Measurement 131 (2019) 590–596

Contents lists available at ScienceDirect

Measurement

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

Automatic reconstruction method for large scene based on multi-site point cloud stitching

Haonan Xu [a], Lei Yu [a,*], Junyi Hou [a], Shumin Fei [b]

[a] School of Mechanical and Electric Engineering, Soochow University, Suzhou, China
[b] School of Automation, Southeast University, Nanjing, China

ARTICLE INFO

Article history:
Received 14 November 2017
Received in revised form 3 September 2018
Accepted 7 September 2018
Available online 10 September 2018

Keywords:
Large scene three-dimensional reconstruction
Kinect sensor
Bundle adjustment
Point cloud registration
Automatic reconstruction method

ABSTRACT

At present, the three-dimensional (3D) reconstruction system is based on hand-held camera which causes certain problems, such as needing for a large number of human-computer interactions, demanding on high quality image data and insufficient accuracy of 3D reconstruction. Aiming at these problems, a fully automatic reconstruction of large scene based on multi-site point cloud stitching is presented in this paper. The proposed method uses the Kinect sensor for image acquisition. Since, there are multiple sites in the room, each site of image data is processed using a separate model in order to get good 3D point cloud data. Then, these sites are used to constitute a local area network, and the method of bundle block adjustment is employed to stitch each site point cloud data. The proposed method achieves a high-degree automation and provides a high-precision 3D reconstruction which has two main advantages: (i) the reconstruction process is fully automatic, without any human-computer interaction; (ii) the automatic reconstruction is robust. Experimental results show that proposed automatic reconstruction method is convenient and practical, and can provide better 3D reconstruction model than commonly used methods. Moreover, it can be applied to virtual reality shopping malls, Virtual Reality (VR) and other fields.

1. Introduction

The three-dimensional reconstruction and display of scenes are widely used in many areas, such as design simulation, virtual reality and human-computer interaction. The existing three-dimensional reconstruction technology is mainly based on use of hand-held camera to obtain data for three-dimensional reconstruction. However, this method is not easy to operate, and quality of obtained image data and effect of three-dimensional reconstruction are not satisfactory. If a three-dimensional reconstruction model for photos and video is easy to obtain and works well, then that model could be applied to many areas, such as business web site platform, virtual reality shopping malls and other shopping malls. Therefore, the achievement of a fully automated reconstruction and an ideal three-dimensional reconstruction model is very important goal of computer vision.

The current automated reconstruction system is mainly based on hand-held camera, and it provides good results in reconstruction of small scenes, but with slower and large human-computer interaction characteristics [1–8,23]. Henry et al. used the Kinect to implement an interactive 3D reconstruction system that selects only key frames for iterative closest point (ICP) cloud registration [2]. They designed a laboratory-wide three-dimensional model, but the model is rough and not complete enough, thus further improvement of accuracy and algorithm performance is needed. ShahramIzadi et al. constructed the Kinect-Fusion system, using Kinect for three-dimensional reconstruction of small-scale scenes and individual objects [3]. Compared to method presented by Henry et al., the results are more elaborate, but because almost all operations are done in the display and limited size of memory, the rebuild range and reconstruction accuracy cannot be good simultaneously, which makes reconstruction difficult. In view of the shortage of Kinect-Fusion system, Raul Mur-Artal et al. proposed a real-time reconstruction method named the simultaneous localization and mapping based on oriented brief feature recognition (ORB-SLAM) which uses parallel tracking and mapping (PTAM) architecture [4]. The ORB-SLAM adds the function of map initialization and closed-loop detection, and optimizes the method of key frame selection and map construction. It achieves good results in processing speed, tracking effect and map precision. However, ORB-SLAM still has some shortcomings [9–13]. The whole SLAM system is calculated by feature points, therefore it is necessary to determine the ORB features for each image, which is

---

* Corresponding author.
E-mail address: slender2008@163.com (L. Yu).

https://doi.org/10.1016/j.measurement.2018.09.022
0263-2241/

time-consuming while the map construction is sparse [14,19]. The most of three-dimensional reconstruction methods for indoor scenes need a hand-held camera for shooting, moreover the operation is not convenient. The amount of human-computer interaction is large when a hand-held camera is used to acquire data. The quality of the data acquired by the camera will directly affect the results of the 3D reconstruction. The model obtained by the method of holding the camera for 3D reconstruction is not ideal.

