



A Curvelet-based approach for textured 3D face recognition[☆]



S. Elaiwat^{a,*}, M. Bennamoun^a, F. Boussaid^b, A. El-Sallam^c

^a School of Computer Science and Software Engineering, University of Western Australia, 35 Stirling Highway, Crawley, WA, Australia

^b School of Electrical, Electronic and Computer Engineering, University of Western Australia, 35 Stirling Highway, Crawley, WA, Australia

^c School of Sport Science, Exercise and Health, University of Western Australia, 35 Stirling Highway, Crawley, WA, Australia

ARTICLE INFO

Article history:

Received 23 February 2014

Received in revised form

7 June 2014

Accepted 11 October 2014

Available online 23 October 2014

Keywords:

Face recognition

Keypoint detection

Local features

Digital Curvelet transform

ABSTRACT

In this paper, we present a fully automated multimodal Curvelet-based approach for textured 3D face recognition. The proposed approach relies on a novel multimodal keypoint detector capable of repeatedly identifying keypoints on textured 3D face surfaces. Unique local surface descriptors are then constructed around each detected keypoint by integrating Curvelet elements of different orientations, resulting in highly descriptive rotation invariant features. Unlike previously reported Curvelet-based face recognition algorithms which extract global features from textured faces only, our algorithm extracts both texture and 3D local features. In addition, this is achieved across a number of frequency bands to achieve robust and accurate recognition under varying illumination conditions and facial expressions. The proposed algorithm was evaluated using three well-known and challenging datasets, namely FRGC v2, BU-3DFE and Bosphorus datasets. Reported results show superior performance compared to prior art, with 99.2%, 95.1% and 91% verification rates at 0.001 FAR for FRGC v2, BU-3DFE and Bosphorus datasets, respectively.

© 2014 Elsevier Ltd. All rights reserved.

1. Introduction

Despite decades of research in automatic face recognition, such a task still remains challenging in the presence of large variations in illumination, pose, facial expression or occlusion [1,2]. 2D face recognition approaches have been extensively investigated to handle these challenges [1]. However, they still suffer from sensitivity to variations in illumination, pose and facial expressions. Recent developments in low cost 3D imaging devices have the potential to address such face recognition challenges. This is because 3D approaches have been shown to be less sensitive to illumination and pose variations [2]. In addition, 3D facial images provide structural information such as geodesic distances and surface curvatures, which can greatly benefit the recognition task [3]. However, facial expressions still remain a major challenge for 3D face recognition approaches because they result in notable facial deformations [2].

Face recognition approaches combining both 2D and 3D facial images can achieve more robust recognition compared to approaches using either 3D or 2D modality alone [2]. Generally, the matching process is performed separately with respect to the data type (2D and 3D faces) before the results are fused at the score

level. For example, Chang et al. [9] fused the matching scores obtained from applying a PCA-based approach for each individual modality (3D and 2D facial images). For a small dataset (951 images with neutral expressions), reported results were 93% and 99% recognition rates for 3D and multimodal approaches, respectively. Although Bowyer et al. [2] showed that the combination of 2D and 3D data (multimodal 2D+3D approaches) gives greater performance compared to any single modality, it is still not clear whether 3D approaches outperform 2D approaches. In this paper, we propose a novel Curvelet-based multimodal approach for textured 3D face recognition.

1.1. Related works

In general, face recognition approaches, whether 2D, 3D or multimodal, can be classified into three main categories [1]:

(i) Holistic matching algorithms which extract global features from the whole face. Eigenfaces [4] and Fisherfaces [5] are well-known examples of this category. Other works such as Lu et al. [6], and Mohammadzade et al. [7] applied Iterative Closest Point (ICP) or its modified versions to match face surfaces. In general, the latter is highly affected by variations in illumination, pose, scale and facial expressions [8].

(ii) Feature-based matching algorithms which rely on matching local features or features associated to specific facial regions (e.g. eyes and nose) rather than matching the full face. Zhong et al. [9]

[☆]This research is supported by the Australian Research Council Grant DP110102166.

* Corresponding author.

E-mail address: Elaiws01@student.uwa.edu.au (S. Elaiwat).

employed Gabor features to extract intrinsic discriminative information from 3D faces. A Learned Visual Codebook (LVC) was then constructed using K-means clustering. Berretti et al. [3] extracted isogeodesic stripes from 3D faces and then represented these stripes using a 3D Weighted Walkthroughs (3DWWs) descriptor. For matching, a graph-based matching algorithm was applied to different faces. Creusot et al. [10] and Berretti et al. [11] proposed methods for extracting distinctive keypoints/landmarks for robust recognition. In general, feature-based matching algorithms are known to be more robust to facial expressions, pose, illumination, scale and occlusions variations [8]. This is because they can exclude those facial regions that could be most affected by perturbations such as changes in facial expression or spurious elements [3]. However, there is no set of face regions (local features) that is perfectly invariant across all facial expressions [2].

(iii) Hybrid matching algorithms which exploit a combination of both holistic and local feature-based matching. Huang et al. [12] proposed a Multi-Scale Local Binary Pattern (MS-LBP) depth map to represent the 3D facial surface in conjunction with the Shape Index (SI) map. SIFT algorithm was then applied on both maps to extract local features. The hybrid matching algorithm then carried out feature-based matching on local feature and holistic matching based on the facial component and configuration constraint. The combination of holistic and local feature-based matching has the potential to give better performance, but at the cost of greater computational cost.

Multi-resolution algorithms have been widely used in conjunction with the aforementioned feature-based matching, Holistic matching and hybrid matching. Jing et al. [13] proposed a 2D face recognition approach based on Fractional Fourier transform and discrimination analysis technique. They firstly adjusted the angle value of Fractional Fourier transform using 2-dimensional separability judgement. A reformative Fisherface method was then applied to extract features. Fauqueur et al. [14] and Bendale et al. [15] employed Complex Wavelet Transform (DTCWT) for keypoint detection. Mandal et al. [16] applied Curvelet transform along with two different dimensionality reduction algorithms (PCA-LDA) for 2D face recognition. A set of coefficients characterized by high variance was firstly selected and then projected by PCA-LDA to a lower dimensional space. Rziza et al. [17] proposed to extract local features from 2D faces mapped to Curvelet domain by dividing each Curvelet subband into a set of equally sized blocks. In order to define local features, each block was represented by its mean, variance and entropy values. All features were then combined and projected by LDA.

Compared to other transforms, Curvelet transform is strongly anisotropic and its needle-shaped elements provide a high directional sensitivity to represent curved singularities in images. In contrast, Wavelet transform exhibits a good representation only at point singularities, because it has a poor directional sensitivity (isotropic base function). Other directional transforms such as Dual-Tree Complex Wavelet Transform (DTCWT) and Gabor Wavelets perform better than Wavelets but still have limited directional selectivity. Finally, the Ridgelet transform is only suitable for representing global straight-line singularities in objects, which are rarely found in real applications [18,19].

1.2. Paper contributions

Although Curvelet transform provides a powerful framework to extract distinctive surface features. Curvelet-based face recognition approaches have been so far mainly limited to holistic matching of 2D global features extracted from the whole face [16,20]. Recently, we proposed a multimodal face identification approach based on Curvelet transform to extract features from semi-rigid regions (eyes-forehead and nose) [21]. Since these

regions are less sensitive to facial expressions, the proposed approach achieves good identification rates under different facial expressions. However, these regions were segmented using static masks which cannot accurately extract eyes-forehead and nose regions for all faces, especially from different datasets. Furthermore, the locality of the extracted features is low, making them sensitive to deformations resulting from facial expressions.

To address these limitations, this paper proposes a novel multimodal approach that introduces the following contributions:

- A multimodal keypoint detector to extract robust and distinctive keypoints from textured 3D faces. The identification of these keypoints is carried out in the Curvelet domain after decomposing each face (depth and texture information) into a set of scale and angle decompositions. Identified keypoints are associated to local face surfaces rich with geometrical and textures features. Because these keypoints are extracted separately across different frequency bands, this allows us to identify more distinctive keypoints on local surfaces associated with high variations. As a result, our keypoint detector is shown to exhibit high repeatability in textured 3D faces.
- A method has been proposed to measure the repeatability of the detected keypoints in the Curvelet domain. The proposed method builds an accumulated map combining all Curvelet coefficients from different subbands. This map facilitates finding the repeatability between detected keypoints in different subbands without requiring to inverse back these keypoints to the spatial domain. A keypoint is considered to be repeatable if it appears at nearly the same location in two accumulated maps (corresponding to two faces) of the same subject.
- A multimodal local surface descriptor to capture highly descriptive local features around extracted 2D and 3D keypoints. In contrast, previous works using Curvelet transform, such as [16], extracted 2D features with holistic matching. Here, our algorithm extracts both 2D and 3D local features around the detected keypoints by including all directional decompositions in the mid-bands to precisely represent geometric/texture features while minimizing sensitivity to noise. As a result, our multimodal surface descriptor is shown to achieve superior performance.

Preliminary results of this approach appeared in [22], which reports only the 3D modality. Furthermore, extensive experiments have been carried out for each feature scale (scale 2, scale 3 and combined scale 2+3) and each modality (2D, 3D and multimodal 2D+3D) on three datasets (FRGC v2, BU-3DFE and Bosphorus) under different scenarios including facial expressions, pose variations and time laps between target and query faces.

1.3. Paper organization

The rest of this paper is organized as follows. Section 2 gives a brief overview of the Curvelet transform and its digital form. Section 3 details the proposed 3D/2D keypoint detection algorithm. Section 4 describes the extraction and construction of our 3D/2D local features. Details of the matching algorithm are given in Section 5. Experimental results are reported in Section 6. Finally, conclusions are drawn in Section 7.

