Neurocomputing 184 (2016) 117–130

Contents lists available at ScienceDirect

Neurocomputing

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

# A unified 3D face authentication framework based on robust local mesh SIFT feature ☆

Yue Ming [a,*], Xiaopeng Hong [b]

[a] Beijing Key Laboratory of Work Safety Intelligent Monitoring, School of Electronic Engineering, Beijing University of Posts and Telecommunications, Beijing 100876, PR China
[b] Department of Computer Science and Engineering, University of Oulu, Finland

ARTICLE INFO

*Article history:*
Received 2 January 2015
Received in revised form
24 May 2015
Accepted 18 July 2015
Available online 11 December 2015

*Keywords:*
3D face authentication
Depth estimation
Facial region segmentation
Local Mesh Scale-Invariant Feature Transform (LMSIFT)
Interactive education platform

ABSTRACT

In this paper, we design a unified 3D face authentication system for practical use. First, we propose a facial depth recovery method to construct a facial depth map from stereoscopic videos. It effectively utilize prior facial information and incorporate the visibility term to classify static and dynamic pixels for robust depth estimation. Secondly, in order to make 3D face authentication more accurate and consistent, we present an intrinsic scale feature detection for interesting points on 3D facial mesh regions. Then, a novel feature descriptor is proposed, called Local Mesh Scale-Invariant Feature Transform (LMSIFT) to reflect the different face recognition abilities in different facial regions. Finally, the sparse optimization problem of visual codebook is used to 3D face learning. We evaluate our approach on publicly available 3D face databases and self-collected realistic scene databases. We also develop an interactive education system to investigate its performance in practice, which demonstrates the high performance of the proposed approach for accurate 3D face authentication. Compared with previous popular approaches, our system has consistently better performance in terms of effectiveness, robustness and universality.



## 1. Introduction

Face authentication and recognition has been a hot topic in computer vision field for decades. It has a wide variety of applications, for example surveillance, automated screening, and human–computer interaction, since human face images are easy-to-access, and the acquisition of human faces are usually non-intrusive and user-friendly. Over the past several decades considerable efforts have been devoted to 2D face authentication and recognition. As a result, accuracy for 2D face authentication and recognition has been substantially improved [1–6]. However, 2D face authentication and recognition are still difficult because of the inherent flaws in handling extreme illumination, pose variations, complex backgrounds and Intra-subject deformation [7].

Due to the deficiency of using only the intensity as the 2D face authentication and recognition algorithms do, the lowered cost of 3D acquisition devices, such as Kinect, SR4000 and 3D scanners, makes a strong push for performance improvement by adding 3D depth information [8]. 3D images can capture different transformations and actual facial anatomical structure. The distinctive advantages of 3D face authentication have improved the effectiveness of recognition [9].

Though great strides have been made in 3D face authentication, it is still challenging to obtain reliability in 3D face identification and verification, particularly in unconstrained "in the wild" scenarios. Especially, accurate depth information recovery from binocular images and effective 3D facial representations are two of the crucial but unsolved issues.

Firstly, traditional binocular vision based depth information recovery methods only depend on the geometric relationship between the two parallel images from the camera and do not incorporate the reference model of the target, therefore it can only coarsely reconstruct the 3D information (e.g. for large and uniform scenes). However, they have difficulties in building a fine depth image in which the detail of scenes, like the slight depth variations on a human face, can be well retained.

Secondly, traditional 3D face descriptors, including depth map [10], spin image [11], meshSIFT [12], etc., are usually based on the assumption that the sampling of 3D face surface is nearly uniform or made uniform by resampling. However, it is extremely difficult for this assumption to be satisfied in real-world since occlusion

☆The work presented in this paper was supported by the National Natural Science Foundation of China (Grants no. NSFC-61402046), President Funding of Beijing University of Posts and Telecommunications.

* Corresponding author.
*E-mail addresses:* myname35875235@126.com (Y. Ming), xhong@ee.oulu.fi (X. Hong).

http://dx.doi.org/10.1016/j.neucom.2015.07.127
0925-2312/

and pose variations may result in loss of facial local information. As a result, 3D information is inevitably lost. Moreover, they usually suffer the problem of high computation complexity.

Concentrating on the major challenges in 3D face generation and facial representations, we propose a unified 3D face authentication framework based on stereoscopic generation. More specially, we propose a facial depth recovery approach for robust depth estimation, and a novel scale and rotation invariant feature descriptor for 3D face description to effectively overcome the obstacles of computation expense and complex backgrounds. Finally, we apply the proposed system to a real-world, interactive education environment to conveniently identify students, thereby providing the benefit of personalized service.

