Research on 3D reconstruction technology based on Kinect

Haojie Duan*

Key Laboratory of Metallurgical Equipment and Control Technology, Ministry of Education, Wuhan University of Science and Technology, Wuhan 430081, China

*13659885455@163.com

Keywords: 3D reconstruction; Kinect; 3D point cloud; ICP algorithm

Abstract: With the development of machine vision technology, 3D reconstruction technology has been widely used. Aiming at the shortcomings of 3D laser scanning imaging technology, this paper proposes a method of acquiring depth image by kinect, denoising the depth image, preprocessing the depth image, acquiring 3D point cloud, and completing the 3D reconstruction of the object by ICP algorithm.

1. Introduction

With the development of machine vision theory and the advancement of hardware technology, 3D reconstruction technology has been widely used in modern industry [1], film and television industry, human-computer interaction and other fields. In the industry, its applications mainly cover both digital-analog detection and digital model reconstruction [2]. In the fields of clothing, art, archaeology, etc., 3D reconstruction is the basic method to convert the scanning point cloud into a real model for rapid digitization [3]. At present, there are excellent ones: three-dimensional reconstruction based on laser scanners and depth scanners. It uses 3D laser scanning technology for 3D modeling. However, the use of scanning instrument three-dimensional reconstruction, although very mature and effective, but the equipment is very expensive, the amount of data is too large and is not conducive to the processing of subsequent data [4]. The depth image and color image collected by Kinect are studied, combined with the related image processing technology, and the solid image is reconstructed by using the depth image and the color image fusion [5]. Compared with the three-dimensional laser scanning technology, the method is simple in operation, inexpensive in equipment and easy to process data.

2. Kinect image processing

Figure 1 is a three-dimensional reconstruction process.

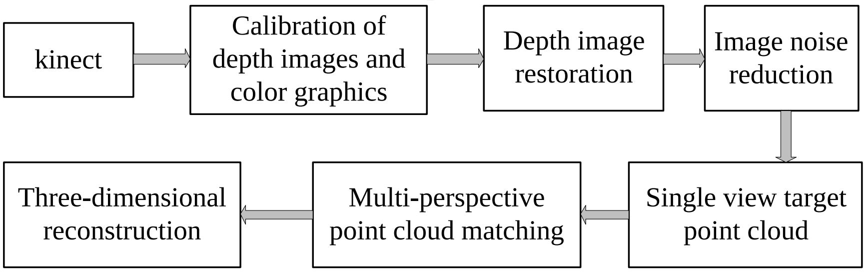

Figure 1 Kinect 3d surface reconstruction flow chart

2.1. Image calibration.

The depth image and the color image should be calibrated first, using Smisek's checkerboard calibration [6]. The method is to capture a chessboard scheme placed in different directions at different positions under natural light, extract corner points of the obtained image, and use multiple checkerboard icons of different viewpoints to obtain an image and obtain an internal reference of the camera. Since the depth image cannot extract the texture information and the planar image

DOI: 10.25236/isrme.2019.025

cannot be recognized, the infrared image of the board is used for calibration.

2.2. Depth image restoration.

By analyzing the cause of the Kinect depth image error, the repair of the depth image is mainly to fill the cavity and remove noise. Using a combination of pixel filter and median filter to repair the depth image.

The depth mean of the hole area and the occlusion area in the depth image is determined by searching each pixel on the image to find a pixel having a depth value of 0, and determining the area to be repaired.

Let Image represent the binary image of the depth map to be repaired. For each pixel in the depth map D, set the gray value $I(u,v)$ of the image corresponding to the pixel with the depth $D(u,v)$ to 0 to 255, and set the rest to 0.

$$\mathrm{I(u,v)} = \begin{cases} 0, D(u,v) > 0 \\ 255, D(u,v) = 0 \end{cases} \quad (1)$$

These pixels are processed using a pixel filter. Pixel filtering is a spatial correlation-based image restoration method. The core idea is to estimate the pixel depth value to be repaired by using the pixel points in the peripheral area of the area to be repaired.

$$\mathrm{S}\left(\mathrm{u}\ ,\ \mathrm{v}\right) = \begin{cases} D\left(u\ ,\ v\right)\ ,\ D(u,v) \neq 0 \\ 0, D(u,v) = 0 \cap (c1 < T1) \cap (c2 < T2) \\ M, D(u,v) = 0 \cap (c1 \geq T1) \cap (c2 \geq T2) \end{cases} \quad (2)$$

Where, $S\left(u\ ,\ v\right)$ represents the depth value after processing by the pixel filter; $D\left(u\ ,\ v\right)$ is the original depth value; c1 and c2 are the thresholds in the inner layers of the filter Ω1 and Ω2, respectively; and M represents the non-zero pixel in the field Ω.

A pixel with a depth value of 0 in the original depth map is defined as a candidate filtering object, and a two-layer filter (inner layer Ω1, outer layer Ω2) is defined by using pixels in the domain Ω. Then find the pixel whose depth value is not 0 in Ω, and compare the number of non-zero pixels in Ω1 and Ω2 with the threshold corresponding to the two. If the number of non-zero pixels in any layer is greater than the threshold, the depth value of the candidate filtered pixels is replaced by the statistical mode of the depth values of all non-zero pixels in the filter, otherwise, the depth value of the pixel remains unchanged.

2.3. Denoising.

Median filtering is a kind of nonlinear filtering, which is mainly used to filter real-valued discrete signals, and can give the best estimation of signals under the condition of minimum absolute error. The median filter can overcome the blur caused by the linear filter to the image, and effectively remove the impulse noise while maintaining good edge characteristics, so as to obtain better processing results.

