ARTICLE IN PRESS
JID: NEUCOM [m5G;April 20, 2017;8:47]

Neurocomputing 000 (2017) 1–7

Contents lists available at ScienceDirect

Neurocomputing

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

Real-time 3D face modeling based on 3D face imaging

Shu Zhan[a,*], Lele Chang[a], Jingjing Zhao[a], Toru Kurihara[b], Hao Du[a], Yucheng Tang[a], Jun Cheng[c]

[a] Hefei University of Technology, Hefei, 230009, China
[b] Kochi University of Technology, Kochi, 782–8502, Japan
[c] Shenzhen Institues of Advanced Technology, Chinese Academy of Sciences, Shenzhen 518055, China

ARTICLE INFO

Article history:
Received 16 September 2015
Revised 26 September 2016
Accepted 17 October 2016
Available online xxx

Keywords:
3D face modeling
Fast segmentation algorithm
Head area segmentation
Sparse iterative closest point(SICP)
Depth map

ABSTRACT

Traditional Iterative Closest Point (ICP) can not properly process the noise, outliers and missing data in face imaging, which would result in low accuracy of face image, face image registration error and much more noise in face image, to solve the above problems, an enhanced sparse ICP to register the 3D point clouds in face imaging is proposed. Sparse Iterative Closest Point (SICP) addressed these problems by formulating the registration optimization, which used sparsity inducing norms, moreover, a fast segmentation algorithm for head area segmentation in depth image was proposed. Based on the proposed fast segmentation algorithm and sparse ICP, a new real time 3D face modeling system was set up, which could generate real time 3D face models with high quality by using a depth camera (such as Kinect) even the background of face imaging was complicated.

1. Introduction

Face modeling has become an active research topic and has received much more attention in various fields such as computer graphics, medical applications, model-based facial analysis, etc., a complete overview can be found in Parke and Water's book [1], several successful applications in animation, video telephone and face recognition could be found in literature [2–11]. According to the ways of obtaining the information of depth image, existing 3D face imaging technology can be divided into two categories: passive technology and active technology. Passive technology calculates the 3D coordinate information of the face in the picture according to a series of algorithms, which is difficult to be real-time effect [12–17]. Active modeling techniques typically use laser range scanning or structured light projection. xages, triangulation is used to determine the 3D geometry of the object. In a structured light system, a pattern of light is projected onto the object. A camera captures the illuminated surface, and uses the deformation of the light pattern to determine the depth of the object [18–20]. Active technical equipments contain a light source to measure the depth information of the object directly, which can achieve the real-time effect, because the structured light technique is much cheaper and has better real-time performance.

Meyer et al. [21] propose a 3D face modeling system that utilizes the depth i mages to create a 3D face model. They put forward a head segmentation algorithm which can segment the head area quickly. This method owns the advantages of cheap instruments and real-time imaging. But sometimes, this method can not segment head completely because the depth image acquired by Kinect often lose useful information (Fig 1). M. Hernandez et al. [22] propose a system that generates a 3D face model from a sequence of depth images. They use a face detection algorithm to segment the user's head from the depth images that is captured by Kinect camera, then the segmented depth images are registered and integrated into a 3D model. This method accumulates the registered depth images in an unwrapped cylindrical 2D image which allows them to use 2D spatial filters to remove noise, but this method is a little complex and time-consuming. [23] uses the Kinect camera's color sensor to detect facial features such as the eyes and nose which can be used to align a generic face model to a depth image. Afterwards, they deform the generic model to fit the depth image. [24] puts forward a 3D face modeling system that builds a 3D face model by integrating many depth images. They use a real-time head pose estimation to give a new initial assumption to the Iterative Closest Point (ICP) algorithm when it fails. However, this method is not appropriate when the head pose has larger variation on angle rotation even if the initial assumption has been given according to the head pose estimation.

Registration is one of the most important steps in 3D face modeling, the Iterative Closest Point (ICP) algorithm is often used to

* Corresponding author.
E-mail address: shu_zhan@hfut.edu.cn (S. Zhan).

http://dx.doi.org/10.1016/j.neucom.2016.10.091
0925-2312/© 2017 Published by Elsevier B.V.

