Acquisition and Processing of Depth Image Based on Kinect

Jinrong Tian [1,a,*], Ying Sun[1,2,b], and Chao Xu [2,c]

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

[2]Hubei Key Laboratory of Mechanical Transmission and Manufacturing Engineering, Wuhan University of Science and Technology, Wuhan 430081, China

[a]2571882913@qq.com; [b]sunying65@wust.edu.cn; [c]726334256@qq.com

Keywords: Curved surface; Depth image; Point cloud; Three-dimensional reconstruction

Abstract: The development of science and technology puts forward higher requirements for manufacturing industry. In order to meet the functional requirements, complex surface can be designed and manufactured. However, the processing of complex surface has the problems of difficulties in mechanical processing and low operation efficiency and low precision in manual. In recent years, machine vision has developed rapidly, and one of its branches, three-dimensional reconstruction technology, has attracted more and more attention. The acquisition of feature parameters from 3D reconstruction of complex surface and simulation and analysis are the foundation and premise of subsequent processing. The application of machine vision technology in mechanical processing is the inevitable trend of the development of automation and intelligence in manufacturing industry. Therefore, this paper proposes a Kinect-based method to obtain the depth image, preprocess and transform the depth image to get the three-dimensional point cloud of the workpiece surface. ICP algorithm is used to register the point cloud, obtaining the three-dimensional coordinates of the surface space, and realizing the three-dimensional reconstruction of the surface.

1. Introduction

In the manufacturing industry of today's world, modern equipment is mostly composed of solid parts with complex curved surfaces. However, the existence of curved surfaces makes the traditional equipment difficult to process, and the manual operation efficiency is low, the accuracy is poor, and there is visual fatigue. In addition, the manufacturing accuracy of parts is the basic requirement of parts manufacturing, and it is also one of the core technologies. Therefore, the detection of parts and the real-time monitoring of processing are essential. In recent years, with the rapid development of machine vision, it has the advantages of real-time on-line monitoring, suitable for any environment, no visual fatigue. The method of surface detection based on computer vision has attracted more attention. Among them, the use of three-dimensional reconstruction technology to model and detect complex surfaces has become a hot spot.

Three-dimensional reconstruction technology can be divided into active and passive modes. Passive modes include SFS, binocular stereo vision method and SfM [1]. SFS method calculates the height of object surface by gray level change in image, and realizes three-dimensional reconstruction. Its advantage is that it can reconstruct object's three-dimensional shape by single image, but this method has certain illumination environment requirements [2]. Binocular stereo vision, which imitates the human eye, matches multiple images from multiple perspectives to obtain the distance between the object and the camera, completing the reconstruction. This method requires higher requirements for the camera equipment [3]. The SfM method estimates the three-dimensional structure of an object by using the motion of the feature points and the projection points corresponding to the image. The above methods all depend on the acquisition of two-dimensional images, and the acquisition of image information is limited, even a large number of image information will be lost [4]. The three-dimensional reconstruction of active method

  DOI: 10.25236/isrme.2019.022

includes time-of-flight method and structured light scanning method. Time-of-flight method relies on the time difference between the sensor and the receiver to obtain the three-dimensional coordinates of the object surface. This method has high accuracy but the equipment is expensive [5].

This paper proposes a three-dimensional reconstruction method based on Kinect. The depth image is acquired based on Kinect. The depth image not only reflects the geometric shape of the visible surface of the scene, but also takes the distance from the image collector to the points in the scene as the pixel value. After coordinate transformation, it can be calculated as the point cloud data, and the point cloud is registered, that is, the three-dimensional coordinate information of the points on the object is obtained, so as to realize the three-dimensional reconstruction of the surface.

2. Kinect Sensor

2.1. Kinect depth image acquisition.

Kinect is a device for games developed by Microsoft. It contains a variety of sensors and lenses. It can realize dynamic capture, voice recognition, impact identification, and online interaction between players. Because Kinect can realize real-time three-dimensional reconstruction, it has been used in computer vision research. Kinect includes infrared CMO camera, infrared camera lens and color camera, which together constitute a 3D structured light depth sensor. Kinect captures a range of color videos and images through color cameras [6]. At the same time, infrared cameras are used to transmit and receive infrared spectra, and infrared CMO cameras are used to analyze infrared spectra and create depth images. The acquisition of depth image is based on the principle that the pattern of structured light will deform with the shape and position of the object. At different distances, Kinect emits infrared laser to the rough object to obtain speckle patterns with different shapes. It has three-dimensional randomness and has a one-to-one correspondence, so it forms spatial stereo coding. After the light source calibration is completed, the depth information of the object surface is obtained according to the size and shape of the speckle, that is, the depth image is acquired to realize the three-dimensional restoration of the object.

