Proceeding of the IEEE
International Conference on Robotics and Biomimetics
Dali, China, December 2019

Stair Feature Recognition and Modeling based on Binocular Vision*

Chao Wang[1)], Jiabo Feng[2)], Weixiang Yang[1)], Weijun Zhang[#2)] and Chunbo Xie[2)]

[1)] *Kunming Power Supply Bureau*
Yunnan Power Grid Co.,Ltd.
Kunming, China
1349979306@qq.com

[2)]*School of Mechanical Engineering*
Shanghai Jiao Tong University
Shanghai, China
fjbfei@163.com

***Abstract*—Autonomous obstacle crossing for mobile robots is an important area of research. The premise of autonomous obstacle crossing is the ability of a robot to sense the environment and determine the size of obstacles. This paper uses stairs as the research object, which is the most common and complex obstacle model in structured environments. By analyzing the structural features of stairs, a method to extract the geometric features of stairs from a point cloud is proposed. From the geometric dimensions of the stairs, the data related to robot obstacle crossing are extracted to construct an obstacle model. The experimental results show that this method can effectively extract the characteristics of stairs and provide an obstacle model that can meet the requirements for obstacle crossing.**

Keywords—Obstacle Crossing, Stair Identification; Binocular Vision;

I. INTRODUCTION

Environmental information acquisition and modeling are key processes for mobile robots to autonomously negotiate obstacles. When moving in dangerous indoor environments, such as nuclear environments and fire scenes, a robot mainly faces structured environments, such as corridors, stairs, and steps. Staircases are the most common obstacles in most circumstances and are also one of the most difficult obstacles to negotiate. In order for a robot to climb stairs autonomously, accurate information on the location and size of the stairs needs to be acquired.

At present, there are many sensors for mobile robots to perceive the surrounding environment and obstacles, including millimeter wave radars, laser radars, ultrasonic sensors, and vision sensors. Compared with the other sensors, the largest amount of environmental information can be obtained through a vision sensor. In addition to the dimension parameters, such as size, shape and depth, information on the color and texture can also be obtained.

Many studies have been carried out to identify obstacles, such as stairs, in a structured environment using a vision sensor. Stephen Se[1] used a monocular camera to extract the edges of stairs to identify stairs and used the information to assist blind people when they walk. Js Gutmann[2] used stereoscopic vision to look down a set of stairs from the top to the bottom. Through the reconstruction of the stairs, the stairs were identified. Then, the information obtained by vision was used to guide a humanoid robot as it moved up and down the stairs. Xiong[3] et al. used vision to extract the horizontal lines on stairs to estimate the posture of a vehicle body and finally guided an unmanned autonomous vehicle in a study on stair tracking.

In this paper, a binocular vision sensor is used to acquire environmental information based on indoor lighting conditions. By rasterizing the space divisions of a point cloud, the index of the point cloud is established, and the normal vector of the point cloud is calculated. The regions of the point cloud are segmented based on the normal vector parameters of the point cloud, and the plane features are extracted. The extracted plane is used to calculate the position of the edge of the staircase, and then the dimension data are provided for the robot to cross the obstacle. The experimental results show that this method can identify the characteristics of stairs well and has a specific anti-interference ability.

II. POINT CLOUD DATA

A 3d point cloud refers to discrete points of surface features collected by measuring instruments. Point cloud data often contain the 3d coordinates of each discrete point on an object, RGB information, size, positional accuracy, spatial resolution, normal vector and curvature of the surface, etc. The data can directly reflect the spatial distribution and surface characteristics of targets [4]. Compared with other sensor data, point cloud data are denser and can better describe the surface of an object. Point clouds are the most important and commonly used method to describe 3d environment data.

A. Binocular Vision

Methods for obtaining 3d point cloud data can be divided into two categories according to different measurement methods: noncontact measurement and contact measurement. Environment modeling mainly adopts binocular vision, monocular vision, depth cameras and other noncontact measurement methods to obtain three-dimensional environment information. Each measurement method has its own advantages, disadvantages and suitable environments. The research object of this paper is an indoor structural environment. Both the light intensity and environmental surface characteristics are suitable for the use of binocular cameras.

Binocular vision uses the parallax method to calculate the depth information of the object (Fig.1). This method is sensitive to distance, and the measurement accuracy decreases rapidly with increasing distance. Therefore, the environmental information acquired by a binocular camera is

* This work is partially supported by China Southern Power Grid (YNKJXM20170208)
# Corresponding author, zhangweijun@sjtu.edu.cn

high-quality only when the distance is small. The point cloud density and calculation accuracy of the distant object are extremely poor, so it is useless for environment modeling.

