to non-rigid deformations. We also propose Gabor Topography Wavelet (GTW) for direct feature extraction from the range data. Both features are combined using a multi-view sparse representation to achieve higher discrimination capability while matching non-rigid 3D surfaces. The performance of the proposed method is evaluated extensively on the Bosphorus face database, the FRGC v2 face database, and the PolyU contact-free hand database and the results are compared to state-of-the-art methods. Experimental results show the effectiveness and superiority of the proposed method in recognizing objects under non-rigid surface deformations.

© 2016 Elsevier Ltd. All rights reserved.

1. Introduction

Deformable 3D surface matching has many applications in computer vision and computer graphics including non-rigid object registration [1], surface classification [2], statistical shape analysis [3], expression invariant face recognition, and shape blending [4]. Deformation is a change in the shape of a surface due to an applied force. Deformable surface matching is a challenging problem because non-rigid deformations change the intrinsic geometry as well as the appearance of the surface texture. Algorithms that rely only on the appearance of surface texture are unable to match deformable surfaces. The use of 3D shape information can improve the performance of non-rigid surface matching. This is especially useful for the non-elastic deformations, where the geodesic distances between points on the surface remain invariant.

Nowadays, the use of 3D information in object recognition algorithms has increased due to the improvement in 3D scanners. 3D images used in surface matching can be divided into two categories. The first category consists of single-view range images captured by commercial 3D scanners. A single-view range image has the location of the surface points in 3D space defined as a single range value (z-direction) on a uniform x–y grid [5]. It can also be represented as an xyz point-cloud of the partial surface. The second category consists of 3D models obtained after merging several point-clouds from different viewing angles. Static 3D sensors can practically capture only a partial view of a surface in a single scan.

In non-rigid 3D surface matching, deformation invariant representation is of particular importance as rigid surface matching systems are sensitive to surface deformations. Different approaches have been proposed for 3D deformation-invariant surface matching. However, these methods rely solely on shape information without considering the distinctive textural pattern on the 3D shape (see Table 1). How to effectively develop multimodal techniques for recognizing non-rigid 3D surfaces remains an open research problem. In this paper, we present a novel Surface Geodesic Pattern (SGP) representation that uses both texture and shape data for deformable 3D surface matching. The proposed geodesic path regulated texture coding can fuse texture and shape information at the data level to tackle the problems of surface deformations. Instead of using the complete 3D images, SGP...
A multi-view feature-level sparse fusion method is proposed to categorize the existing 3D surface matching methods. The algorithm adopted a distance obtained from the hypergraph to estimate combining different views with labeling information. The main contributions of the proposed method are:

Rigid methods

- Oceguera et al. [6]
- Passalis et al. [7]
- Efraty et al. [8]
- Smeets et al. [9]
- Faltamier et al. [10]
- Frome et al. [11]
- Huang et al. [12]
- Li et al. [13]
- Yu et al. [15]

Non-rigid methods

- Elad et al. [19]
- Bronstein et al. [20]
- Berretti et al. [21]
- Bronstein et al. [22]

2. Related work

Existing 3D surface recognition techniques can be broadly classified into two categories according to the type of data they use: purely shape-based and multimodal methods. Shape-based methods extract discriminative features directly from surface shape [11,23]. Hilaga et al. [24] explored inter-geodesic distances between points on a surface to match non-rigid surfaces. They defined a scalar function on the surface namely distribution function which was computed by integrating the geodesic distances from the given point to the rest of the points on the surface. The skeletal structure of the scalar function was constructed and analysed via Multidimensional Reeb Graphs (MRGs) [25]. The distance between MRGs was used as the dissimilarity metric. Sundar et al. [26] compared 3D surfaces by encoding the geometric and topological information in a skeletal graph and used graph matching techniques to match the skeletons. Osada et al. [27] proposed a method for computing shape signatures for 3D polygonal models. They represented the signature of an object’s surface as a shape distribution sampled from a shape function measuring global geometric properties of the object’s surface. Frome et al. [11] proposed two regional shape descriptors of 3D shape contexts and harmonic shape contexts, and evaluated their performance on recognizing vehicles in range images of scenes using a database of 56 cars. Yu et al. [55] proposed the Sparse Patch Alignment Framework (SPAF) algorithm to learn manifolds. They used an optimization strategy to construct local patches of the manifolds based on the patch alignment framework and adopted a sparse representation method to select a few neighbors of each data point spanning a low-dimensional affine subspace around that point. Finally, the complete alignment strategy was used to construct the manifold.

Multimodal methods use both texture and shape information of the 3D surface for matching and hence have higher performance compared to methods which use shape data alone [28]. Mian et al. [14] used a fully automatic multimodal face recognition algorithm to perform hybrid matching. They used a 3D Spherical Face Representation (SFR) and the Scale-Invariant Feature Transform (SIFT) [29] descriptor to form a rejection classifier that was combined with a region-based ICP algorithm. The results of all matching engines were fused at the metric level to achieve higher accuracy. They reported 99.7% and 98.3% verification rates at a 0.001 FAR (False Acceptance Rate) and 99% and 95.4% identification rates for neutral versus neutral and non-neutral versus neutral expression experiments respectively, on the FRGC v2 database [30]. Hajati et al. [15] proposed Patch Geodesic Distance (PGD) for expression and pose invariant 3D face recognition. In PGD, 3D face images were partitioned into equal-sized square patches in a non-overlapping manner. Local geodesic paths within patches and their global geodesic paths were combined to encode the shape adjusted textures into feature descriptors. They achieved rank-1 recognition rates of 84.8% and 69.1% on the BU-3DFE [31] and the Bosphorus [32] databases, respectively. Zhang et al. [16] proposed a multimodal method for palmprint verification. They extracted surface curvature features from the shape, while Gabor features were used for texture representation. An Equal Error Rate (EER) of 0.0022% on a database of 108 subjects was achieved. Li et al. [17] extracted principal line features and palm shape features to align the palmprints. They achieved an EER of 0.025% on the PolyU 3D palmprint database which contains 8000 samples. Kanhangad et al. [18] introduced two representations for contactless hand verification: finger surface curvature and unit normal vector. They also introduced a new feature representation called SurfaceCode for 3D palmprint which resulted in a significant reduction in the template size. They achieved an Equal Error Rate (EER) of 0.56% on the PolyU contact-free hand database [18].

