Pattern Recognition 48 (2015) 1235–1246

Contents lists available at ScienceDirect

Pattern Recognition

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

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 repeatably 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.

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 app roaches 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).

http://dx.doi.org/10.1016/j.patcog.2014.10.013
0031-3203/

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