2. Digital Curvelet transform

The Curvelet transform, originally developed in 1999 by Donoho and Duncan [23], is a multi-scale and multi-directional representation with highly anisotropic behaviour. The second generation of Curvelet transform, reported in [24], uses a frequency partition

technique to make it not only simpler but also faster and less redundant than its first generation counterpart.

Given a function $f(\mathbf{x}) \in L^2(\mathbb{R}^2)$, $\mathbf{x} = (x_1, x_2)$, the continuous Curvelet transform is defined by [24]

$$C(\zeta, \theta, v) = \int_{\mathbb{R}^2} f(\mathbf{x}) \varphi_{\zeta, \theta, v}(\mathbf{x}) d\mathbf{x}, \quad \zeta > 0, \theta \in [0, 2\pi), v \in \mathbb{R}^2 \quad (1)$$

$\varphi_{\zeta, \theta, v}$ represents the continuous “mother Curvelet”, which projects $f(\mathbf{x})$, to provide the continuous Curvelet coefficient (CCT) $C(\zeta, \theta, v)$ at a scale ζ , orientation θ and location v [24].

For ease of understanding the transform and its properties, it is normally represented at discrete intervals formed by sampling the CCT at dyadic intervals $\zeta_a = 2^{-a}$, $a = 1, 2, \dots$ scales, $\theta_h = 2\pi 2^{-a/2} h$, $h = 0, 1, \dots$ orientations (subbands), and equispaced sampling on a rotated anisotropic grid in space $v_b = (i, j) \in \mathbb{Z}^2$, i.e.

$$C(a, h, b) = \int_{\mathbb{R}^2} f(\mathbf{x}) \varphi_{a, h, b}(\mathbf{x}) d\mathbf{x} \quad (2)$$

This sampled formulation can be represented by the discrete Curvelet formula:

$$u(a, h, b) = \sum_{0 \leq x_1, x_2 < N} f[x_1, x_2] \varphi_{a, h, b}^D[x_1, x_2] \quad (3)$$

where $u(a, h, b)$ are the discrete Curvelet coefficients, $\varphi_{a, h, b}^D$ are the discrete mother Curvelets and $f(\mathbf{x}) = (x_1, x_2)$ is the sampled data (e.g. an image intensity). In order to implement the discrete Curvelet transform, the discrete mother Curvelet waveform is defined by means of its Fourier transform at a parabolic window (wedge) $\mathcal{U}_{a, h}$. This window smoothly localizes the Fourier transform near the sheared wedges which obey the parabolic scaling [24]. In this paper, the wrapping implementation of FDCT is adopted because it is considered to be the fastest Curvelet transform algorithm available [25].

3. Keypoint detection

To extract keypoints from textured 3D faces. Each facial image is decomposed, using Curvelet transform (Eq. (3)), into a set of Curvelet coefficients at each scale and angle θ (subband) as shown in Fig. 1. Notice that the angle decomposition in Curvelets produces identical subbands at angle θ and $\pi + \theta$ within, the same scale. Subbands at angle $(\pi + \theta)$ can thus be excluded. For keypoint selection, Curvelet coefficients' magnitude are examined at each scale by evaluating:

$$u'_{a, h, b} = u_{a, h, b} |c_{a, h, b}| \geq \bar{u}_a, \quad \forall h, \forall b_j \quad (4)$$

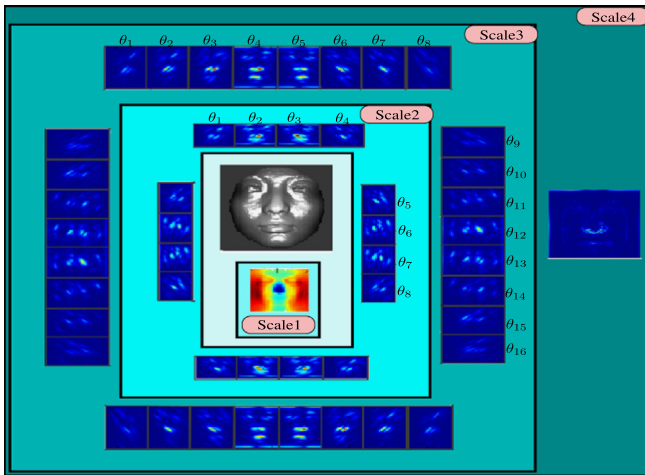


Fig. 1. Illustration of a 4 scales Curvelet decomposition.

where $u'_{a, h, b}$ and \bar{u}_a represent a keypoint at scale a , subband h and Curvelet position $b = (i, j)$, and the mean value of all coefficients at scale a , respectively. The mean value has been used before as a threshold on Wavelet and Curvelet transforms and other multi-resolution algorithms for including image enhancement, de-noising, feature extraction and coefficients filtering [18]. The mean value was also used as the best threshold yielding the best F-measure in [26]. The latter dealt with detecting and estimating dominant peaks of onset frequencies of signals. In our case, the mean value (Fig. 2A) offers an automated and efficient way to detect the most dominant coefficients (keypoints), which are mainly the Curvelet harmonics representing curvature variations.

Extracting keypoints from different subbands within the same scale could produce several keypoints with the same spatial position. This would result in redundant keypoints and thus redundant features (Fig. 2B and C). To address this issue, each detected keypoint $u'_{a, h, b}$ is examined w.r.t other keypoints in all other subbands within the same scale. A given keypoint $u'_{a, h, b}$ will be discarded in two scenarios:

Scenario1:

$$u'_{a, h, b} < \max(u'_{a, h', b'}) |x_{h', b'} \approx x_{h, b}, \quad \forall h', \forall b' \quad (5)$$

where $u'_{a, h', b'}$ represents any keypoint within the same scale of the examined keypoint $u'_{a, h, b}$. $x_{h', b'}$ and $x_{h, b}$ are the spatial positions of keypoints $u'_{a, h', b'}$ and $u'_{a, h, b}$, respectively. The first occurs for the case of two or more keypoints found at the same spatial position within the same scale (Fig. 2C scenario 1).

Scenario2:

$$u'_{a, h, b} < \lambda \max(u'_{a, h', b'}) |x_{h', b'} \in w(x_{h, b}), \quad \forall h', \forall b' \quad (6)$$

where w represents a patch centered around the spatial position $x_{h, b}$ of the examined keypoint. λ is a weighting factor whose value (set to 0.8) is used to only keep the most significant keypoints in the patch. Fig. 2C illustrates scenario 2 (after applying scenario 1), where the keypoint $u'_{a, h, b}$ under consideration is filtered out if its magnitude is λ times lower than the maximum magnitude of any of the keypoints $u'_{a, h', b'}$ that are present within the same patch. We have found that up to 60 keypoints can be filtered out using scenarios 1 and 2.

In our experiments, we applied four scales Curvelet transform to represent geometric facial features. Higher scale decomposition would increase the number of Curvelet components (coefficients), but would lead to a poorer representation. In this work, both the coarsest and finest scales have been discarded because they capture information in the lowest (smooth surfaces) and the highest (noise and image boundaries) frequency bands. In addition, both of these scales do not have an angle decomposition. Furthermore, all Curvelet coefficients falling on the face boundaries and mouth area, including chin/beard (areas below nose) were automatically excluded. The motivation behind this comes from two important findings reported in [2]. Firstly, the upper part of the face has been shown to be more significant for recognition compared to the lower part. Secondly, the eyes, the forehead, and the nose have been shown to be less sensitive to facial expressions compared to the mouth and cheeks.

Several previous works including [21,27,28] applied a special mask to separate these regions (semi-rigid regions) from the whole face in order to achieve robust recognition. However, extracting only the rigid regions from the entire face is a complex task that is difficult to automate. Instead, we chose here to exclude the parts which are the most affected by facial expressions (mouth and chin/beard regions) rather than using special masks for the nose, eyes and forehead. This was achieved by defining a horizontal reference line \perp at a distance d below the nose tip. All Curvelet coefficients below line \perp are excluded because they belong to the mouth region. Fig. 3A shows a face and its excluded