The basic idea of median filtering is to replace a point value in a digital image or a sequence of numbers with the median of the values of the points in its field. For a one-dimensional sequence $x_1, x_2, x_3, \dots\dots, x_n$, all data is sorted by size $x_{i1} \leq x_{i2} \leq x_{i3}, \dots\dots, x_{in}$.

$$\mathrm{y} = \mathrm{med}\{X_{in}\} = \begin{cases} X_{i\left(\frac{n+1}{2}\right)}, n(odd) \\ \frac{1}{2}\left[X_{i\left(\frac{n}{2}\right)} + X_{i\left(\frac{n}{2}+1\right)}\right], n(even) \end{cases} \quad (3)$$

In the specific implementation, a filter window with a length of 2n+1 is taken, and n is a positive integer. The window is slid over the data. The output of the median filter is the median value of each pixel in the window instead of the pixel value at the center of the window.

The digital image is a two-dimensional signal. For the two-dimensional median filtering of the filtering window A, the output can be expressed as

$$y_{ij} = med\{x_{ij}\} \tag{4}$$

In the middle, $\{x_{ij}\}$ is a two-dimensional data sequence.

3. Point cloud processing

3.1. Point cloud registration.

Because the angle of view of the Kinect camera is different, a depth image of two angles of view can be obtained. At this time, it is necessary to register these depth images so that they can be matched to the same coordinate system. The commonly used registration method is the ICP algorithm.

The ICP algorithm is an optimal matching method based on the least squares method. Through the iterative transformation, the rotation matrix R and the translation vector t of the target point set and the reference point set are found, so that the two point clouds satisfy the optimal matching under certain conditions. Suppose that the point cloud to be matched is $\mathrm{P} = \{p_i\}_{i=1}^{N_p}$ and $\mathrm{X} = \{x_i\}_{i=1}^{N_p}$, and for each point p_i in the point set P, find the nearest neighbor point on the point set X as the corresponding point, assuming that the corresponding point set of $\mathrm{P} = \{p_i\}_{i=1}^{N_p}$ is $\mathrm{Y} = \{y_i\}_{i=1}^{N_p}$, the point cloud is matched. The quasi is to find the optimal transformation relationship between P and Y, that is, to find the smallest rotation matrix R and translation vector t to minimize the objective function f.

$$\mathrm{f} = \frac{1}{N}\sum_{i=1}^{N_p}\left\|y_i - (R_{p_i} + t)\right\|^2 \tag{5}$$

3.2. Point cloud streamlining.

After completing the above steps, a large number of redundant point cloud features will be generated, and directly triangulating will consume a large amount of memory and time, so the point cloud information is streamlined.

Here, the point cloud is filtered by a downsampling filter, and the point cloud space is divided into a plurality of small cubes of equal volume by a 1 cm cube. Then, the points of these cubes are replaced by the center of gravity of the cube. The following formula:

$$\begin{cases} \overline{x} = \frac{1}{n}\sum x_i \\ \overline{y} = \frac{1}{n}\sum y_i \\ \overline{z} = \frac{1}{n}\sum z_i \end{cases} \tag{6}$$

4. Experimental results

On the window10 system, the face image reconstruction relying on point cloud information is realized by using the VS and PCL libraries. Experimental method: The color image and depth image of the surrounding scene are acquired by the Kinect sensor, and the depth image is processed and the hole is repaired. According to the calibration result of Kinect sensor, the 3D point cloud information is obtained, and the point cloud information is accurately registered by the ICP algorithm, and then triangulated to obtain a 3D scene. The result is shown in the figure2.

Figure2. 3D reconstruction rendering

5. Conclusion

Through the depth image and color image collected by Kinect, the camera is first calibrated, then the depth image is denoised, and then the precise 3D point cloud of the scene is captured by point cloud matching to reconstruct the 3D model of the scene. It can be seen from the experiment that the three-dimensional reconstruction model obtained by this method is more specific and can be seen in detail. Compared with the laser reconstruction technology, the three-dimensional reconstruction method has a simple process and is easy to operate. However, because of the RGB_D restoration, the reconstruction effect may not be partial error, and the main reconstruction error exists in the algorithm model construction and depth image extraction effects, which are still to be improved.

References

[1] H.S. Huang. *Application of ground 3D laser scanning technology in cultural relic protection,* Chang'an University. (In Chinese)

[2] D. Ai, G.B. Ni and M. Wang. *Overview of 3D reconstruction technology based on Kinect*, Journal of Sensors and Microsystems, 2017, 36(8):1-6. (In Chinese)

[3] W. Zou, Q.Q. Zheng and B. Li. *Static Gesture Recognition Based on Kinect Sensor and HOG Features*, Software Guide, 2018(2). (In Chinese)

[4] W. Cui, M.Y. Zhu and K. Yan. Application of 3D reconstruction technology based on Kinect in 3D display, Electronic Measurement Technology, 2017, 40(3): 113-116. (In Chinese)

[5] H. Geng, M.M. Zhang and J. Tong. Real-time stable 3D multi-finger tracking algorithm based on Kinect, Journal of Computer-Aided Design & Computer Graphics,2013,25(12):1801-1809. (In Chinese)

[6] J. Smisek, M. Jancosek and T. Pajdla. 3D with Kinect, Consumer Depth Cameras for Computer Vision. Springer London. 2013:3—25.