2.2. Kinect-based three-dimensional reconstruction process.

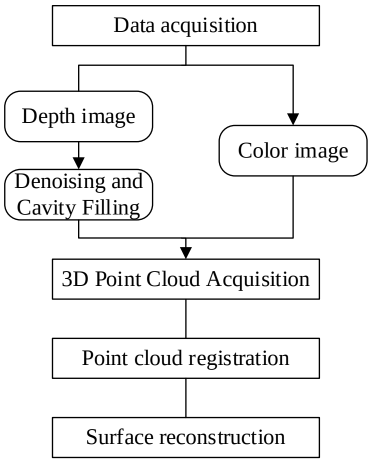

Figure 1. 3D reconstruction process based on Kinect

2.3. Depth Image Preprocessing

The depth image acquired by Kinect sensor will lead to the loss of image content for various reasons. The reasons for the loss of depth image content are analyzed as follows: Because of electronic devices, noise often exists in the acquired depth image. Because of the limited measuring distance and range of Kinect, there exists an "unmeasurable area", that is, the depth information of some areas can not be obtained. Because Kinect uses infrared reflection to obtain depth information, when specular reflection occurs on the surface of an object, it is easy to lead to inaccurate depth information acquisition in some areas, that is, cavitation occurs. When there is an object occluding the surface to be measured, part of the area cannot be measured. This kind of area presents the shape of the depth image consistent with the contour of the object [7].

Therefore, in order to obtain more accurate three-dimensional feature parameters of curved surface, the depth images obtained by Kinect should be denoised and filtered, and the common denoising algorithms include median filtering, gauss filtering and joint bilateral filtering. For the pixels in the image, median filtering can eliminate the isolated noise points and retain the edge information by replacing the original values with the median values in the field of pixels. Gauss filter weights the whole image according to the weight of correlation between pixels. Joint bilateral filtering is introduced to calculate the weight of the pixels in the image. The color image acquired by Kinect is complete. Considering the combination of color image and depth image to make up for the voids in depth image, this paper adopts joint bilateral filtering algorithm to denoise and repair the voids in depth image.

The weights of joint bilateral filtering pixels are composed of the spatial weights of depth image ω_s and color image ω_r. The weights are calculated by Eq.1[8].

$$\begin{aligned} \omega &= \omega_s \times \omega_r \\ \omega_s &= \exp\left(-\frac{(i-x)^2+(j-y)^2}{2\sigma_s^2}\right) \\ \omega_r &= \exp\left(-\frac{(G(i,j)-G(x,y))^2}{2\sigma_r^2}\right) \end{aligned} \quad (1)$$

σ_s and σ_r are based on the standard deviation of Gauss function. The gray value of color image converted into gray image is $G(i,j)$ at the pixel point (i,j), and $G(x,y)$ is the gray value of depth image at the pixel point (x,y).

For the hole position of the depth image, $G(i,j)=0$, the depth information is repaired and the filter weight is modified by Eq.2.

$$\omega = \begin{cases} \omega_s \times \omega_r & G(i,j) \neq 0 \\ 0 & G(i,j) = 0 \end{cases} \quad (2)$$

3. Acquisition and Registration of Point Clouds

3.1. Acquisition of point cloud.

The transformation relationship between depth image and point cloud is shown in Fig. 2. The depth image is transformed into point cloud, that is, the m-point coordinates on depth image are transformed into the world coordinate point M.

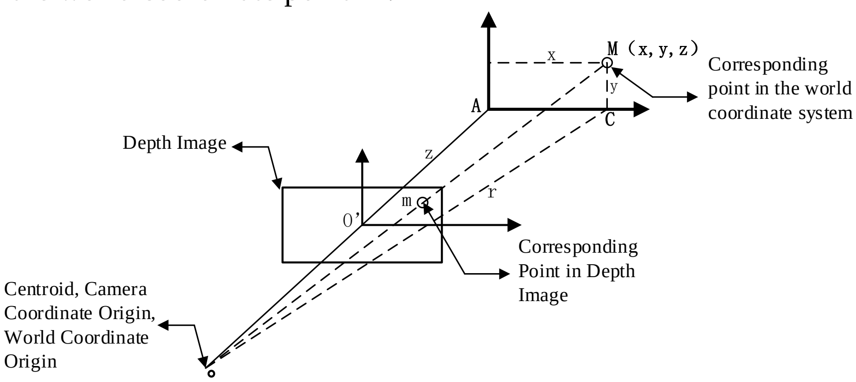

Figure 2. Conversion diagram of point cloud data and depth image

Where m point coordinates (u,v) assuming that the center point of the depth image is O', its coordinate is (u_0,v_0), based on the similarity theorem.