Fig. 1. The general process of face segmentation.

register the 3D point clouds. ICP algorithm is firstly proposed by Besl et al [25], which has been used widely in many fields such as robot navigation, surface reconstruction, shape registration and face recognition. Fast ICP [26] is the improvement and extension of traditional ICP algorithm which is raised by Rusinkiewicz et al. There are six steps with corresponding performance and results in Fast ICP algorithm on the basis of these factors. Fast ICP analyzes the impact of various factors or strategies in different stages on the entire ICP performance and results, which can be used to select a different algorithm for the registration of point sets or surfaces and thus achieve better registration. Because the images acquired by Kinect lack much information, users are usually asked to change their poses in front of the camera to get captured the depth images [27–34] and then, a complete 3D face model is built. However the large angle rotation of head will result in wrong registration. What is more, twisting head in front of the camera will make the adjacent two frames have overlapping area, which will affect selecting the corresponding points. Traditional ICP [22] algorithm and fast ICP algorithm [25,26] are used to register the 3D points respectively, which may outcome 3D face model with low quality because the depth images obtained by Kinect have noise, outliers and even miss data. Sparse Iterative Closest Point (SICP)[35] algorithm is adopted to register the point cloud to handle these weaknesses by formulating the registration optimization using sparse inducing norms. In the stage of image preprocessing we use the bilateral filter to get smooth images [36]. Experiment results show that we can obtain a 3D face model with better quality using sparse ICP algorithm in point cloud registration.

2. Fast face segmentation

Firstly, a raw depth image D_τ could be acquired at time τ which consists of a set of pixels(u, v) and each pixel contains a depth measurement $D_\tau(u, v)$ in millimeters.Then we segment the depth image into foreground and background regions, where the foreground contains object, and the background is the environment. Connected component analysis is used to determine the foreground region. Two neighboring depth pixels are considered being connected when their depth difference is under a threshold. Foreground region is considered as a large area with the shortest distance from the camera. A method is developed to separate the head from the foreground region quickly in the depth image. We need to find a horizontal scan line that separates the foreground into head and torso regions. In order to accomplish this task, we generate a row histogram by

$$h(v) = \sum_u M_\tau(u, v) \tag{1}$$

Here, M_τ is the foreground mask, and each bin $h(v)$ contains the width of the foreground region in row v. The bins in the row histogram can be split into sets H and J

$$H = \{v | v \le y \ and \ h(v) \ne 0\} \tag{2}$$

$$J = \{v | v > y \ and \ h(v) \ne 0\} \tag{3}$$

Here, y is the horizontal scan line to separate the binary codes. The image captured by Kinect contains both depth information and color information, and we can detect face [37] in the scene by utilizing the color information. Face detection is the first step in 3D face modeling process. We denote the height of the head as y and classify rows within head region and torso region into set H and J set, respectively(Fig 2).

A segmented depth image D'_τ,

$$D'_\tau(u, v) = \begin{cases} D_\tau(u, v) & iff \quad M_\tau(u, v) = 1 \ and \ v \le y \\ 0 & otherwise \end{cases} \tag{4}$$

Because the Kinect could emphasis on the performance and efficiency when extracting the depth image, the quality of the extracted depth image is poor and ”user segmentation” often has jagged [38]. In order to ensure the subsequent application, it is necessary to conduct the depth image edge detection and noise reduction [39].We eliminate depth measurements with a high local variance $\delta(u, v)^2$

$$D''_\tau(u, v) = \begin{cases} D'_\tau(u, v) & iff \quad \delta(u, v)^2 < \delta^2_{max} \\ 0 & otherwise \end{cases} \tag{5}$$

where δ^2_{max} is the threshold of pixel's local variance. In addition, we used the bilateral filter [40] to reduce noise when preserving depth discontinuity. To visualize the differences, we displayed the normal maps corresponding to the raw measurements (a), the measurements filtered by Gaussian filter (b), the measurements filtered by Bilateral filter (Fig 3). In Fig 3, we can see that the measurements filtered by Bilateral filter is better than the other two.

3. Sparse ICP

Sparse ICP algorithm can address these problems of noise, outliers and missing data appeared in image registration, which uses sparse inducing norms to formulate the registration optimization [35,41–43] .

Given two surfaces X, Y embedded in a k-dimensional space, we formulate the pairwise registration problem as

$$\arg\min_{R,t} \int_X \varphi(Rx + t, \psi) dx + I_{so(k)}(R), \tag{6}$$

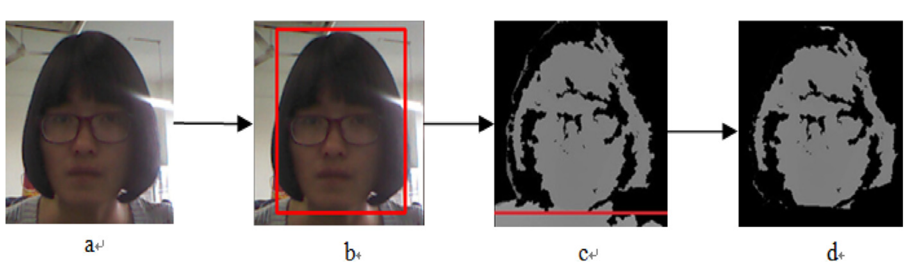

Fig. 2. The general process of face segmentation. (a) the face image of RGB (b)the label of the user's head height (c) find the horizontal scan line in depth image (d) the depth image after segmenting the user's head.