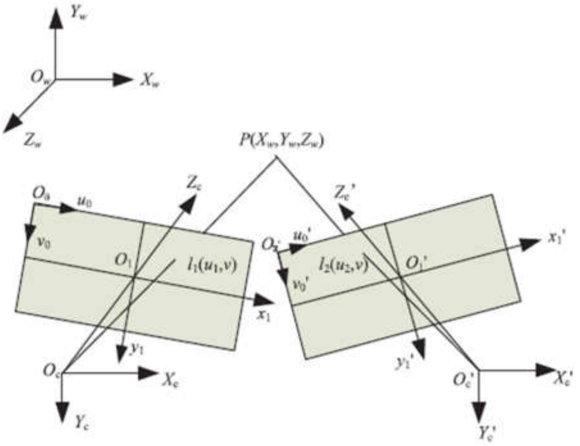
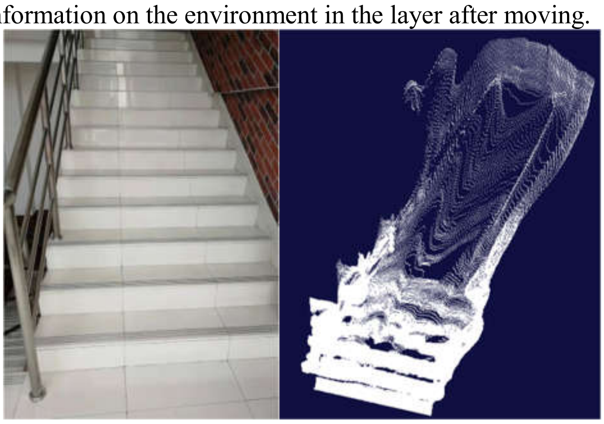


Fig. 1. Geometric model of a binocular camera.

Binocular vision theory is well established, and it is not discussed in this paper. Fig. 2 (a) shows the stair object to be detected. The surface of the stairs is smooth tile, which easily reflects light. As a result, interference easily occurs with a tof camera or a structural light camera, which is unsuitable. In this paper, a binocular camera is chosen as the sensor to measure the stairs. The calculated point cloud data are shown in fig. 2 (b). The characteristics of the first three steps (approximately 1 m away) are obvious. The distant point clouds are very sparse, and it is difficult to obtain the structural features. The useful data can only provide the size of the terrain for a small section of the robot's path. Therefore, the robot needs to continuously acquire information on the environment in the layer after moving.



(a) (b)

Fig. 2. Stairway 3-D point cloud diagram

B. Space division based on raster

The point cloud obtained through binocular stereo vision matching is a scattered point cloud. There are no direct topological relationships between points, and direct computing requires the consumption of many resources. Therefore, it is necessary to establish indexes first to facilitate data processing and reduce the computational time. Common methods include the spatial cell method, octree method, k-d tree method, etc. [5].

In this paper, a 3d raster is used to divide the point cloud. With this method, the point cloud data only needs to be traversed once. Fig. 3 is a schematic diagram of the 3d raster space. The steps of this method are as follows:

- Determine the detection space range. The width range(X_{MIN}, X_{MAX}), height range (Y_{MIN}, Y_{MAX}) and depth range(Z_{MIN}, Z_{MAX}) of the environment to be detected are determined according to the characteristics of the camera and the motion planning requirements.
- Create a cube box with points A(X_{MIN}, Y_{MIN}, Z_{MIN}) and B (X_{MAX}, Y_{MAX}, Z_{MAX}) as the corner points. Rasterize the cube with a grid with a side length of a = 10 mm, as shown in Fig. 3. Each grid has a three-dimensional (X_{INDEX}, Y_{INDEX}, Z_{INDEX}) index associated with the coordinates. Establish a linear relationship between the raster index and raster coordinates.
- Traverse all points and calculate the raster index using the point coordinate data, according to the linear relationship established in the previous step. Each point will be placed into the corresponding raster according to the index.
- Count the number of points in each raster to evaluate the quality of the corresponding point cloud. The raster with a point cloud of high quality has a large weight in the subsequent calculation.