Table 1

<table>
<thead>
<tr>
<th>Shape-based methods</th>
<th>Multimodal methods</th>
</tr>
</thead>
<tbody>
<tr>
<td>Oceguera et al. [6]</td>
<td>Mian et al. [14]</td>
</tr>
<tr>
<td>Passalis et al. [7]</td>
<td></td>
</tr>
<tr>
<td>Efraty et al. [8]</td>
<td></td>
</tr>
<tr>
<td>Smeets et al. [9]</td>
<td>Hajati et al. [15]</td>
</tr>
<tr>
<td>Faltamier et al. [10]</td>
<td></td>
</tr>
<tr>
<td>Zhang et al. [16]</td>
<td></td>
</tr>
<tr>
<td>Frome et al. [11]</td>
<td>Li et al. [17]</td>
</tr>
<tr>
<td>Huang et al. [12]</td>
<td>Kanhangad et al. [18]</td>
</tr>
<tr>
<td>Li et al. [13]</td>
<td></td>
</tr>
<tr>
<td>Yu et al. [15]</td>
<td></td>
</tr>
<tr>
<td>Elad et al. [19]</td>
<td></td>
</tr>
<tr>
<td>Bronstein et al. [20]</td>
<td></td>
</tr>
<tr>
<td>Berretti et al. [21]</td>
<td></td>
</tr>
<tr>
<td>Bronstein et al. [22]</td>
<td>N/A</td>
</tr>
</tbody>
</table>
From the deformability viewpoint, existing 3D surface matching algorithms can be divided into rigid and non-rigid algorithms. Rigid algorithms utilize rigid transformations for feature extraction [9,11,13,23,33–36]. Kazhdan [23] used rotation-invariant spherical harmonics as shape descriptors in 3D object recognition. Vranic [33] proposed depth buffer-based, silhouette based, and ray-based methods for similarity measurement between 3D objects. Laga et al. [34] applied spherical parameterization and geometry images for 3D object matching which was invariant to rotation and scaling. They compensated the rotations in 3D images using the principal axis of the object. Allen et al. [35] proposed a method for fitting high-resolution template meshes to detailed human body range scans with sparse 3D markers. They formulated an optimization problem in which the degrees of freedom were an affine transformation at each template vertex. The objective function was a weighted combination of three measures: proximity of transformed vertices to the range data, similarity between neighboring transformations, and proximity of sparse markers at corresponding locations on the template and target surface. Falltner et al. [10] introduced a 3D face recognition method by fusing scores from 28 independently matched small regions. They reported a rank-1 recognition rate of 97.2% and verification rate of 93.2% at 0.1% FAR on the FRGC v2 database [30]. Oceguera et al. [6] presented a Markov Random Field (MRF) model to analyse 3D meshes and determine both discriminative and non-discriminative vertices. Efraty et al. [8] proposed a silhouetted face profile framework for 3D face recognition. 3D data of subjects were used to create profiles under different rotations and train a classifier using the extracted features.

Smeets et al. [9] proposed meshSIFT features for expression-invariant face recognition. In meshSIFT, salient points on the 3D facial surface are detected as mean curvature extrema in scale space and orientations are assigned to each. Then, the neighborhood of each salient point is described as a feature vector consisting of concatenated histograms of shape indices and slant angles. Finally, the feature vectors of two 3D facial surfaces are matched by comparing the angles in the feature space. They achieved rank-1 recognition rates of 93.7% and 89.6% for the Bosporus [32] and the FRGC v2 [30] databases, respectively. They also demonstrated that symmetric meshSIFT descriptors had acceptable performance in partial 3D data matching. Passalis et al. [7] proposed a pose-invariant 3D face recognition approach using facial symmetry and automatic landmark detection. They fitted an annotated face model to the input face scan to overcome the challenges of missing data. An average rank-1 recognition rate of 83.7% was reported on datasets of the University of Notre Dame and the University of Houston containing pose variations.

Huang et al. [37] proposed Multi-Scale Local Binary Pattern (MS-LBP) for 3D facial surface representation. They achieved a rank-1 recognition rate of 96.1% on the FRGC v2 database [30]. In [12], a geometric representation for 3D range images based on Multi-Scale Extended Local Binary Patterns (MS-ELBP) achieved a rank-1 recognition rate of 97.2% and a verification rate of 98.4% at 0.001 FAR respectively on the FRGC v2 database [30]. Li et al. [13] proposed a mesh-based approach for 3D face recognition using local shape descriptor and a SIFT-like matching process. They employed both the maximum and the minimum curvatures estimated in the 3D Gaussian scale space to detect salient points. To comprehensively characterize 3D facial surfaces and their variations, they used weighted statistical distributions of multiple order surface differential quantities, including histogram of mesh gradient (HoG), histogram of shape index (HoS), and histogram of gradient of shape index (HoGS) within a local neighborhood of each salient point.