$$\frac{OO'}{O\bar{A}} = \frac{mO'}{MA} \quad (3)$$

The formula for its transformation is:

$$\begin{cases} x_w = z_c \cdot (u-u_0) \cdot dx/f \\ y_w = z_c \cdot (v-v_0) \cdot dy/f \\ z_w = z_c \end{cases} \quad (4)$$

Among them, the three-dimensional coordinate points in the world coordinate system are x_w, y_w, z_w. The same object in the camera coordinate system and the world coordinate system has the same depth, that is, $z_c = z_w$. So point cloud data can be obtained.

3.2. Point cloud registration.

Kinect camera obtains depth images from different angles. In order to match different depth images to the same coordinate system, different depth images need to be registered. ICP algorithm is used to improve the registration accuracy. The ICP algorithm iteratively transforms the point sets in the two coordinate systems until it finds the rotation matrix R and translation vector t of the target point set and the reference point set satisfying the convergence accuracy of registration, that is, the two point clouds satisfy the optimal matching under certain conditions [9]. The flow chart is shown in Fig.3.

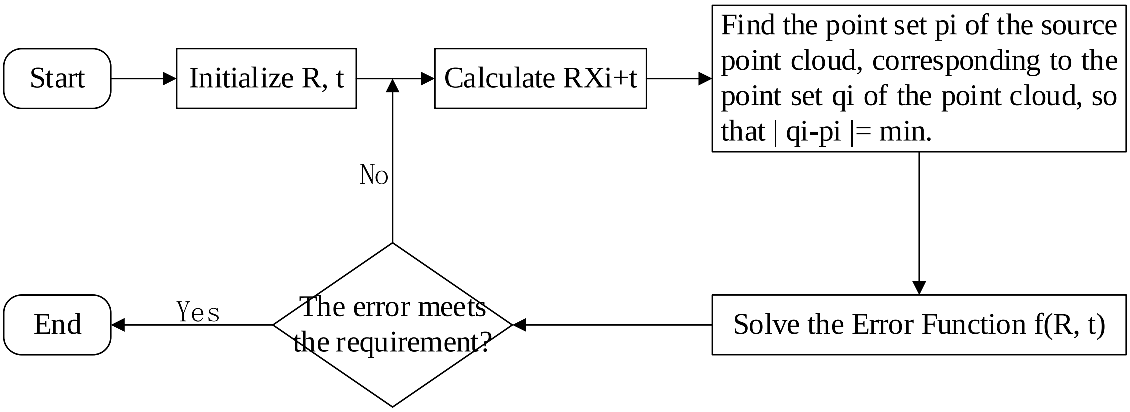

Figure 3. Point cloud registration flow chart

4. Conclusion

Aiming at the difficult problem of workpiece surface processing, this paper presents a method based on Kinect to acquire object color image and depth image, which combines color image and depth image to denoise and fill holes in depth image, calculates object three-dimensional coordinates by using the spatial geometric relationship between depth image and point cloud, and registers multiple point clouds in the same coordinate system using ICP algorithm. Finally, three-dimensional surface reconstruction is realized.

References

[1] G.S. Li and W.S. Chou: High Technology Letters, Vol. 24 (2018) No.2, p125.

[2] C.C. Hsu and C.L. Huang: Journal of Electronic Science and Technology of China, Vol. 14 (2016) No.4, p377.

[3] C.M. Chang, C.H. Chang and C.L. Huang: Journal of Electronic Science and Technology of China, (2014) No.4, 429.

[4] X.F. Dan: Real-time 3D reconstruction of indoor scene based on Kinect(Ph.D., Xi`an University Of Science And Technology, China 2017) p.2.

[5] A.Kamel, B.W. Liu and P. LiAn: International Journal of Human-Computer Interaction, Vol. 35 (2019) No.4-5, p427.

[6] J.W. Xin, Y. Luo and P.G. Cheng: Geomatics & Spatial Information Technology, (2017) No.7, p.65.

[7] S.Y. Sun, C.F. Kuang and X. Liu: Journal of Applied Optics, (2017) No. 2, p.210.

[8] S.W. Guo, X.H. Zhang, Y. Fan and Y.Z. Deng: Journal of Wuhan Institute of Chemical Technology, (2016) No. 2, P.185.

[9] A.J. Wan and X.J. Zhao: Electro-Optic Technology Application, (2017) No.3, p.62.