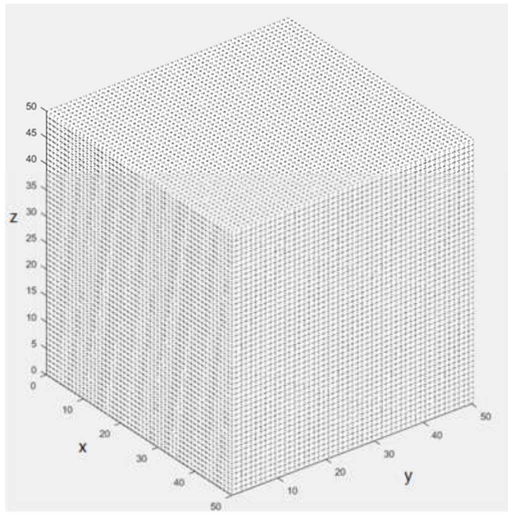


Fig. 3. 3-D grid space map

For the task of climbing stairs, the required parameters only include the height and width of the stair steps in the space ahead of the robot, as well as the distance from the robot. The location of the wall can be obtained by a laser radar. Moreover, the point cloud data describe a series of nonoccluded surfaces. For this reason, we simplified the 3d point cloud index method to reduce computational and memory requirements. The dimension of the raster index is reduced by proposing an empty raster or poor-quality raster of the point cloud. The specific steps are as follows:

- Take all rasters with the same indexes of X_{INDEX} and Y_{INDEX} as a group.
- In each group of points, select the raster with the best point cloud quality as the effective raster. Discard the remaining grids.

- Rebuild the raster index by preserving only X_{INDEX} and Y_{INDEX} and deleting Z_{INDEX}.

By rasterizing the point cloud, we obtained a set of ordered point clouds, which created conditions for the subsequent calculation. The advantages of the raster method are the short search time and the ability to evaluate point cloud quality. The results from practice trials show that this method can process point clouds in real time. The disadvantage is that point cloud density is required, which can be satisfied by the advancement of vision sensor technology.

III. Stair Feature Extraction

To obtain three-dimensional data of the stairs, the corresponding geometric features, including corner points, straight lines and planes, need to be extracted from the point cloud according to the shape of stairs. In the structured scene, theoretically, we can obtain the size data of the scene mainly by identifying edges in the environment. However, due to the measurement principle and the performance of the sensor, it is difficult to obtain a sufficient number of points to compute the coordinates of the corners. Due to the quality of the point cloud mass, the corner point may not even have point cloud data. Therefore, it is difficult to extract the characteristics of the contour. To avoid this problem, we select the plane features with dense point clouds as the objects to be extracted. The stairs have obvious plane and geometric features and contain abundant point clouds, which can effectively prevent interference and reduce position errors.

A. Normal vector of the point cloud

After the point cloud grid is completed, the order of all points is known, and the plane can be constructed. To obtain the correlation between these points and determine whether a point and the surrounding points are points on a plane, the normal vector method is adopted. The plane fitting method and triangular mesh method are usually used to calculate the normal vector of the point cloud. The plane fitting method is usually used to fit all the points in the neighborhood of the point during the calculation. The normal direction information of the fitted plane is used as the estimated value of the normal vector at the point. The triangulation method uses the scattered point cloud to construct an irregular triangulation network, calculates the normal vector values of all the adjacent triangular faces, and then assigns the normal vector of these adjacent faces as the normal vector of this point. Using the above methods to calculate the normal vector requires a large amount of system resources, especially to establish the triangular grid and store the topological relations, which reduces the efficiency of the program.

In this paper, the least square method is used to determine the fitting plane. The least square method is used to identify a plane that approximates the tangent plane near a point. Moving infinitely closer to the tangent plane of the point will allow one to identify the vector normal to that point and the vector normal to the point cloud.

Given any point $P\left(p_x \quad p_y \quad p_z\right)$ in the space, the equation of a plane to be fitted near point P is:

$$F(x, y, z) = ax + by + cz + d = 0 \tag{1}$$

The purpose is to solve the coefficients a, b, c and d in the equation. Substitute the coordinates of the neighboring points of point p into equation (1). A system of linear equations are obtained and used for solving the plane equations. Parameter $\left(x_i \quad y_i \quad z_i\right)$ is the coordinate of the points near point P.

$$\begin{bmatrix} x_1 - p_x & y_1 - p_y & z_1 - p_z \\ x_2 - p_x & y_2 - p_y & z_2 - p_z \\ \vdots & \vdots & \vdots \\ x_k - p_x & y_k - p_y & z_3 - p_z \end{bmatrix} \begin{bmatrix} a \\ b \\ c \end{bmatrix} = 0 \tag{2}$$

The vector whose parameters are unitized is the unit normal vector. The vector direction of the scattered point cloud method calculated by the micro tangent plane is arbitrary, as it is the direction from a point to the inside of the point cloud surface and from a point to the outside of the surface, so we adjust the direction to ensure that the curvature calculation does not cause confusion.