The above rigid object recognition algorithms, however, are inherently sensitive to surface deformations. Since most of the objects observed in the real world are non-rigid, non-rigid shape representation has gained more attention in computer vision [19–21,38]. Elad and Kimmel [19] presented a method to construct a bending invariant signature for surfaces based on the geodesic distances between points. They measured the inter-geodesic distances between uniformly distributed points on the surface. Then, a Multi-Dimensional Scaling (MDS) technique was applied to extract coordinates in a finite dimensional Euclidean space in which geodesic distances are replaced by Euclidean ones for matching. Bronstein et al. [20] proposed an expression-invariant representation of faces called Canonical Image Representation which modelled deformations resulted from facial expressions. The canonical images were built by calculating the geodesic distances between the points of the facial surface. Their experimental results demonstrated that a smaller embedding error leads to better recognition.

Berretti et al. [21] encoded the geometric information of the 3D face into a compact representation in the form of a graph where the nodes were represented with equal-width iso-geodesic stripes. They achieved recognition rates of 99% and 94.1% on the SHREC’08 and the FRGC v2 databases, respectively. Litman et al. [39] proposed a diffusion-geometric framework for volumetric stable component detection and description in deformable shapes. They evaluated their method on the SHREC’11 [40] database and SCAPE [41] human body database and demonstrated high discriminability of the proposed volumetric features. Bronstein et al. [22] proposed a method for computing the similarity of non-rigid shapes based on the distributions of diffusion distances defined with explicit invariance properties.

3. Surface geodesic pattern (SGP)

In this section, we propose a new Surface Geodesic Pattern (SGP) representation that directly encodes deformation invariant texture patterns of a surface in the 3D space. The important notations used in this section are summarized in Table 2. For a textured surface $I(x,y)$ in 3D space, $Z_0(x_0,y_0)$ is the center point of a set of geodesic rings, and $Z_{0q}(x,y), r = 1, ..., R; q = 0, ..., N - 1$, is the $q$th neighboring point around $Z_0(x_0,y_0)$ on the $r$th ring (see Fig. 1). 3D texture information is carried by the changes in magnitude of $k(Z)$ among points $Z_0$ and $Z_{0q}$, which are spatially distributed on the geodesic rings of the surface. We represent the texture variation of the surface using a novel Geodesic Inner-Derivative (GID) function. GID can be computed in different orders encoding the relative changes of texture variations within the neighborhood. GID measures concavity and convexity of the texture variations using the product of the texture derivatives along a spiral path within the neighborhood.

The first-order Geodesic Inner-Derivative (GID) at $Z = Z_0$, $I_{1r, q}(Z_0)$, is defined as

<table>
<thead>
<tr>
<th>Table 2</th>
<th>SGP’s important notations and their descriptions.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Notations</strong></td>
<td><strong>Descriptions</strong></td>
</tr>
<tr>
<td>$i_{r, q}(Z)$</td>
<td>$r$-th order GID at point $Z$ using the $r$th geodesic ring and $q$th neighboring point</td>
</tr>
<tr>
<td>$r$</td>
<td>$r$-th geodesic ring</td>
</tr>
<tr>
<td>$q$</td>
<td>$q$th neighboring point</td>
</tr>
<tr>
<td>$W_p$</td>
<td>Patch size</td>
</tr>
<tr>
<td>$a$ and $b$</td>
<td>Patch labels</td>
</tr>
<tr>
<td>$x$ and $y$</td>
<td>Patch coordinates</td>
</tr>
<tr>
<td>$F_i$</td>
<td>Spiral derivative function</td>
</tr>
<tr>
<td>$Z_0(x_0, y_0)$</td>
<td>Geodesic ring’s center point</td>
</tr>
<tr>
<td>$SGP(Z_0)$</td>
<td>Surface Geodesic Pattern at point $Z_0$</td>
</tr>
</tbody>
</table>
\[ l_{r,q} (Z_0) = \left( l_{r,q} (Z_0) - l_{r-1,q-1 \mod (N_0)} \right) \]
\[ \left( l_{r+1,q+1 \mod (N_0)} - l_{r+q \mod (N_0)} \right); \]
\[ r = 1, \ldots, R & q = 0, \ldots, N - 1 \]

where \( l_{r,q} (Z_0) \) denotes the texture value of the neighboring point \( Z_{r,q} \) around the center point \( Z_0 \), and \( mod \) is the remainder operator.

The second-order Geodesic Inner-Derivative (GID) at \( Z = Z_0 \), \( l'_{r,q} (Z_0) \) is defined as
\[ l'_{r,q} (Z_0) = \left( l'_{r,q} (Z_0) - l'_{r-1,q-1 \mod (N_0)} \right) \]
\[ \left( l'_{r+1,q+1 \mod (N_0)} - l'_{r+q \mod (N_0)} \right); \]
\[ r = 1, \ldots, R & q = 0, \ldots, N - 1 \]

where \( l'_{r,q} (Z_0) \) is the first-order Geodesic Inner-Derivative (GID) computed using Eq. (1).

In general, the \( n \)-th order Geodesic Inner-Derivative (GID) at \( Z = Z_0 \) can be defined as
\[ l^{(n)}_{r,q} (Z_0) = l^{(n)}_{r,q} (Z_0) - l^{(n-1)}_{r,q} \]
\[ r = 1, \ldots, R & q = 0, \ldots, N - 1 \]

where \( l^{(n)}_{r,q} (Z_0) \) is the \( n \)-th-order Geodesic Inner-Derivative (GID) at \( Z = Z_0 \), and \( n \) is the order of the derivative.