As already mentioned, in current three-dimensional reconstruction of large scenes, the degree of automation is low and the reconstruction model is not ideal. This paper presents a fully automated reconstruction system for multi-site point cloud stitching intended for large scenes. The main feature of this work include: (i) This paper presents a fully automated data acquisition platform, and uses advanced calibrated camera and rotating head for image acquisition. (ii) In the room with multiple sites, each site of image data is processed using an independent adjustment method to obtain a three-dimensional point cloud data. (iii) Each site constitutes a local area network and the bundle block adjustment method is used for each site point cloud data to get a three-dimensional reconstruction model and to achieve full automation.

The outline of this paper is as follows. In Section 2, the mathematical model of point cloud stitching and optimization is introduced, including mathematical model of single point cloud, multiple point cloud stitching and bundle block adjustment. Section 3 is devoted to the solution of mathematical model of cloud stitching and its optimization. Section 4 includes simulation results to illustrate feasibility and effectiveness of the proposed strategy. Finally, we give a conclusion in Section 5.

2. Mathematical model of point cloud stitching and optimization

2.1. Mathematical model of single point cloud

In order to obtain the three-dimensional reconstruction model of a large scene, this paper uses a certain number of sites to form images, and then splices them into site point cloud [24–25]. In a single site point cloud, this paper uses the method of rotating head. The camera placed on the rotating head rotates with a step of 10° to collect images, thus 36 pairs of color images and depth images are obtained. The data obtained by the scan are not in the same coordinate system, thus they need to be converted to the same coordinate system. In order to unify the data of each group, the space coordinate of the n^{th} coordinate system is switched to the first coordinate system, and it needs to go through (n-1) rotation and translation, but the excessive number of rotations and translations causes the accumulative error. In order to get a good model, the error should be evenly distributed, so the independent adjustment should be used [16–17,20–22].

The adjustment is based on determination of parameters of adjacent model, achievement of seven spatial similarities of parameters of three corner elements of exterior orientation (angle of rotation of the X, Y, and Z coordinate axes), namely Φ,Ω, and K, three translations and a scale of the zoom:

$$\begin{bmatrix} X_1 \\ Y_1 \\ Z_1 \end{bmatrix} = \lambda \begin{bmatrix} a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \\ c_1 & c_2 & c_3 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} + \begin{bmatrix} \Delta X \\ \Delta Y \\ \Delta Z \end{bmatrix} \tag{1}$$

where, a_1,b_1,c_1,a_2,b_2,c_2,a_3,b_3, and c_3 are the direction cosines of the function of corner elementsΦ, Ω, and$K$; ΔX, ΔY, and ΔZ are the translations of the coordinate origin in three coordinate directions; and λ is the scaling factor. Then, we introduce the corrections of X, Y and Z, and coordinate each point of gravity. Therefore, the error equation is defined by:

$$\begin{bmatrix} \mathrm{v_x} \\ \mathrm{v_y} \\ v_z \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & \bar{X} & -\bar{Z} & 0 & -\bar{Y} \\ 0 & 1 & 0 & \bar{Y} & 0 & -\bar{Z} & \bar{X} \\ 0 & 0 & 1 & \bar{Z} & \bar{X} & \bar{Y} & 0 \end{bmatrix} \begin{bmatrix} \mathrm{d}\Delta X \\ d\Delta Y \\ d\Delta Z \\ d\lambda \\ d\Phi \\ d\Omega \\ dK \end{bmatrix} - \begin{bmatrix} I_X \\ I_Y \\ I_Z \end{bmatrix} \tag{2}$$

where, (v_x, v_y, v_z) is the correction of the observed value of (X, Y, Z); $d\Delta X$, $d\Delta Y$, $d\Delta Z$, $d\lambda$, $d\Phi$, $d\Omega$, and $d\mathrm{K}$ are the corrections of the myopic value of parameter to be determined; $\bar{X}$, $\bar{Y}$, and $\bar{Z}$ are the coordinates of gravity center of spatial point in coordinate system to be rotated, and I_X, I_Y, and $I_Z$are the constant terms of error equation.