B. Regional Segmentation

When the normal vectors of the adjacent grid are the same, they are suspected to be of the same plane of the obstacle. To further determine the plane of the obstacle, we use the flooding method to identify the connected area, and the connected area of the adjacent grid is a complete plane.

Flood fill, also called seed fill, is an algorithm that determines the area connected to a given node in a multidimensional array [6]. The principle of this method is to start from a point inside the target area, namely, a seed, and spread to all the pixels in the area in a certain direction. The definition of the boundary adopted by the flooding method is that all pixels on the boundary of the region have a certain color value, all pixels inside the region do not have this particular color, and the pixels outside the boundary can have the same color value as the boundary. Here, the color value of each point is marked by 0, 1 and 2.

The detailed procedure is given as follows:

- Find and mark the pixel points of seed (x, y) according to certain rules;
- Detect the color of this point. If the color is different from the boundary color and the fill color, fill the point with the fill color; otherwise, do not fill it;
- Detect the adjacent position of this point and continue the last step until all pixel points within the boundary range of the region have been detected.

When performing steps (2) and (3) to search for pixel points, two types of detected adjacent pixels will appear: a 4-adjacent connection and an 8-adjacent connection. A 4-adjacent connection starts from a point in the region and searches through four directions: up, down, left and right. An 8-adjacent connection is similar to a 4-adjacent connection but has the following additional four directions: right up, right down, left up and left down.

Finally, the obtained results are grayscale processed. The graying process is relatively simple. Only the RGB values in the color images need to be obtained and calculated by using a formula. The formula is as follows:

$$GREY = RED \times 0.299 + GREEN \times 0.387 + BLUE \times 0.144 \quad (3)$$

C. *Stair Edge Extraction*

By dividing the data into regions, we obtain a series of planes. However, we cannot use these planes to describe the geometry of the stairs directly. One reason they cannot be used directly is that the planes are disorderly and need to be processed further, and another reason is that edges are better suited to describe the geometry of the stairs than are flat surfaces. Moreover, due to the limited angle of the camera, these planes are the vertical planes of the steps. These planes cannot be used for obstacle-crossing planning, either for walking robots or for tracked robots. In this research, we describe the geometric dimensions of the staircase using an ordered set of angular edges.

(1) Binarization

In the process of image processing, grayscale images usually do not meet the requirements for feature value extraction. Therefore, the grayscale images need to be processed to filter the invalid information. Through binary processing, the target features become more prominent.

(2) Edge detection

In the obtained grayscale image, there is often an edge, which is near the edge at which the gray value changes unevenly and there are large changes in the gray value. Based on this property, we can identify the place where the gray value changes by a large value by using the derivative method. At present, the commonly used methods mainly include first-order differentiation, and the commonly used operators include the Robert operator, Sobel operator and Prewitt operator [25]. Among the operators, the Sobel operator works best, and we choose this operator for calculation.

(3) Line extraction

After the edge detection process is completed, it is necessary to extract the straight line features, which are the edges that are the straight lines of the stairs. In this paper, Hough transform is used to extract the straight lines. The detected line segments are merged and screened according to the characteristics of stairs, and the line segments used to describe the characteristics of the edges of the stairs are obtained.

D. *Obstacle model*

Depending on the mechanical mechanism of the walking mechanism, the obstacle model is different for different robots. In this study, the walking mechanism of the robot is a tracked flipper. For the crawler structure, the edge of the staircase and the crawler of the robot are in contact when the robot is climbing stairs. The walking path of the robot is an intangible slope formed by edges. Therefore, a virtual ramp can be constructed using the pose data of the edges, which is used as the basis for obstacle climbing planning for the robot.

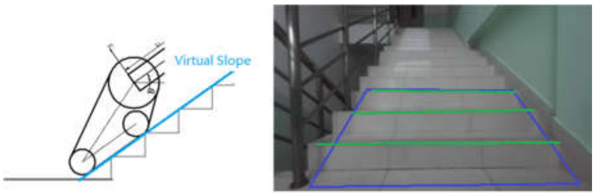


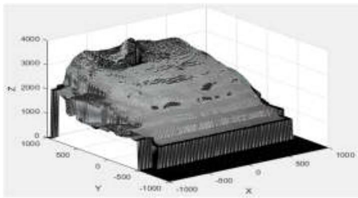
Fig. 4. Virtual ramp based on the edges of the stairs

IV. EXPERIMENT

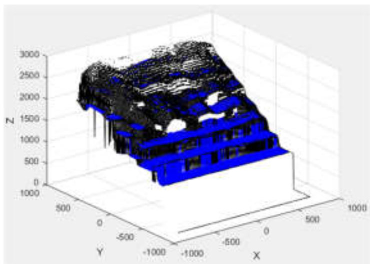
A. *Acquire Stair Characteristics*

The stairs involved in the experiment are shown in fig. 5 (a). The rasterized point cloud is shown in fig. 5 (b). All the normal vectors of the inner point cloud are calculated. The grid with the normal vector that meets the requirements is rendered, as shown in fig. 5 (c). The location of the nearest three steps can be obtained when the connection area determined by the flood fill method of flooding water is searched, as shown in fig. 5 (d). Through edge extraction and straight line extraction, the straight line features of the stair edges are obtained, as shown in fig. 5 (e).