After computing the \( n \)-th order Geodesic Inner-Derivative (GID) at \( Z = Z_0 \), we encode the computed inner-derivatives of the point \( Z = Z_0 \) into a decimal value called Surface Geodesic Pattern (SGP) using the unit step function. Essentially, we are only encoding the direction of inner-derivative which is more robust to illumination changes compared to the derivative value itself. We represent a surface texture value at \( Z = Z_0 \) with an \( n \)-th order Surface Geodesic Pattern (SGP) as
\[ SGP^{(n)}_r (Z_0) \]
\[ = \sum_{k=0}^{N-1} 2^k u^{(n)}_{r,k} (Z_0) \]
\[ r = 1, \ldots, R \]

where \( u^{(n)}_{r,k} (Z_0) \) is the \( n \)-th-order Geodesic Inner-Derivative (GID) at \( Z = Z_0 \) computed from the neighboring texture values located on the geodesic rings, and \( u() \) denotes the unit step function. The result of the SGP operator for an example 3D face is illustrated in Fig. 2. It can be observed from Fig. 2 that as the operator order increases, finer details are captured on the textured surface.

The SGP operator is used for surface texture representation. The proposed method applies SGP operator on each point to extract discriminative features from its neighborhood. We model the distribution of Surface Geodesic Patterns (SGPs) by local histograms, due to their inherent robustness against variations in pose and illumination. For this purpose, we partition the SGP images into non-overlapping square patches and compute the histogram for each patch, independently. For an \( M \times M \) SGP image and a patch size of \( W_p \), the number of patches is \((M/W_p)^2\). Here, we label the patch points with \( a \) and \( b \) indexes defined as
\[ a = \lfloor x/W_p \rfloor + 1; \quad 0 \leq x < M \]
\[ b = \lfloor y/W_p \rfloor + 1; \quad 0 \leq y < M \]

where \( a \) and \( b \) are integers ranging from 1 to \( M/W_p \), and the symbol \( \lfloor \cdot \rfloor \) denotes the floor function.

Using \( a \) and \( b \) indexes, the \( ab \)-th patch of the Surface Geodesic Pattern \( SGP^{(n)}_{r,a,b}(Z) \), is represented as
\[ SGP^{(n)}_{r,a,b}(Z(x,y)) = SGP^{(n)}_r(Z(x,y) + x_{ab}, W_p(b - 1) + y_{ab})) \]

where \( x_{ab} \) and \( y_{ab} \) are the coordinates in the \( ab \)-th patch.

We use the spatial histogram for each patch and concatenate the histograms to achieve a feature vector as
\[ HSGP^{(n)}_r(Z) = \left\{ H[SGP^{(n)}_{r,a,b}(Z)] | a = 1, \ldots, M/W_p, b = 1, \ldots, M/W_p \right\} \]

where \( H[\cdot] \) denotes the spatial histogram operator.

Since the proposed SGP is a geodesic path regulated texture pattern, it is intrinsically robust to surface deformations. This is also evidenced by the consistently greater discriminative capability of SGP compared to the original intensity data under various deformations. Fig. 3 compares the SGP and intensity differences of an expressive face against faces with neutral expression in the Bosphorus database [32]. As can be seen, the difference between the average inter-class distance and the intra-class distance of SGP is nearly doubled to that of the original intensity data, showing a very strong discriminability of SGP in recognizing 3D objects under surface deformations.

4. Gabor topography wavelet and sparse fusion

In addition to SGP features, we also extract pure shape features directly from the topography maps of the surface. These feature are then fused with the SGP features using a sparse multi-view representation. The distinctive characteristic of a topography map is that the shape of the surface is shown by level curves (contour lines). Level curves are lines that join points of the equal height on the surface above or below a reference point (usually the highest point of the surface). The important notations used in this section are summarized in Table 3. Given a surface \( Z(x,y) \), the level curve with the height index \( h \), \( L_h \), is defined as
\[ L_h = \{ (x,y) | z_{max} - h w_z < Z(x,y) < z_{max} - (h-1) w_z \}; \]
\[ h = 1, 2, \ldots, z_{max}/w_z + 1 \]

where \( z_{max} \) is the maximum height (reference) point in the surface, \( w_z \) is the step size of the level curve, \( h \) is the height index of the level curve, and \( \lfloor \cdot \rfloor \) denotes the floor function.

Using the level curves \( L_h \), \( h = 1, 2, \ldots, z_{max}/w_z + 1 \), the topography map \( TM(Z(x,y)) \) can be defined as
After extracting the topography map of the given surface, Gabor wavelets [42] are applied to create the Gabor topography map, $G_{W}\left(Z(x, y)\right)$, by convolving the topography map with the Gabor kernel, $\varphi_{r, \theta}(Z(x, y))$, as

$$
G_{W}\left(Z(x, y)\right) = TM\left(Z(x, y)\right) \ast \varphi_{r, \theta}(Z(x, y))
$$

where $\omega$ is the radial center frequency of a sinusoidal wave plane, $\theta$ is the anti-clockwise orientation of the Gaussian envelope of the

$$
\varphi_{r, \theta}(Z(x, y)) = \frac{e^{-\frac{\omega^2}{4\sigma^2}}}{4\pi^{\frac{3}{2}}} e^{-\frac{(x - x_0)^2 + (y - y_0)^2 + (z - z_0)^2}{2\sigma^2}} e^{j(\omega x \cos \theta + \omega y \sin \theta)}
$$
GTW’s notations and their definitions.