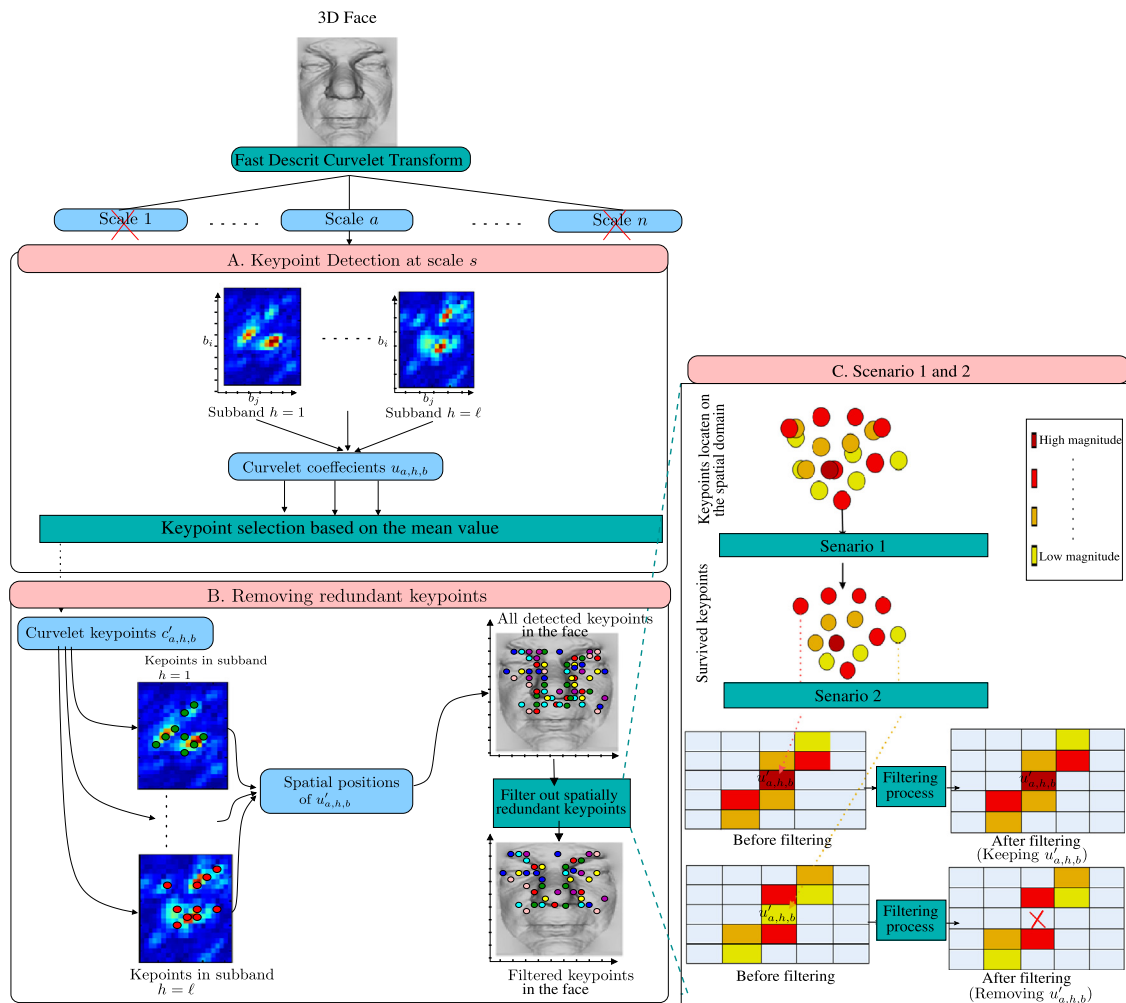


Fig. 2. Block diagram of the keypoint detection algorithm (figure best seen in color). (For interpretation of the references to color in this figure caption, the reader is referred to the web version of this paper.)

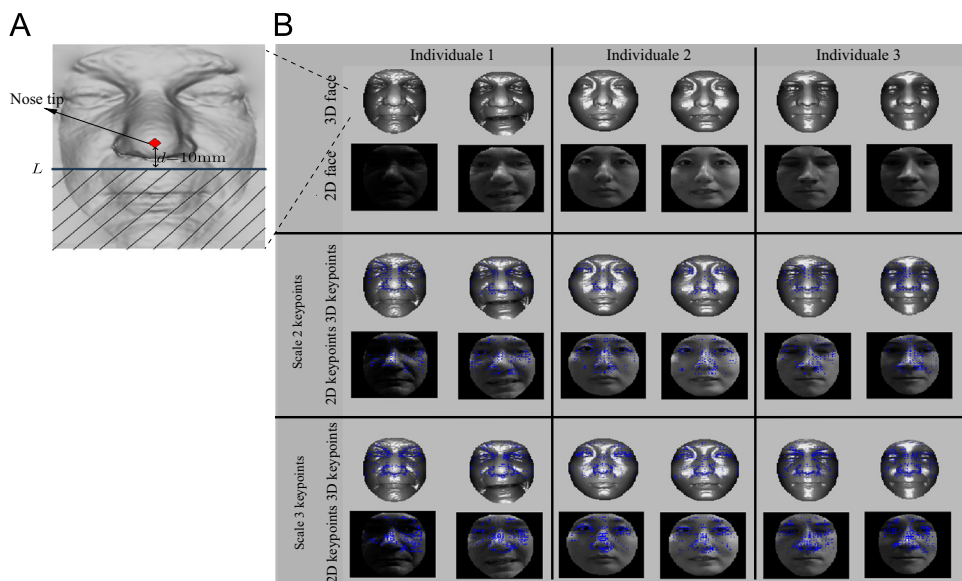


Fig. 3. (A) Excluded mouth region, (B) keypoints of scales 2 and 3 on the 2D and 3D faces.

regions. The value of d has been chosen equal to 10 mm to ensure that the excluded part does not include any part of the nose. Fig. 3B illustrates the detected keypoints in 2D and 3D faces.

We notice that the detected keypoints are located on the local surfaces which are associated with high curvatures such as the eyes and nose surfaces. During our experimental tests, we observed that

the nose, eyes and forehead regions represent the most important parts for correct matching in the case of neutral faces. For non-neutral faces, we observed that the matches resulting from the nose regions are the least affected by facial expression. As a result, they give better matches compared to the eyes and forehead regions. Since we are proposing a fully automated algorithm, we did not score/weight the matches based on the location of the keypoints.

3.1. Keypoints repeatability

A keypoint is said to be repeatable if it appears at nearly the same location and scale in two onsets (e.g. two faces) of the same subject, with one onset being a transformed version of the other. We have investigated the repeatability of our keypoint detector on the FRGC v2 dataset. However, the ground truth for keypoints of the 3D/2D faces of the dataset is not available. To address this issue, two approaches are typically adopted in the literature: (1) generate synthetic data from the ground truth, where known transformations are applied e.g. translation, scaling, rotation or noise [29]. Keypoints are then detected and their transformations are compared with those generated synthetic ones, or (2) estimate the ground truth data (keypoints data) by registering different faces using an accurate registration algorithm, where errors between the two registrations are calculated and used to indicate the repeatability of the detected keypoints [8]. In this work, we followed the second scenario as it has some insights and gives more realistic measures [8].

Given that both keypoints and local descriptors are built in the Curvelet domain (Sections 3 and 4); the repeatability of the keypoints was measured in the Curvelet domain. since the keypoints are falling in different subbands within the same scale, all Curvelet coefficients of all subbands within the same scale were mapped to one map (accumulated map). This map facilitates finding the repeatability between keypoints in different subbands. Building an accumulated map using straight forward methods (e.g combination) as other transforms such as Dual-Tree Complex Wavelet transform [14] is not applicable in our case due to Curvelet subbands have different sizes. To address this, we could either expand (up-sample and interpolate) all subbands to the original image size before combining them into a single map, or map all coefficients directly into the accumulated map based on their spatial positions and magnitudes. To avoid the computation complexity and non-linearities associated to up-sampling, we mapped all coefficients into the accumulated map using their spatial positions and magnitudes.

In cases when there were two or more coefficients falling in the same position (from different subbands), the highest magnitude would be selected. Fig. 4 illustrates the accumulated maps of scales 2 and 3.

The accumulated maps of different images (at a certain scale) of a given individual are then automatically registered using the ICP algorithm. Registration errors are evaluated using the distances between nearest neighbors of keypoints at each scale (one from each face). Fig. 5 reports the cumulative repeatability ratio of facial images with 160 mm height and 160 mm width (1 mm \equiv 1 pixel) as a function of increasing distance for neutral and non-neutral expressions. The repeatability of the detected 3D keypoints was found to reach 76% and 88% for scales 2 and 3, with neutral expression at an error of 4 mm, respectively. The total repeatability of scales 2 and 3 was 84%. In the case of non-neutral expressions, the total repeatability (scale 2 + 3) of 3D keypoints at 4 mm, dropped by only 2% to reach 82%. The total repeatability of (scales 2 + 3) for the detected 2D keypoints reaches 78% with neutral expression at an error of 4 mm. For non-neutral expressions, the detected 2D keypoints achieved the same repeatability than neutral expressions. This is to be expected since 2D data is less affected by facial expressions than 3D data.

As a result, the key characteristics of our keypoint detector can be summarized as follows: (i) the detected 2D/3D keypoints are repeatable when considering range images of the same person. (ii) the keypoints are identified in different frequency scales and orientations, allowing us to capture of different geometrical and texture features. (iii) the detected keypoints vary between two different individuals (people).

4. Feature extraction

After identifying keypoints on both 2D and 3D facial images, local surface descriptors were constructed around each detected keypoint in the Curvelet domain, considering all subbands of the scale in which the keypoint was detected (Fig. 6). Assume that we have a keypoint $u_{a,h,b}$ with point $q_j = (q_{j_1}, q_{j_2})$ referring to the corresponding position (coefficient) of the keypoint in subband j . We firstly extract a (5×5) subband patch $\mathcal{W}_{a,j}$ centered around q :

$$\mathcal{W}_{a,j} = u_{a,j,x}, \quad \forall x = (q_j - 2:q_j + 2) \quad (7)$$

where $u_{a,j,x}$ represents a Curvelet coefficient at scale a , subband j , and position x . The size of each subband patch determines the degree of locality of the features. In general, large subband patches could result in high sensitivity to facial expressions, while small subband patches give a weak descriptiveness.