$$\begin{matrix} I_X = X_1 - \Delta X - \lambda X' \\ I_Y = Y_1 - \Delta Y - \lambda Y' \\ I_Z = Z_1 - \Delta Z - \lambda Z' \end{matrix} \begin{bmatrix} X' \\ Y' \\ Z' \end{bmatrix} = \begin{bmatrix} a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \\ c_1 & c_2 & c_3 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} \tag{3}$$

Then, (1) and (2) are combined to determine two adjacent images of transformation parameters. Using the (m-2)$^{\text{th}}$ picture that corresponds to the camera coordinate system, we convert the (m-1)$^{\text{th}}$ picture and the m^{th} picture to the point of the (m-1)$^{\text{th}}$ picture to the camera coordinate system that corresponds to the (m-2)$^{\text{th}}$ picture.

2.2. Mathematical model of multiple point cloud stitching

Once we have obtained the 3D model data of one site, we can use the same method to get the data of the second site. The stitching of sites also uses coordinate transformation to unify the data, i.e. to translate the data into the same coordinate system. But due to the site increase, the error gradually accumulates, the reconstructed 3D model drifts and so on. In this paper, The idea of photogrammetry is applied. The site layout is the same as in the navigation area, and the model is optimized by constructing the bundle block adjustment method.

The relationship between two adjacent sites based on the three-dimensional coordinate conversion model is as follows:

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} X_S \\ Y_S \\ Z_S \end{bmatrix} + \lambda R \begin{bmatrix} x \\ y \\ z \end{bmatrix} \tag{4}$$

where, X_S, $Y_S$and Z_S represent the translation parameters, which represent the three-dimensional coordinates of the scanning site translated into the reference coordinates of location, and they are the point cloud positioning parameters. And R matrix reflects the automatic orientation of the point cloud. According to the nature of Rodriguez matrix, the i^{th} site is directed to its adjacent site h with an unknown number of adjustment equations. Based on (4), we get:

$$F = \left(I - \hat{S}_{\mathrm{i}}\right) \begin{bmatrix} \hat{X}_{\mathrm{h}} \\ \hat{Y}_{\mathrm{h}} \\ \hat{Z}_h \end{bmatrix} - \left(I - \hat{S}_{\mathrm{i}}\right) \begin{bmatrix} \hat{X}_{Si} \\ \hat{Y}_{Si} \\ \hat{Z}_{Si} \end{bmatrix} - \lambda\left(I + \hat{S}_{\mathrm{i}}\right) \begin{bmatrix} \hat{x}_{ih} \\ \hat{y}_{ih} \\ \hat{z}_{ih} \end{bmatrix} = 0 \tag{5}$$

where, I is the third order unit matrix,$\hat{S}_{\mathrm{i}}$ is the adjustment value of the antisymmetric matrix of the i^{th} site, $\hat{X}_{\mathrm{h}}$, $\hat{Y}_{\mathrm{h}}$, and $\hat{Z}_{\mathrm{h}}$ are the measured coordinates of the site h, $\hat{X}_{\mathrm{Si}}$, $\hat{Y}_{\mathrm{Si}}$, and $\hat{Z}_{\mathrm{Si}}$ are the coordinates of the i^{th} site scanning position, $\hat{x}_{ih}$, $\hat{y}_{ih}$, and $\hat{z}_{ih}$ represent the adjusted scan coordinates of the i^{th} site; and λ is the scaling factor.

Since the above formula can be obtained for two adjacent sites point cloud stitching, then we can use the same method to get multi-point site cloud stitching, such that between multiple sites a local area network is constituted. And the bundle block adjustment is used to deal with each site point cloud stitching.