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>( t^h )</td>
<td>Level curve with the height index ( h )</td>
</tr>
<tr>
<td>( h )</td>
<td>Height index</td>
</tr>
<tr>
<td>( w_0 )</td>
<td>Step size of the level curve</td>
</tr>
<tr>
<td>( z_{\text{max}} )</td>
<td>Maximum height in the surface</td>
</tr>
<tr>
<td>( \omega )</td>
<td>Radial center frequency of a sinusoidal wave plane</td>
</tr>
<tr>
<td>( \theta )</td>
<td>Anti-clockwise orientation of the sinusoidal Gaussian envelope</td>
</tr>
<tr>
<td>( \sigma )</td>
<td>Spatial width of the Gaussian envelope along ( x ) and ( y ) axes</td>
</tr>
<tr>
<td>( \lambda )</td>
<td>Scaling factor</td>
</tr>
<tr>
<td>( m )</td>
<td>Number of frequencies</td>
</tr>
<tr>
<td>( n )</td>
<td>Number of orientation</td>
</tr>
<tr>
<td>( D )</td>
<td>Dictionary matrix</td>
</tr>
<tr>
<td>( d )</td>
<td>Vector element of the dictionary matrix</td>
</tr>
<tr>
<td>( \gamma )</td>
<td>Balancing factor</td>
</tr>
<tr>
<td>( V_w )</td>
<td>Sparse representation vector</td>
</tr>
<tr>
<td>( \langle \cdot, \cdot \rangle )</td>
<td>Inner product operator</td>
</tr>
<tr>
<td>( v )</td>
<td>Gabor kernel</td>
</tr>
<tr>
<td>( AAD )</td>
<td>Average Absolute Deviation</td>
</tr>
<tr>
<td>( GTW )</td>
<td>Gabor Topography Wavelet</td>
</tr>
</tbody>
</table>

sinusoid, and \( \sigma \) is the spatial width of the Gaussian envelope along the \( x \) and \( y \) axes. Fig. 4 illustrates an example Gabor topography map.

The Gabor topography maps are obtained by a filter bank of Gabor functions consisting of a number of scales and orientations. To select the appropriate discrete frequencies that define the scales, we use the following exponential sampling [43]:

\[
\omega^k = \frac{\omega}{\lambda^k}, \quad k = 0, \ldots, m - 1
\]

where \( \omega^k \) is the \( k \)-th frequency, \( \lambda > 1 \) is the frequency scaling factor, and \( m \) is the number of frequencies.

Rotation angle, \( \theta \), is uniformly spaced as

\[
\theta^l = \frac{2\pi l}{n}, \quad l = 0, \ldots, n - 1
\]

where \( n \) is the number of orientations.

The Gabor topography maps are partitioned into equal-sized \( W_{\text{GTW}} \times W_{\text{GTW}} \) patches as

\[
GTW_{\omega,\theta}(Z(x_{ab}, y_{ab})) = GTW_{\omega,\theta}(W_{\text{GTW}}(a - 1) + x_{ab}, W_{\text{GTW}}(b - 1) + y_{ab})
\]

where \( x_{ab} \) and \( y_{ab} \) are the coordinates in the \( ab \)-th patch, and \( W_{\text{GTW}} \) is the patch size.

After partitioning, we calculate the Average Absolute Deviation (AAD) from the mean of the patch’s gray levels for all patches. For the \( ab \)-th patch of the Gabor topography map, the Average Absolute Deviation (AAD) from the mean of the gray levels is defined as

\[
AAD_{\omega,\theta}(Z) = \frac{1}{W_{\text{GTW}}} \sum_{x_{ab}, y_{ab}=0}^{W_{\text{GTW}}} |GTW_{\omega,\theta}(Z(x_{ab}, y_{ab})) - GTW_{\omega,\theta}(Z)|
\]

(17)

where \( GTW_{\omega,\theta}(Z) \) is the \( ab \)-th patch of the Gabor topography map defined in (16), \( W_{\text{GTW}} \) is the number of pixels in the \( ab \)-th patch and \( GTW_{\omega,\theta}(Z) \) is the mean of the \( ab \)-th patch’s gray levels.

By concatenating the AADs, we build a feature vector called Gabor Topography Wavelet (GTW) for the given subject as

\[
\begin{aligned}
\mathbf{d}_{\omega,\theta} &= \{ \text{AAD}_{\omega,\theta}(Z) \}_{a=1}^{M/W_{\text{GTW}}}, \quad b = 1, \ldots, M/W_{\text{GTW}}, \quad k = 0, \ldots, m - 1, \quad l = 0, \ldots, n - 1
\end{aligned}
\]

(18)

where \( W_{\text{GTW}} \) is the patch size. Fig. 5 illustrates the Gabor topography maps and the Average Absolute Deviations for a sample topography map.

Once the GTW feature vectors are computed, we fuse them with the SGP feature vectors using the sparse multi-view representation method. For this purpose, we define a dictionary matrix which models input 3D images using a training set. Suppose there are \( S \) sample images per subject in the training set. To make the columns of the dictionary matrix, we concatenate the SGP and the GTW vectors of the training samples using a balancing factor. The column of the dictionary matrix for the \( s \)-th sample of the \( t \)-th subject, \( \mathbf{d}_{s,t} \), is defined as

\[
\mathbf{d}_{s,t} = [HSGP_{\omega,\theta}(Z) ; GTW_{\omega,\theta}(Z)]^T ; s = 1, \ldots, S \text{ & } t = 1, \ldots, T
\]

(19)

where \( \gamma \) is a scalar balancing factor which is determined experimentally.