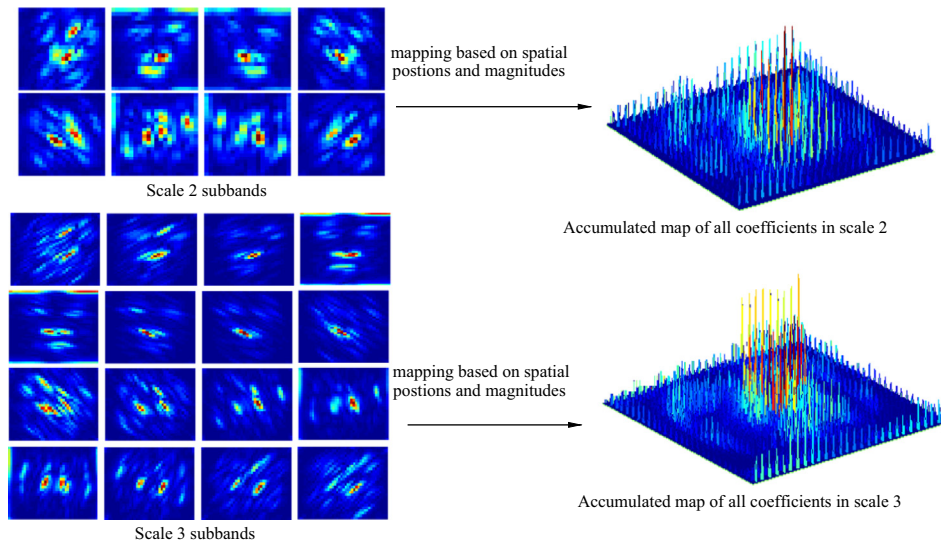


Fig. 4. Accumulated maps of scales 2 and 3.

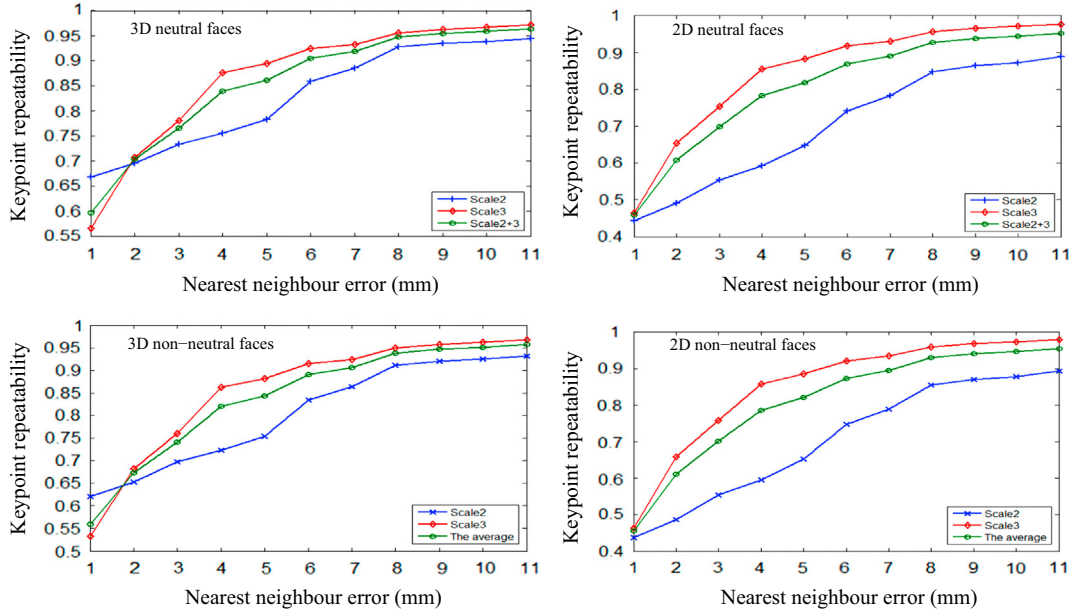


Fig. 5. Keypoints repeatability (3D + 2D faces) for neutral expressions.

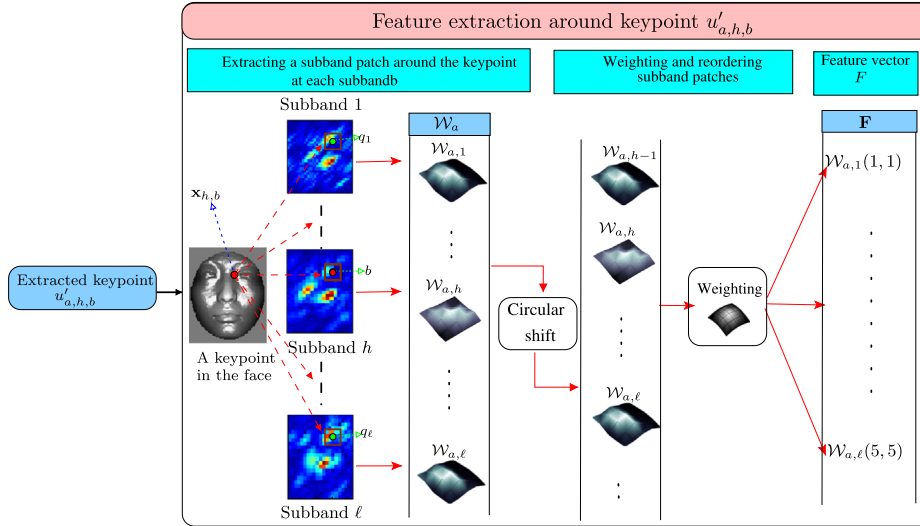


Fig. 6. Block diagram of the feature extraction algorithm.

We fixed the size of the patches to 5×5 after testing the algorithm with varying subband patch sizes varying from 3×3 to 7×7 . 7×7 subband patches were found to exhibit a better performance (descriptiveness) for neutral cases (5% higher than 5×5 patches) compared to the case when facial expressions are present (4% lower than 5×5 patches). 5×5 subband patches offered the best trade-off between descriptiveness and sensitivity to facial expressions in all cases. Note that these patches are extracted from the Curvelet subbands which are down-sampled (from the original size) during the decomposition process. As a result, a small patch with size $n \times n$ in the Curvelet domain is equivalent to a patch whose size is at least double ($2n \times 2n$) in the spatial domain (image). The overall patch \mathcal{W}_a of keypoint $u_{a,h,b}$ is then constructed from all subband patches:

$$\mathcal{W}_a = \{\mathcal{W}_{a,1}, \mathcal{W}_{a,2}, \dots, \mathcal{W}_{a,\ell}\}^T \quad (8)$$

where ℓ correspond to the number of subband patches.

Although Curvelet coefficients provide a powerful representation of geometrical features, they are sensitive to rotation [30]. To address this, all subband patches in \mathcal{W}_a are reoriented based on

the keypoint orientation (subband) using circular shift:

$$\begin{aligned} &[\mathcal{W}_{a,1}, \dots, \mathcal{W}_{a,h-1}, \mathbf{W}_{a,h}, \dots, \mathcal{W}_{a,\ell}] \\ &\quad \downarrow \text{circularshift} \downarrow \\ &[\mathcal{W}_{a,h-1}, \mathbf{W}_{a,h}, \dots, \mathcal{W}_{a,\ell}, \mathcal{W}_{a,1}, \dots] \end{aligned} \quad (9)$$

Note that the keypoint subband patch is denoted by $\mathbf{W}_{a,h}$ and is located at a fixed position (e.g. the second position) to extract the same features structure from different keypoints.

The elements (coefficients) in each subband patch j in \mathcal{W}_a are then scaled w.r.t the center point q_j using a weighting kernel function. The scaled patch \mathcal{W}'_a is defined by

$$\mathcal{W}'_a = \mathcal{W}_{a,j} e^{-(1/2)(\Delta/(\mathcal{N}/2)\sigma)^2} |_{\forall j=1:\ell, \mathcal{N}=5, \sigma=1} \quad (10)$$

where $e^{-(1/2)(\Delta/(\mathcal{N}/2)\sigma)^2}$ is the weighting Gaussian function with \mathcal{N} window size while Δ represents the Euclidean distance of a coefficient to the point q_j . Thus more emphasis is given to those coefficients that are closer to the keypoint.

Finally, all elements of the patch \mathcal{W}'_a are concatenated into one feature vector F . This feature extraction is applied for scales 2 and 3 of 2D and 3D faces (Fig. 6).

5. Feature matching

During online recognition, sets of features in a probe face are matched with corresponding sets of features in all gallery faces. Let $\mathbf{F}_{P,k}$ and $\mathbf{F}_{G,k}$ denote the set of features of a probe and a gallery face using modality $k \in (3D \text{ scale}2, 3D \text{ scale}3, 2D \text{ scale}3, 2D \text{ scale}3)$, respectively:

$$\mathbf{F}_{P,k} = \{\mathbf{f}_1, \mathbf{f}_2, \dots, \mathbf{f}_m\}, \quad \mathbf{F}_{G,k} = \{\mathbf{f}'_1, \mathbf{f}'_2, \dots, \mathbf{f}'_n\} \quad (11)$$

A feature of a probe and a gallery face are represented by \mathbf{f} and \mathbf{f}' , respectively. m is the number of features in the probe face and n is the number of features in the gallery face.