The main contributions of this paper are summarized in the following items:

1. *Reliability*: We propose a reliable facial depth recovery technique, which is adapted to estimate the 3D face shape and geometric information. Different from previous research, our method is based on prior facial information, and incorporates the visibility term with static and dynamic segmentation for robust dense depth estimation. An accurate, refined model has significant advantages in handling occlusions and complex backgrounds.
2. *Effectiveness*: We derive a descriptor named Local Mesh Scale-Invariant Feature Transform (LMSIFT) for 3D face authentication. LMSIFT is able to transform the depth and gray information into a one-dimensional vector without the too strong assumption of uniform density. Compared with commonly-used 3D facial descriptors, our representation has better discrimination, and easily handles scale-rotation changes and partial mesh matching.
3. *Universality*: The performance improvement of our 3D face authentication system which was demonstrated on realistic scenes, where the 3D facial data was collected at various times, different environments, and from individuals from different countries, and with a large range of ages, poses, occlusions, and complex backgrounds.

The rest of this paper is organized as follows. In Section 2 we survey related work on face authentication using local features in mesh and 3D face data for authentication. Section 3 introduces our proposed framework. In Section 4 consistent facial depth recovery from a binocular framework is described. In Section 5 our proposed feature detection and descriptor scheme is presented. In Section 6 we derive the learning scheme for the local mesh scale-rotation invariant descriptor. In Section 7 we verify the use of our approach and apply our system to an interactive education system. In Section 8 we conclude the paper.

## 2. Related works

3D data processing, especially 3D data generation and learning, has been a long continuously challenging topic in computer vision and graphics due to its widely applications in for example 3D movies (e.g. Avatar) and 3DTV [13].

The problem of analyzing and recognizing faces is an important branch of 3D data processing. The actual facial anatomical structure is usually recovered by stereo matching for compensating data missing during 2D projection and overcoming the difficulty of 3D affordable acquisition devices. In realistic situations, the 3D face recognition method can match partial scans, especially for analyzing students' states in the interactive learning system. From the perspective of theoretical analysis, in 2005, Phillips et al. [14] first compared the recognition performance between 2D and 3D face data, which demonstrated that 3D images can preserve more discriminative information.

### 2.1. Depth information recovery from binocular images

There are two kinds of devices for depth information acquisition: the depth camera and the stereo camera. The most representative product of the depth camera might be the Kinect developed by Microsoft Corp. Kinect measures the distance in depth by infrared, which has the advantage in real-time situation. However, it is influenced by sunshine easily so that its detective range is limited. In contrast, the stereo camera system collects the depth information through building the matching relationship between the pixels in a pair of images obtained by two cameras. It has much more immunity to sunshine than the early one and thus results in a larger range of detective distance. With the development of technology, the precision of the stereo camera based approaches namely the stereo matching algorithm are getting much higher. As a result, they become a popular method to collect depth information [15,16].

The researches in the US and Europe started the early research of stereo matching algorithm [17]. Scharstein and Szeliski reorganized the past researches in 2002, presented a completed standard of stereo matching algorithms, which is still being used nowadays, and maintain a website for stereo matching algorithm evaluation [17]. Stereo matching algorithm can be roughly divided into the global algorithms and the local ones. The global stereo matching algorithms show high performance but their running costs are usually high. On the contrary, the local algorithm computes efficiently but its performance was inferior. To improve its performance, Yoon and Kweon proposed an adaptive weighting scheme [18]. In 2008, the mutual information was introduced by Hirschmuller, to refine the local algorithm [19]. More recently, the researches on stereo matching gradually blur the concept of local algorithm and global algorithm, and combine them to achieve a semi-local-semi-global stereo matching scheme. Moreover, parallel stereo matching algorithms are well investigated in [20,21].

However, the above-mentioned methods do not incorporate the reference model of the target, therefore it can only coarsely reconstruct the 3D information (e.g. for large and uniform scenes), which is difficult to reflect fine depth variations (e.g. for human face).

### 2.2. 3D facial representations

Because of the aforementioned advantages, also with the advent of new capture devices, the 3D face recognition has triggered increased interest. Feature representations have also been developed for 3D facial images [8]. Huang et al. [10] proposed multiscale extended Local Binary Patterns (eLBP) to represent the facial geometry. SIFT-based matching scheme combined local and holistic analysis and proved robust to expressions, occlusions, and pose variations. Spherical harmonic features (SHF) was presented by Liu et al. [11], which demonstrated the outstanding advantages for representing gross shape and fine surface details on 3D face surface. Our previous work [22] proposed Sparse Bounding Sphere for 3D Face Recognition, which demonstrated the excellent performance for pose variations. Although facial surface features can efficiently describe facial geometric structures, they unfortunately lost the regional discrimination. Our past researches have proved effectiveness of the rigid area for 3D face recognition [23].

Darom and Keller [24] applied the popular SIFT feature to the mesh domain to capture local regions of interests in different manifestations over a similar support. Maes et al. [25] presented the meshSIFT detector that detects the local feature as scale space extreme for 3D face surface. Smeets further extended the meshSIFT to the symmetric surface-feature [12,26,27], which shows promising