By repeating the above process for all subjects in the training set, we obtain the desired dictionary matrix for the sparse fusion. In this way, the dictionary matrix contains \( S \times T \) columns, called atoms, as

\[
\mathbf{D} = [\mathbf{d}_{1,1}, \mathbf{d}_{1,2}, \ldots, \mathbf{d}_{1,T}, \ldots, \mathbf{d}_{S,T}]
\]

(20)

Once the dictionary matrix \( \mathbf{D} \) is obtained, we compute the sparse representation of each subject in the gallery and probe sets as a sparse linear combination of the dictionary atoms as

\[
\mathbf{V} = \mathbf{DV}_{\text{sp}}
\]

(21)

where \( \mathbf{V} \) is the concatenation of SGP and GTW feature vectors of the subject, and \( \mathbf{V}_{\text{sp}} \) is the sparse representation vector. Sparsity implies that only a few components of \( \mathbf{V}_{\text{sp}} \) are non-zero and the rest are zero. This suggests that \( \mathbf{V} \) can be decomposed as a linear combination of only a few column vectors in \( \mathbf{D} \). Such vectors are called the basis of \( \mathbf{V} = \mathbf{D} \) which is over-complete (there are fewer rows than columns).

Given a feature vector \( \mathbf{V} \) and the dictionary \( \mathbf{D} \), the sparse decomposition problem aims to find the representation of \( \mathbf{V} \) as a linear combination of a few dictionary elements (columns). The sparse decomposition equation is defined as

\[
\min_{\mathbf{V}_{\text{sp}}} \| \mathbf{V}_{\text{sp}} \|_0 \text{ s. t. } \mathbf{V} = \mathbf{DV}_{\text{sp}}
\]

(22)

where \( \| \cdot \|_0 \) is the pseudo-norm (\( L_0 \) norm) which counts the number of non-zero components of \( \mathbf{V}_{\text{sp}} \). Because this problem is NP-Hard [61], we use a convex relaxation of the problem by taking the \( L_1 \) norm instead of \( L_0 \) norm as

\[
\| \mathbf{V}_{\text{sp}} \|_1 = \sum_{t}^{T} \| \mathbf{V}_{t} \|_1
\]

(23)

One solution of (21) is the Basis Matching Pursuit (BMP) algorithm [44]. Assuming \( \mathbf{d}_0 \in \mathbf{D} \), the vector \( \mathbf{V} \) can be decomposed into...
\[ V = \langle V, d_0 \rangle d_0 + R_V \]  

where the symbol \( \langle \cdot, \cdot \rangle \) denotes the inner product operator and \( R_V \) is the residual vector over approximating \( V \) in the direction of \( d_0 \).

Since \( d_0 \) is orthogonal to \( R_V \), we rewrite Eq. (22) as

\[ \| V \|^2 = \| \langle V, d_0 \rangle \|^2 + \| R_V \|^2 \]  

In order to minimize \( \| R_V \| \), we choose \( d_0 \) such that \( \| \langle V, d_0 \rangle \| \) is the maximum. The BMP algorithm decomposes the residue \( R_V \) iteratively by projecting \( R_V \) on a column vector of \( D \) with the best match of \( R_V \). The procedure is repeated until the residue becomes smaller than a predefined threshold.

5. Experimental results

To evaluate the effectiveness of the proposed approach, experiments are conducted on the Bosphorus [32] and FRGC v2 [30] 3D face databases for face recognition under facial expressions. The PolyU contact-free hand database [18] is used for 3D hand identification and verification. Since the palmprints have been captured in a contact free fashion, there are some deformations of the hand in the database.

The Bosphorus database [32] contains face texture (RGB) and range (depth) data for 105 persons, including 4666 pairs of faces with different expressions, poses, and occlusions. The range images have a resolution of 0.3 mm, 0.3 mm, and 0.4 mm in \( x, y, \) and \( z \) (depth) dimensions, respectively, and the color texture maps have a resolution of 1600 \( \times \) 1200 pixels. To allow for a more direct comparison between the faces, they are normalized and aligned using 24 landmarks provided in the database. Spikes in the range maps are removed and holes are filled as in [14]. Then, faces are resampled on a uniform square grid at 1 mm resolution and cropped to the size of 120 \( \times \) 120 in the way that the distance of the nose tip is 60 mm from the sides and 70 and 50 mm from the top and the bottom, respectively.

The FRGC v2 database [30] contains 4950 face texture and range images. The database is divided into three sets, namely Spring2003 (943 scans of 277 individuals), Fall2003 and Spring2004 (4007 scans of 466 individuals). The range and texture images are both 480 \( \times \) 480 with one-to-one pixel correspondence. Similar to the Bosphorus database, the FRGC v2 face images are normalized and cropped to the size of 120 \( \times \) 120. The PolyU contact-free hand database [18] includes 3540 hand data from 177 subjects aged 18–50 years, with multiple ethnic backgrounds. Each subject has contributed 5 pairs of hand range and texture images acquired simultaneously in the first session, followed by another 5 in the second session. During the image acquisition no constraints have been employed to confine the position and deformation of the hand. The resolution of the images (both range and texture) is 640 \( \times \) 480 pixels. All hand scans have been cropped to 120 \( \times \) 120 regions of interest.

5.1. Parameters

In this section, we examine the parameters involved in the proposed methods and empirically show that SGP and GTW are robust to a wide range of parameter values. To determine the SGP’s parameters, the radius \( r \), the patch size \( W_p \), and the order, an experimental investigation on the SGP’s recognition accuracy is conducted with different values of the SGP’s order \( \in \{1, 2, 3\} \), the SGP’s radius \( \in \{2 \text{ to } 10 \text{ mm}\} \) and the SGP’s patch size \( \in \{3 \text{ to } 120\} \) using a training set selected from the Bosphorus database [32]. To build the training set, one frontal face per subject is considered as the gallery and one of the remaining faces (rotated/expansive) is selected randomly for each subject as the probe. The results are displayed in Fig. 6. The recognition accuracy increases rapidly when SGP’s patch size decreases from 120 to 15. It remains flat at high level till 6, then drops gracefully. The accuracy decreases when the SGP’s order increases from 1 to 3. The best performance is achieved when the SGP’s radius, the SGP’s patch size, and the SGP’s order are 6, 10, and 1, respectively.