To calculate the similarity between two faces, corresponding features of each modality k are matched using cosine rules given by [8]

$$\mathbf{M}_{P,G,k} = \cos(\mathbf{F}_{P,k}^T, \mathbf{F}_{G,k}^T) \quad (12)$$

where $\mathbf{M}_{P,G,k}$ corresponds to the similarity between probe face \mathcal{P} and gallery face \mathcal{G} for modality k . When the two features $f (f \in \mathbf{F}_{P,k})$ and $f' (f' \in \mathbf{F}_{G,k})$ are exactly equal, $\mathbf{M}_{P,G,k}(i,j)$ will be 1, corresponding to a perfect match. For each probe feature (row) of $\mathbf{M}_{P,G,k}$, the best match is extracted to form one vector $\mathbf{s}_{P,G,k}$ constructed by concatenating all best matches:

$$\mathbf{s}_{P,G,k} = (\max(\mathbf{M}_{P,G,k})) \quad (13)$$

The max operator is used to find the best match for each probe feature in $\mathbf{M}_{P,G,k}$. In the same manner, the best matches between probe face \mathcal{P} and all gallery faces are extracted to form a matrix $\mathbf{S}_{P,k}$:

$$\mathbf{S}_{P,k} = [\mathbf{s}_{P,1,k}, \dots, \mathbf{s}_{P,g,k}]^T \quad (14)$$

where g refers to the size of gallery. Each column i in $\mathbf{S}_{P,k}$ represents the best matches between probe feature i and all gallery features. Prior to scoring matches of each gallery face, the matrix $\mathbf{S}_{P,k}$ is normalized in each column using min-max rule, which rearranges the values between 0 and 1. The normalized matrix $\hat{\mathbf{S}}_{P,k}$ is defined by

$$\hat{\mathbf{S}}_{P,k}(:,i) = \frac{\mathbf{S}_{P,k}(:,i) - \min(\mathbf{S}_{P,k}(:,i))}{\max(\mathbf{S}_{P,k}(:,i) - \min(\mathbf{S}_{P,k}(:,i))) - \min(\mathbf{S}_{P,k}(:,i))} \quad (15)$$

The score matching between probe face \mathcal{P} and gallery face \mathcal{G} is then calculated as the mean value of the best matches (in row \mathcal{G} of matrix $\hat{\mathbf{S}}_{P,k}$) that are greater than or equal to a pre-determined threshold value β .

This matching algorithm results in a similarity vector \mathbf{r}_k found by matching a probe face with all gallery faces using modality k . Given that all similarity vectors have a positive polarity, the highest scores will correspond to the best matches. To fuse different similarity vectors (different modality), each similarity vector is first normalised

using the min-max rule:

$$\hat{\mathbf{r}}_k = \frac{\mathbf{r}_k - \min(\mathbf{r}_k)}{\max(\mathbf{r}_k - \min(\mathbf{r}_k)) - \min(\mathbf{r}_k - \min(\mathbf{r}_k))} \quad (16)$$

The overall similarity vector \mathbf{r} between the probe face and gallery faces is calculated by fusing the similarity vectors from different modalities using a confidence weighted sum rule [8]:

$$\mathbf{r} = \sum_{k=1}^n \alpha_k \mathbf{r}_k \quad (17)$$

where α_k represents the confidence measure of each modality. α_k can be either chosen based on experimental results or can be set automatically by [8]:

$$\alpha_k = \frac{\max(\mathbf{r}_k) - \bar{\mathbf{r}}_k}{\max_2(\mathbf{r}_k) - \bar{\mathbf{r}}_k} \quad (18)$$

The operator \max_2 is used to find the second maximum value in the vector. Finally, the resulting vector \mathbf{r} is used to make the decision. In our experiments, we fused similarity vectors in different scenarios based on data type (2D, 3D) and scales (scale 2 and scale 3). More details are provided in the next section.

Because our implementation of the Curvelet transform is based on FFT [24], our proposed keypoint detector and surface descriptor exhibit a relatively low computational complexity compared to other keypoint detectors [29]. The matching process can be expensive in the case of 3D+2D modality, since we need $m \times n$ matches for each $k \in (3D \text{ scale}2, 3D \text{ scale}3, 2D \text{ scale}3, 2D \text{ scale}3)$. To reduce the complexity of matching, The keypoint (feature) matching is only carried out between those keypoints which fall within the same region. In other words, if the Euclidean distance between them is less than a threshold value ϑ , we will match them. By testing 500 random matches, we observed that the 3D+2D modality took an average of 0.36 s to achieve one full match (between two faces). This evaluation of the computational cost was carried out on a standard desktop with an Intel Core i7 3.40 GHz processor with 8.0 GB RAM. Further improvements can be achieved using FPGAs or GPUs real-time implementations of the proposed algorithms. However, our main focus in this paper is more on the accuracy and robustness compared to the computational complexity.

6. Experimental results

We performed our experiments on three datasets, namely FRGC v2 (experiment 3), BU-3DFE and Bosphorus. The FRGC v2 dataset consists of 4950 3D scans along with their texture maps (2D), in the presence of large variations in facial expressions and illumination conditions, but limited pose variations. This dataset is divided into three main sets: Spring2003, Fall2003 and Spring2004. The first subset is used for training while the remaining two subsets are used for validation. The validation set contains 4007 scans (3D along with

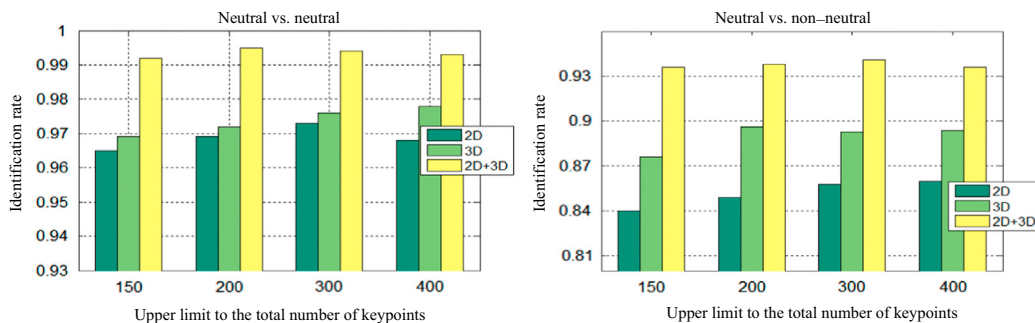


Fig. 7. Identification rates as a function of the upper limit set for the total number of keypoints.

2D) of 466 subjects partitioned into gallery, neutral and non-neutral expressions. 466 scans under neutral expression were used to build the gallery, while the remaining 3541 scans represent probes categorized into 1944 neutral scans (neutral expressions) and 1597 non-neutral scans (non-neutral expressions). To include more challenging cases, we tested our algorithm on the BU-3DFE and Bosphorus datasets. The BU-3DFE dataset contains six different facial expressions, namely Angry, Disgust, Fear, Happy, Sad and Surprise with four levels of expression intensity for each subject. The total number of facial expressions is 2500 distributed over 100 subjects. In contrast, the Bosphorus dataset contains 35 facial expressions

(1 Neutral, 20 Lower Face Action Units, 3 Action Unit Combinations and 6 Emotional Expressions) and 13 pose variations (7 Yaw Rotations, 4 Pitch Rotations and 2 Cross Rotations). More details about these datasets can be found in [31–33].

Since the raw data (3D data along with the corresponding textured maps) could contain non-facial elements (e.g. background, clothes, ears and neck) together with noise (e.g. holes and spikes), each raw data was normalized using the normalization algorithm in [34,35]. Four main steps were employed to normalize each 3D face and its corresponding 2D face: (1) Face localization based on the nose tip using a coarse to fine approach.

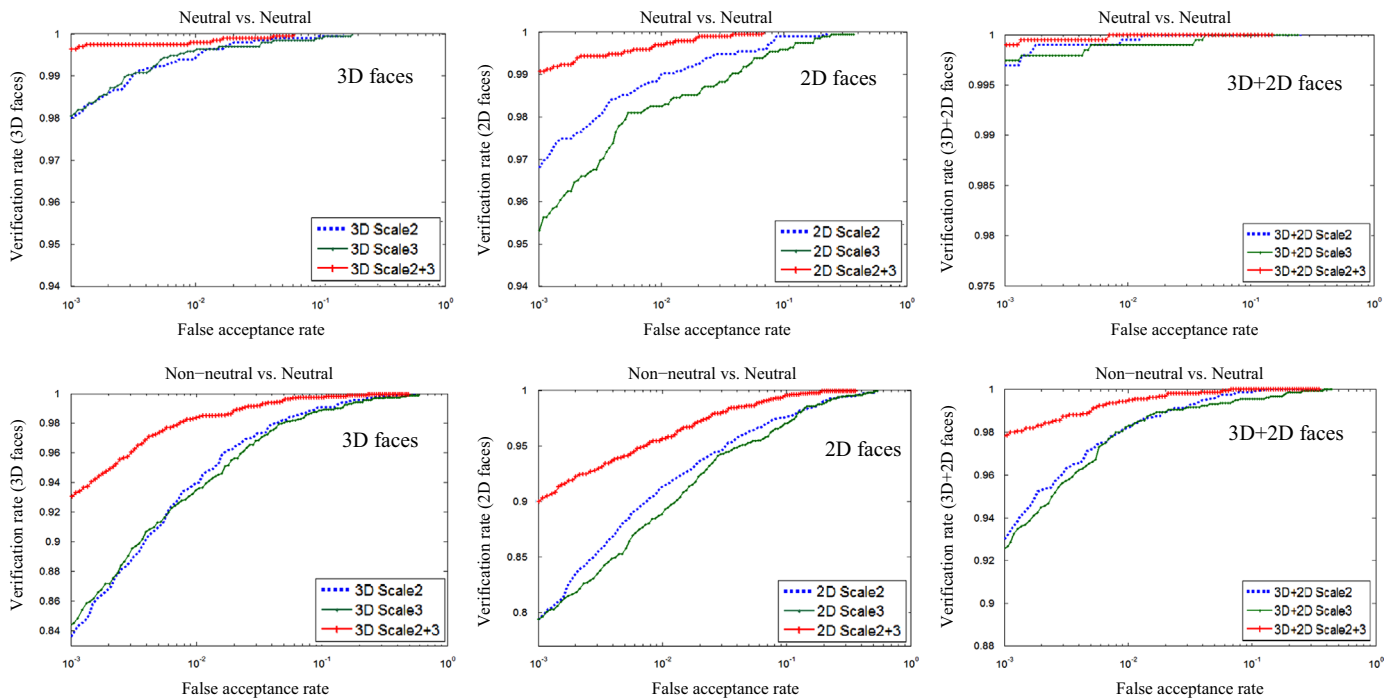


Fig. 8. Verification results for the FRGC v2 dataset.