Fig. 7 illustrates the sensitivity of the GTW to the step size \( w_z \) and the patch size \( W_{GTW} \). The recognition rate increases as \( w_z \) decreases from 30 to 4 and then remains flat till 1. On the other hand, the recognition rate increases when \( W_{GTW} \) reduces from 120 to 10 and then decreases gradually when \( W_{GTW} \) reduces from 10 to 3. In the remaining experiments, we set \( w_z=1.3, W_{GTW}=10, r=6, W_p=10, \) and order=2.

5.2. Face identification under varying pose and expressions in the bosphorus database

The Bosphorus database [32] has a rich repertoire of facial expressions i.e. up to 35 expressions per subject. To evaluate the performance of the proposed method under expression and pose variations, our algorithm is tested against all faces in the Bosphorus database [32] excluding the profile and the occluded ones. Up to 45° pose variations are included in our test data. One neutral expression frontal face per subject is used as the gallery, while the rest are used as probe faces. The test data contains faces with emotions (happy, fear, disgust, anger, sadness, surprise), lower and
upper face actions, combined face actions, and pose variations (Left 45°, Right 10°, Right 20°, Right 30°, Right 45°, Right Down, Right Up, Slightly Up, Strong Up, Slightly Down, Strong Down) (see Fig. 8).

The rank-1 recognition rate of the proposed and the benchmark methods on the Bosphorus database [32] are listed in Table 4. As can be seen, the rank-1 recognition rate of the proposed method with 105 gallery identities and 3696 probes, including scans with expression and rotation changes, is 95.2%. However, the Baseline ICP [45] and the Baseline PCA [45] methods have the recognition rates of 72.4% and 70.6% respectively, using only 47 gallery identities and 1508 expressive probes. Alyuz et al. [46], Dibeklioglu et al. [47], and Dibeklioglu et al. [47] reported the rank-1 recognition rates of 95.3%, 89.2%, and 62.6% respectively, on the Bosphorus database [32] using 47 gallery images. This experiment shows that the proposed algorithm outperforms the benchmarks on the Bosphorus face database [32]. Alyuz et al. [46] achieved similar performance to our technique, using less than half of the database.

5.3. Face identification and verification under varying expressions in the FRGC v2 database

We compare the proposed method with recent published methods that use FRGC v2 [30] as the testing database in their experiments. Comparison is performed on both identification and verification tasks. In line with the FRGC v2 test protocol, we use the Fall2003 and Spring2004 sets containing 4007 scans of 466 subjects for the identification experiment. The gallery set of 466 faces is formed using the first neutral scan of each subject or the first stored scan for the few subjects that have no neutral scan. Two probe sets are generated. The first probe set consists of 1944 neutral (expressionless) faces and the second set consists of 1597 expressive (having a non-neutral expression) faces. Fig. 9 illustrates samples of the face images in the FRGC v2 database [30].

The identification accuracies of the proposed method and the benchmark methods on the FRGC v2 database [30] are shown in Table 5 for three different scenarios: neutral versus neutral (466 scans as the gallery and 1944 scans as the probes), non-neutral versus neutral (466 scans as the gallery and 1597 scans as the probe), and all versus neutral (466 scans as the gallery and 3541 scans as the probe). In neutral versus neutral, the rank-1 recognition rate is 99.4% for the proposed method, while the best accuracy among the benchmarks is 99.2%. In non-neutral versus neutral, the proposed method achieved 95.7%, which is the second best accuracy. In all versus neutral, the proposed method has the best performance (97.7%) compared to the benchmarks. The Cumulative Match Characteristic (CMC) curves for the three scenarios are
deformations. On the faces with expressions, due to its robustness to surface variations, the proposed method maintains a high level of accuracy even in challenging scenarios. The curves show the recognition rates (vertical axis) of the proposed method for identifying a ranked list of subjects (horizontal axes). The results show that the proposed method maintains a high level of accuracy even on the faces with expressions, due to its robustness to surface deformations.

In the verification experiment, we evaluate the proposed method using the FRGC v2 test protocols. First, we compute the True Acceptance Rate (TAR) at 0.001 False Acceptance Rate (FAR) in neutral versus neutral, non-neutral versus neutral, and all versus neutral scenarios. The results are tabulated in Table 6, and the Receiver Operating Characteristic (ROC) curves of the proposed method are illustrated in Fig. 11(a).

Then, we evaluate the proposed method using FRGC v2 database under ROC II and ROC III test protocols [10,30]. ROC II considers both the target and the query from the collection of one academic year (two semesters). However, ROC III test protocol considers the target from one semester, while the query is from another semester. True Acceptance Rates (TAR) and Equal Error Rates (EER) following ROC II and ROC III protocols of the FRGCv2 database are illustrated in Fig 11(b).

### Table 5

The rank-1 recognition rates (%) of the proposed method and the benchmarks in the face identification task on FRGC v2 database [30].

<table>
<thead>
<tr>
<th>Method</th>
<th>Neutral vs. Neutral</th>
<th>Non-neutral vs. neutral</th>
<th>All vs. neutral</th>
</tr>
</thead>
<tbody>
<tr>
<td>Haar et al. [48]</td>
<td>–</td>
<td>–</td>
<td>97</td>
</tr>
<tr>
<td>Berretti et al. [21]*</td>
<td>96.1</td>
<td>91</td>
<td>94.1</td>
</tr>
<tr>
<td>Faltiemier et al. [10]</td>
<td>–</td>
<td>–</td>
<td>97.2</td>
</tr>
<tr>
<td>Kakadiaris et al. [49]*</td>
<td>–</td>
<td>–</td>
<td>97</td>
</tr>
<tr>
<td>Drira et al. [50]*</td>
<td>99.2</td>
<td>96.8</td>
<td>97.7</td>
</tr>
<tr>
<td>Mian et al. [14]</td>
<td>99</td>
<td>95.4</td>
<td>97</td>
</tr>
<tr>
<td>Oceguera et al. [6]</td>
<td>–</td>
<td>–</td>
<td>96.6</td>
</tr>
<tr>
<td>Smeets et al. [9]</td>
<td>–</td>
<td>–</td>
<td>89.6</td>
</tr>
<tr>
<td>Proposed Method</td>
<td>99.4</td>
<td>95.7</td>
<td>97.7</td>
</tr>
</tbody>
</table>

* Results obtained from published papers.

### Table 6

<table>
<thead>
<tr>
<th>Method</th>
<th>Neutral vs. Neutral</th>
<th>Non-Neutral vs. Neutral</th>
<th>All vs. Neutral</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mian et al. [14]*</td>
<td>99.7</td>
<td>98.3</td>
<td>99.3</td>
</tr>
<tr>
<td>Wang et al. [51]*</td>
<td>99.2</td>
<td>97.7</td>
<td>98.6</td>
</tr>
<tr>
<td>Queirolo et al. [52]*</td>
<td>99.5</td>
<td>94.8</td>
<td>–</td>
</tr>
<tr>
<td>Berretti et al. [21]*</td>
<td>97.7</td>
<td>91.4</td>
<td>95.5</td>
</tr>
<tr>
<td>Proposed Method</td>
<td>99.7</td>
<td>96.3</td>
<td>98.5</td>
</tr>
</tbody>
</table>

* Results obtained from published papers.

5.4. 3D hand identification and verification in the PolyU contact-free database

Another challenging surface matching task is in the case of contact-free hand identification and verification. The major motivation of the contact-free hand recognition is that the human hands are not rigid which can be a problem in contactless hand imaging. To evaluate the proposed method, we use the PolyU contact-free hand database [18]. Fig. 12 illustrates 10 hand scans from five different subjects in the database. One scan per subject is from the first session and the other one is from the second session. As in the face recognition, we test both identification and verification scenarios.

For the identification experiment, the first hand scan from each subject is selected to make a gallery set and all the remaining scans are used as probes. In Table 8, the rank-1 recognition rate, the Equal Error Rate (EER), and the True Acceptance Rate (TAR) at 0.001 False Acceptance Rate (FAR) of the proposed method and the benchmarks in the hand identification and verification tasks are reported. The rank-1 recognition rates are 94.2% and 99.2% for the benchmark and the proposed method, respectively. The True Acceptance Rate at 0.001 False Acceptance Rate is 98.6% which is much higher than that of the benchmark approaches (87.5%, 96.2%, 80%, and 88%), while the Equal Error Rate (EER) of the proposed method is close to the best benchmark results.

5.5. Computational complexity analysis

The computational complexity of the proposed method depends on the radius of the Surface Geodesic Pattern (SGP) and the step size of the Gabor Topography Wavelet (GTW). Table 9 provides a comparison of computation times of the proposed method versus the complexity parameters using a Matlab implementation on an Intel Core i5 2.5 GHz machine with 6 GB RAM. Table 9 shows that the computation time is least sensitive to the step size ($\omega_h$). Higher variations in the computation time occur due to changes in the radius ($r$) where the time doubles when the radius increases by four times.

6. Conclusion

Textured 3D object recognition is appealing in real-world applications because most 3D sensors capture shape and texture simultaneously. However, surface deformations are inevitable in...
3D data which significantly reduces the accuracy of the extracted features. This makes rigid 3D object representation and recognition techniques unsuitable for deformable objects. In this paper, we presented Surface Geodesic Pattern (SGP) representation which is robust to non-rigid surface deformations. We also proposed Gabor Topography Wavelet (GTW) features that are extracted directly from the topography map of the 3D shape alone. Finally, we proposed a multi-view algorithm to fuse GTW and SGP features using sparse representation. Experiments on three standard datasets and comparison with benchmarks show the efficacy of the proposed technique and its robustness to non-rigid deformations.

### References

13. H. Li, D. Huang, P. Lemaire, J.-M. Morvan, L. Chen, Expression robust 3D face


Soheila Gheisari received the B.Sc. degree in control engineering from K. N. Toosi University of Technology, Iran, in 2004 and M.Sc. degree in electronic engineering from Central Tehran Branch, Azad University, Iran, in 2013. Currently, she is a PhD student at Faculty of Engineering and Information Technology, University of Technology Sydney, Australia. Her research interests include deep learning, pattern recognition, and computer vision.

Yongsheng Gao received the B.Sc. and M.Sc. degrees in electronic engineering from Zhejiang University, Hangzhou, China, in 1985 and 1988, respectively, and the Ph.D. degree in computer engineering from Nanyang Technological University, Singapore. Currently, he is a Professor with the School of Engineering, Griffith University, Australia. His current research interests include face recognition, biometrics, biosecurity, image retrieval, computer vision, pattern recognition, environmental informatics, and medical imaging.

Ajmal S. Mian received the B.E. degree in avionics from the College of Aeronautical Engineering, Nadirshaw Edulji Dinshaw University, Karachi, Pakistan, in 1993, the M.S. degree in information security from the National University of Sciences and Technology, Islamabad, Pakistan, in 2003, and the Ph.D. degree in computer science with distinction from The University of Western Australia, Crawley, Australia, in 2006. He received the Australasian Distinguished Doctoral Dissertation Award from Computing Research and Education Association of Australia in 2007. He received the prestigious 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 four national competitive research grants and is currently a Research Professor 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.