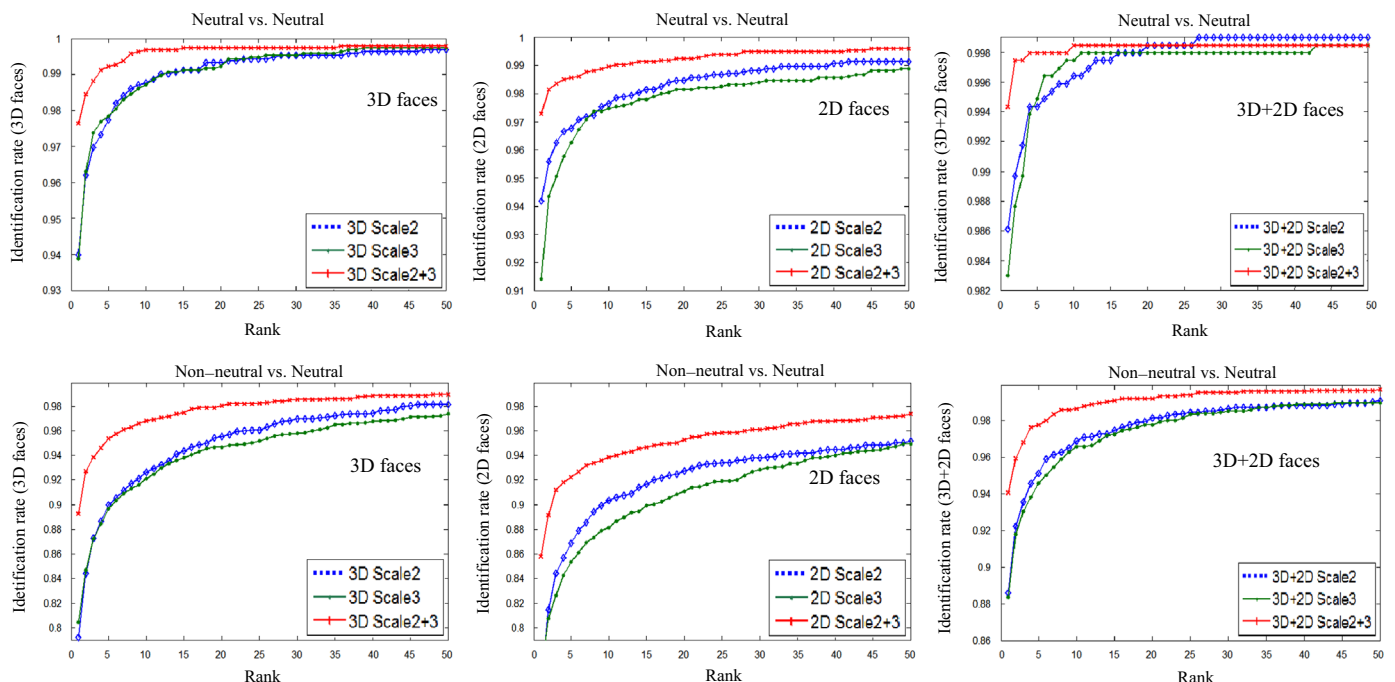


Fig. 9. Identification results for the FRGC v2 dataset.

A sphere centered at the nose tip is subsequently used to crop the 3D face and its corresponding 2D face. (2) Removal of spikes resulting from outliers. To this end, each point of the 3D faces was examined based on the distance to its 8-neighbours. If the distance is greater than a certain threshold value, it is considered as an outlier and removed. Removing outliers could produce holes, which can be filled by means of a cubic interpolation. (3) Pose correction with PCA alignment iteratively applied to correct the pose of the 3D faces and their 2D faces counterparts. In the case of high pose variations (e.g. profile faces), the ICP algorithm is used instead of PCA to align faces (probes) w.r.t the gallery faces [35]. (4) Symmetric filling. After pose correction, we use facial symmetry to fill in the missed parts resulting from pose correction, by mirroring the original point cloud from (X) to $(-X)$ (the nose tip represents the origin point $(0,0)$). A mirrored point is added to the original point cloud if its Euclidean distance (in X and Y only) to the closest point in the point cloud is less than a threshold value ξ . However, the symmetric filling can produce spikes and cannot fill in all holes. These spikes and holes are smoothed out by fitting a surface to the points [35].

To ensure that the recognition results are not biased in favour of the gallery faces which have more keypoints, we tested our recognition results w.r.t different upper limits set for the total number of keypoints in the case of a 2D and 3D modality. Fig. 7 illustrates how the performance varies w.r.t the chosen upper limit set for the total number of keypoints. Although our algorithm ensures that the total number of detected keypoints do not exceed the prescribed upper limit (e.g. 150, 200, etc.), the actual total number of detected keypoints can be much less than that. This condition was applied in the case of the 2D and 3D modalities, separately. 3D+2D modality results from fusing the 3D and 2D modalities at the score level. In practice, we set the value of the total number of keypoints to 200, which provides a good trade-off between performance and computational cost.

Fig. 8 reports our identification results for each modality and each scale. In the case of neutral expressions, the 3D local features achieved identification rates of 94%, 93.1% and 97.6% for scale 2, scale 3 and the combined scales (scale 2 + 3), respectively. The 2D local features achieved identification rates of 94.2%, 91.4% and 97.3% for scale 2, scale 3 and combined scales 2 and 3, respectively. In the case of non-neutral expressions, the identification rates of 79.2%, 80.5%, 89.3%, 75.3%, 76.5% and 85.8% were achieved by scale 2 of 3D features, scale 3 of 3D features, the combined scales of 3D features, scale 2 of 2D features, scale 3 of 2D features and the combined scales of 2D features, respectively. Scale 2 achieved better identification rates than scale 3 because more distinctive features are falling in the frequency band of scale 2. In the case of non-neutral expressions, scale 2 features were shown to be more sensitive to facial expressions than scale 3 features. The identification rates of the combined scales for (3D + 2D) local features are 99.4% and 94.1% for probes with neutral and non-neutral expressions, respectively.

Fig. 9 reports the ROC curves of our algorithm for neutral and non-neutral expressions. At 0.001 FAR (False Acceptance Rate), the 3D features achieved verification rates of 99.6% and 93.1% for the

Table 1
Identification results and comparison on FRGC v2 dataset.

	Algorithm	Neutral vs. Neutral	Non-neutral vs. Neutral	All vs. Neutral
This paper	Curvelet local features	99.4%	94.1%	97.1%
[39]	Radial curves	–	–	97%
[37]	Gauss–Markov posterior marginals	–	–	96.6%
[3]	Isogeodesic stripes	≈ 96%	≈ 91%	≈ 94%

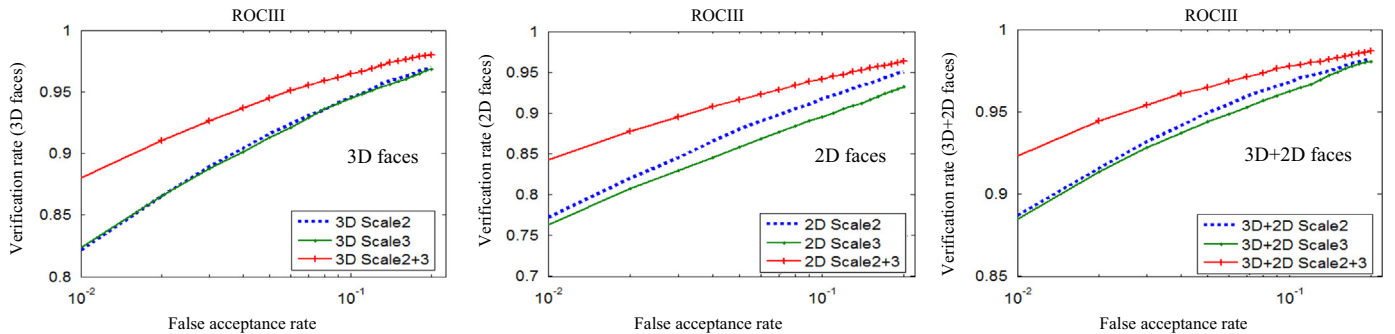


Fig. 10. ROCIII results for the FRGC v2 dataset.

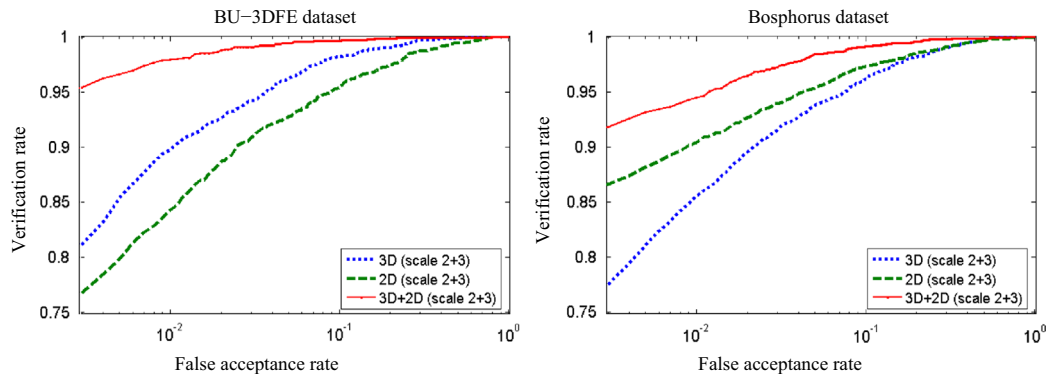


Fig. 11. Verification results on BU-3DFE and Bosphorus.

Table 2
Verification results and comparison on FRGC v2 dataset.

Algorithm		FAR	Neutral vs. Neutral		Non-neutral vs. Neutral		All vs. Neutral		ROCIII
			0.001	0.1	0.001	0.1	0.001	0.1	
This paper	Curvelet local features		99.9%	100%	98%	100%	99.2%	100%	97.8%
[7]*	Iterative closest normal point	–	–	–	–	98.5%	–	–	99.6%
[39]	Radial curves	–	–	–	–	–	–	–	97.14%
[40]	Sparse bounding sphere	–	–	–	–	≈ 92.6%	–	≈ 95.8%	96.03%
[3]	Isogeodesic stripes	97.7%	–	–	91.4%	–	95.5%	–	–
[41]	Simulated annealing	–	–	–	–	–	–	98.2%	96.6%
[42]	Gabor features	–	–	–	–	–	–	97.5%	95.3%

* Ref. [7] uses 7 scans per subject in the gallery while other works use 1 scan per subject.

combined scales with neutral and non-neutral expressions, respectively. In contrast, the 2D features achieved verification rates of 99.1% and 90% for the combined scales with neutral and non-neutral expressions, respectively. The verification rates of the combined scales of combined 3D and 2D features are 99.9% and 98% with neutral and non-neutral expressions, respectively. Fig. 10 reports the ROCIII experiment for each modality. The combined scales of 3D and 2D features achieved 96.5% and 94.2% verification rates at 0.1 FAR, respectively. In contrast, the combined scales of (3D + 2D) features achieved 97.8% verification rate.

To evaluate the robustness of the proposed algorithm, we tested it on the BU-3DFE and the Bosphorus datasets. Fig. 11 shows the verification results on both BU-3DFE and Bosphorus datasets. Our algorithm achieves 98.21%, 95.46% and 99.67% verification rates at 0.1 FAR for 3D, 2D and Multimodal 3D+2D local features, respectively for all expressions (6 expressions \times 4 levels of intensity) on the BU-3DFE dataset. In the case of the Bosphorus dataset, our algorithm achieves 95.7%, 97.3% and 99% verification rates at 0.1 FAR for 3D, 2D and Multimodal 3D+2D, respectively for all expressions and pose variations. Compared to the most recent state-of-the-art works [27,36–38] on these datasets, our results on the BU-3DFE dataset are comparable to the 98.20% reported by Lei et al. [27] (3D modality), who had to rely on special masks to isolate the forehead and the nose regions. Compared to the spherical harmonic features by Liu et al. [36] using the Bosphorus dataset, our verification rate at 0.1 FAR is 2.7% higher using 3D modality and 4.3% using 3D+2D modality. Ocegueda et al. [37] achieved 96.3% and 93.8% verification rates at 0.001 FAR using the BU-3DFE and Bosphorus datasets. Ming [38] achieved 93.95% verification rates at 0.001 FAR using Bosphorus. Our algorithm achieves 95.01% and 91% verification rates at 0.001 FAR using BU-3DFE and Bosphorus datasets, respectively. However, Ocegueda's algorithm assumes that the faces are already accurately registered, and hence that the corresponding vertices between faces are known. This condition limits the practical use of their algorithms to well registered faces. Registration algorithms cannot always offer an accurate registration notably in the presence of high pose variations. In contrast, Ming [38] did not include the pose variation part (13 pose variations in Bosphorus dataset) in his experiments. Our algorithm is thus comparatively robust to high variations in pose and facial expressions.

In general, we can notice that 3D local features achieve a higher recognition rate compared to the 2D local features on both FRGC v2 and BU-3DFE datasets. This is because 3D local features are less sensitive to illumination conditions and more capable of capturing distinctive features compared to the 2D local features. However, for the Bosphorus dataset, which is geared towards facial expressions and high pose variations, we note that the performance of 2D local features outperforms the performance of 3D local features since 2D local features are less sensitive to such variations. As expected, combining 2D and 3D local features results in a superior performance under all variations (expressions, poses and illumination).

Tables 1 and 2 provide a comprehensive comparison between our algorithm and recent published algorithms tested on FRGC v2 dataset. The comparison carried was out both for identification (rank 1) and verification. Verification was done at 0.001 and 0.1 FAR for facial expressions, and at 0.1 FAR for ROCIII experiment. We used 0.001 and 0.1 since they are commonly used in the literature. Tables 1 and 2 show that our algorithm outperforms reported works based on different modalities including geometric features; global and local features and image transformations in terms of different facial expressions (neutral, non-neutral and all expressions) and different semesters (Spring2004 scans vs. Fall2003 scans or ROCIII). Note that we achieved 100% verification rates at 0.1 for both neutral and non-neutral expressions, while the recent work proposed by [7] achieved 98.5% verification rate for non-neutral expressions with 7 scans per subjects used in the gallery. In the case of ROCIII experiment, [7] achieved the highest verification rate at 0.1 FAR in the literature. However, this was obtained by registering each subject in the gallery with 7 scans. This extensive evaluation demonstrates the benefits of constructing local features in the Curvelet domain by integrating all Curvelet elements in each subband within the same scale.

7. Conclusion

In this paper, we have presented a novel Curvelet-based multi-modal algorithm for textured 3D face recognition. The proposed algorithm exploits a novel keypoint detector that can identify salient points (keypoints) on textured 3D face surfaces. The proposed keypoint identification is shown to be highly repeatable and robust because each keypoint is detected across number of frequency bands (mid-bands) and directions (angles). Highly distinctive local surface descriptors are then built around the detected keypoints by integrating all Curvelet elements in each subband within the same scale. Extensive experiments performed on the FRGC v2, BU-3DFE and Bosphorus datasets have shown that the proposed algorithm is robust and accurate for varying facial expressions, with a verification rate of 99.2%, 95.1% and 91%, respectively.

Conflict of interest

None declared.

References

- [1] W. Zhao, R. Chellappa, P.J. Phillips, A. Rosenfeld, Face recognition: a literature survey, *ACM Comput. Surv.* (2003) 399–458.
- [2] K. Bowyer, K. Chang, P. Flynn, A survey of approaches and challenges in 3d and multi-modal 3d + 2d face recognition, *Comput. Vis. Image Underst.* (2006) 1–15.

- [3] S. Berretti, A. Del Bimbo, P. Pala, 3d face recognition using isogeodesic stripes, *IEEE Trans. Pattern Anal. Mach. Intell.* 32 (12) (2012).
- [4] M. Turk, A. Pentland, Face recognition using eigenfaces, in: 1999 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 1999, pp. 586–591.
- [5] P. Belhumeur, J. Hespanha, D. Kriegman, Eigenfaces vs. fisherfaces: recognition using class specific linear projection, *IEEE Trans. Pattern Anal. Mach. Intell.* (1997) 711–720.
- [6] L. Xiaoguang, A. Jain, D. Colbry, Matching 2.5d face scans to 3d models, *IEEE Trans. Pattern Anal. Mach. Intell.* (2006) 31–43.
- [7] H. Mohammadzade, D. Hatzinakos, Iterative closest normal point for 3d face recognition, *IEEE Trans. Pattern Anal. Mach. Intell.* 35 (2) (2013) 381–397.
- [8] A. Mian, M. Bennamoun, R. Owens, Keypoint detection and local feature matching for textured 3d face recognition, *Int. J. Comput. Vis.* (2008) 1–12.
- [9] C. Zhong, Z. Sun, T. Tan, Robust 3d face recognition using learned visual codebook, in: 2007 IEEE Conference on Computer Vision and Pattern Recognition, CVPR '07, 2007 pp. 1–6.
- [10] C. Creusot, N. Pears, J. Austin, A machine-learning approach to keypoint detection and landmarking on 3d meshes, *Int. J. Comput. Vis.* (2013) 146–179.
- [11] S. Berretti, N. Werghi, A. Bimbo, P. Pala, Selecting stable keypoints and local descriptors for person identification using 3d face scans, *Vis. Comput.* (2014) 1–18.
- [12] D. Huang, G. Zhang, M. Ardabilian, Y. Wang, L. Chen, 3d face recognition using distinctiveness enhanced facial representations and local feature hybrid matching, in: IEEE International Conference on Biometrics: Theory Applications and Systems (BTAS), 2010, pp. 1–7.
- [13] X.-Y. Jing, H.-S. Wong, D. Zhang, Face recognition based on discriminant fractional fourier feature extraction, *Pattern Recognit. Lett.* 27 (13) (2006) 1465–1471.
- [14] J. Fauqueur, N. Kingsbury, R. Anderson, Multiscale keypoint detection using the dual-tree complex wavelet transform, in: IEEE International Conference on Image Processing, 2006.
- [15] P. Bendale, B. Triggs, N. Kingsbury, Multiscale keypoint analysis based on complex wavelets, in: Proceedings of the British Machine Vision Conference 2010: BMVC, BMVA Press, Aberystwyth, United Kingdom, 2010, pp. 49.1–49.10.
- [16] T. Mandal, Q. Jonathan Wu, Y. Yuan, Curvelet based face recognition via dimension reduction, *Signal Process.* (2009) 2345–2353.
- [17] M. Riza, M. El Aroussi, M. El Hassouni, S. Ghoulali, D. Aboutajdine, Local curvelet based classification using linear discriminant analysis for face recognition, *Int. J. Comput. Sci.* 4 (1) (2009).
- [18] I. Sumana, M. Islam, Z. Dengsheng, L. Guojun, Content based image retrieval using curvelet transform, in: IEEE Workshop on Multimedia Signal Processing, 2008.
- [19] J. Ma, G. Plonka, A review of curvelets and recent applications, in: IEEE Signal Processing Magazine, 2009.
- [20] G.K. Verma, S. Prasad, G. Bakul, Robust face recognition using curvelet transform, in: Proceedings of the 2011 International Conference on Communication, Computing and Security, ACM, New York, NY, USA, 2011, pp. 239–242.
- [21] S. Elaiwat, F. Boussaid, M. Bennamoun, A. El-Sallam, 3d face identification using curvelet transform, in: 2013 1st International Conference on Communications, Signal Processing, and their Applications (ICCSA), 2013, pp. 1–6.
- [22] S. Elaiwat, M. Bennamoun, F. Boussaid, A. El-Sallam, 3-d face recognition using curvelet local features, *IEEE Signal Processing Letters* (2014) 172–175.
- [23] D. Donoho, M. Duncan, Digital curvelet transform: strategy, implementation and experiments, in: Proceedings of Aerosense 2000, Wavelet Applications VII, SPIE, 1999, pp. 12–29.
- [24] E. Candès, L. Demanet, D. Donoho, L. Ying, Fast discrete curvelet transforms, *J. Multiscale Model. Simul.* (2006) 861–899.
- [25] J. Ma, G. Plonka, A review of curvelets and recent applications, *IEEE Signal Process. Mag.* 27 (2) (2010).
- [26] S. Bock, G. Widmer, Maximum filter vibrato suppression for onset detection, in: Proceedings of the 16th International Conference on Digital Audio Effects (DAFx), Maynooth, Ireland, September 2013, 2013.
- [27] Y. Lei, M. Bennamoun, A.A. El-Sallam, An efficient 3d face recognition approach based on the fusion of novel local low-level features, *Pattern Recognit.* (2013) 24–37.
- [28] F. Al-Osaimi, M. Bennamoun, A. Mian, Spatially optimized data-level fusion of texture and shape for face recognition, *IEEE Trans. Image Process.* (2012) 859–872.
- [29] D. Lowe, Distinctive image features from scale-invariant keypoints, *Int. J. Comput. Vis.* (2004) 91–110.
- [30] M. Islam, Z. Dengsheng, L. Guojun, Rotation invariant curvelet features for texture image retrieval, in: IEEE International Conference on Multimedia and Expo, 2009, ICME 2009, 2009.
- [31] P. Phillips, P. Flynn, T. Scruggs, K. Bowyer, J. Chang, K. Hoffman, J. Marques, M. Jaesik, W. Worek, Overview of the face recognition grand challenge, in: IEEE Computer Society Conference on CVPR, vol. 1, 2005.
- [32] L. Yin, X. Chen, Y. Sun, T. Worm, M. Reale, A high-resolution 3d dynamic facial expression database, in: IEEE International Conference on Automatic Face Gesture Recognition, 2008, pp. 1–6.
- [33] A. Savran, N. Alyüz, H. Dibeklioglu, O. Çeliktutan, B. Gökberk, B. Sankur, L. Akarun, Bosphorus database for 3d face analysis, in: B. Schouten, N. Juul, A. Drygajlo, M. Tistarelli (Eds.), *Biometrics and Identity Management*, vol. 5372, Springer, Berlin, Heidelberg, 2008, pp. 47–56.
- [34] A. Mian, M. Bennamoun, R. Owens, An efficient multimodal 2d–3d hybrid approach to automatic face recognition, *IEEE Trans. Pattern Anal. Mach. Intell.* (2007) 1927–1943.
- [35] B. Li, A. Mian, W. Liu, A. Krishna, Using kinect for face recognition under varying poses, expressions, illumination and disguise, in: IEEE Workshop on Applications of Computer Vision, 2013, pp. 186–192.
- [36] P. Liu, Y. Wang, D. Huang, Z. Zhang, L. Chen, Learning the spherical harmonic features for 3-d face recognition, *IEEE Trans. Image Process.* (2013) 914–925.
- [37] O. Ocegueda, T. Fang, S. Shah, I. Kakadiaris, 3d face discriminant analysis using Gauss–Markov posterior marginals, *IEEE Trans. Pattern Anal. Mach. Intell.* 35 (3) (2013) 728–739.
- [38] Y. Ming, Rigid-area orthogonal spectral regression for efficient 3d face recognition, *Neurocomputing* (2014) 445–457.
- [39] H. Drira, B. Ben Amor, A. Srivastava, M. Daoudi, R. Slama, 3d face recognition under expressions, occlusions, and pose variations, *IEEE Trans. Pattern Anal. Mach. Intell.* (2013) 2270–2283.
- [40] Y. Ming, Q. Ruan, Robust sparse bounding sphere for 3d face recognition, *Image Vis. Comput.* 30 (8) (2012).
- [41] C. Queirolo, L. Silva, O.R.P. Bellon, M. Pamplona Segundo, 3d face recognition using simulated annealing and the surface interpenetration measure, *IEEE Trans. Pattern Anal. Mach. Intell.* 32 (2) (2010) 206–219.
- [42] C. Xu, S. Li, T. Tan, L. Quan, Automatic 3d face recognition from depth and intensity gabor features, *Pattern Recognit.* (2009) 1895–1905.

Said Elaiwat received the B.Sc. degree in computer science from AL-Zaytoonah Private University of Jordan, Amman, Jordan, in 2004 and the M.Sc. degrees in computer science from Al-Balqa Applied University, Amman, Jordan, in 2007. He is currently a Ph.D. student in the School of Computer Science and Software Engineering, The University of Western Australia, Crawley, Australia. His current research interests include object/face recognition, image processing, computer vision and machine learning.

Mohammed Bennamoun received the M.Sc. degree in control theory from Queens University, Kingston, ON, Canada, and the Ph.D. degree in computer vision from Queens/Q.U.T, Brisbane, Australia. He was a Lecturer in robotics with Queens University and joined QUT, in 1993 as an Associate Lecturer. He is currently a Winthrop Professor. He served as the Head of the School of Computer Science and Software Engineering, The University of Western Australia, Crawley, Australia, from 2007 to 2012. He served as the Director of the University Centre at QUT: The Space Centre for Satellite Navigation from 1998 to 2002. He was an Erasmus Mundus Scholar and a Visiting Professor with the University of Edinburgh, Edinburgh, U.K., in 2006. He was a Visiting Professor with the Centre National de la Recherche Scientifique and Telecom Lille1, France, in 2009, Helsinki University of Technology, Helsinki, France, in 2006, and University of Bourgogne and Paris 13, Paris, France, from 2002 to 2003. He is the co-author of *Object Recognition: fundamentals and Case Studies* (Springer-Verlag, 2001) and the co-author of an edited book on *Ontology Learning and Knowledge Discovery Using the Web* in 2011. He has published over 250 journal and conference publications and secured highly competitive national grants from the Australian Research Council (ARC). Some of these grants were in collaboration with Industry partners (through the ARC Linkage Project scheme) to solve real research problems for industry, including Swimming Australia, the West Australian Institute of Sport, a textile company (Beaulieu Pacic), and AAM-GeoScan. He has worked on research problems and collaborated (through joint publications, grants, and supervision of Ph.D. students) with researchers from different disciplines, including animal biology, speech processing, biomechanics, ophthalmology, dentistry, linguistics, robotics, photogrammetry, and radiology. He received the Best Supervisor of the Year Award from QUT. He received an award for research supervision from UWA, in 2008. He served as a Guest Editor for a couple of special issues in international journals, such as the *International Journal of Pattern Recognition and Artificial Intelligence*. He was selected to give conference tutorials from the European Conference on Computer Vision and the International Conference on Acoustics Speech and Signal Processing. He has organized several special sessions for conferences, including a special session for the IEEE International Conference in Image Processing. He was on the program committee of many international conferences. He has contributed in the organization of many local and international conferences. His current research interests include control theory, robotics, obstacle avoidance, object recognition, artificial neural networks, signal/image processing, and computer vision (particularly 3D).

Farid Boussaid (M00SM04) received the M.S. and Ph.D. degrees in microelectronics from the National Institute of Applied Science (INSA), Toulouse, France, in 1996 and 1999, respectively. He joined Edith Cowan University, Perth, Australia, as a Postdoctoral Research Fellow, and a member of the Visual Information Processing Research Group, in 2000. He joined the University of Western Australia, Crawley, Australia, in 2005, where he is currently an Associate Professor. His current research interests include smart CMOS vision sensors, image processing, gas sensors, neuromorphic systems, device simulation, modeling, and characterization in deep submicron CMOS processes.

Amar A. El-Sallam received B.Sc. and M.Sc. in EEng from Assiut University and a Ph.D. from Curtin University, Australia. He received several scholarships including the highly competitive and prestigious IPRS, the Australian Telecommunication Cooperative Research Centre (ATCRC) and the Australian Telecommunication Research Institute (ATRI/WATRI) scholarships. From 2006 to 2009, Amar was a Postdoctoral Research Fellow with Microelectronics Research Group (MRG), at the school of EEng at the University of Western Australia (UWA). Currently, he is a Research Assistant Professor at the School of Computer Science and Software Engineering and the School of Sport Science, Exercise and Health at UWA. His current research covers various disciplines including computer vision, biometrics, signal and image processing and wireless communications, sport science. Amar has over 50 refereed scientific publications, 2 theses, and been an active reviewer for many of the major publishers e.g. IEEE and Elsevier. He secured and worked in a number of major research grants, e.g. with Swimming Australia, Australian Institute of Sport and the Australian Ministry of Health dealing with the treatment of Parkinson disease using a novel clinical system.