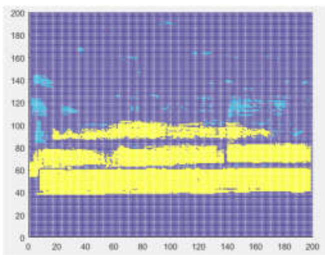
(a) Stairs in the experiment



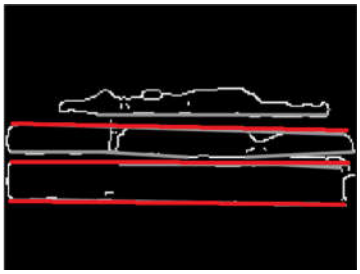
(b) 3D point cloud containing depth information



(c) Plane region of similar normal vectors



(d) The extracted plane region



(e) Line segments describing the stairs edge

Fig. 5. Stair feature extraction process

Two staircases of different sizes are used to test the algorithm. The result is shown in table I. The calculated value is obtained using the algorithm in this paper, and the average of multiple computations is used. The actual value is obtained by measurements. The height and width of a single stair, as well as the distance to the camera, are used as parameters to evaluate the effect of the algorithm.

TABLE I. CALCULATED VALUES AND ACTUAL VALUES

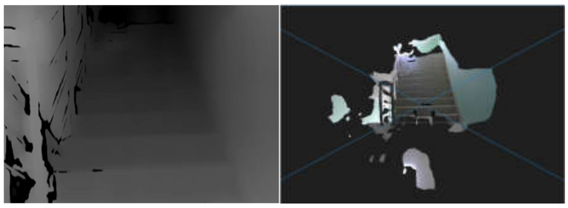
Calculated height	Calculated width	Calculated distance	Actual height	Actual width	Actual distance
15.2cm	27.3 cm	98.1 cm	16 cm	28 cm	100 cm
16.4 cm	27.5 cm	101.5 cm	16 cm	28 cm	100 cm
14.1 cm	26.6 cm	97.2 cm	15 cm	26 cm	100 cm
16.1 cm	25.1 cm	100.8 cm	15 cm	26 cm	100 cm

As shown in the table, the calculation error of height and width is approximately 1 cm. The deviation from 1 m is approximately 3 cm. The accuracy of the algorithm meets the mobile robot motion planning requirements.

According to the analysis of experimental data, the main types of errors are the following: errors from the camera itself, as the distortion of the camera lens and the influence of light will affect the acquired image quality; errors in the visual algorithm, the process of obtaining point clouds by deep calculations and the process of point cloud data processing; and errors from the influence of noise, data screening and calculations. The errors in the control system lead to an inaccuracy in the velocity and swing angle.

B. Stair climbing tests

For an autonomous stair climbing experiment with a track-flipper robot, the first step is the process of identifying the staircase parameters, as shown in fig. 6, which includes the robot's visual processing step. Fig. 6 (a) is an image of the depth information of the staircase, and fig. 6 (b) shows the process of visually generating the 3d point cloud data. The height and depth of the stairs in the experimental site are 16 cm and 28 cm, respectively. With the identified staircase data, the position and angle parameters of the virtual slope were input into the controller. According to the preplanned climbing movement, the angle of the flippers can be controlled to adapt to the staircase structure.



(a) Depth image (b) Point cloud

(c) Preparing to climb (d) Climbing the stairs

Fig. 6. Visual recognition scene

During the experiment, it was found that the track contact may be insufficient due to model inaccuracies, which will cause the track to slip.

CONCLUSION

Autonomous mobile robots that are used to navigate stairs in dangerous environments for reconnaissance, disaster relief or other purposes require the ability to perform basic intelligent behaviors. On the basis of an existing robot, this paper proposes the use of binocular vision for 3d point cloud generation, and the 3d point cloud is used to detect stairs and identify the parameters; this method provides a basis for the realization of robots that can autonomously climb stairs.

REFERENCES

[1] J. Stephen Se, and M. Brady . "Ground plane estimation, error analysis and applications." Robotics and Autonomous Systems 39.2(2002):59-71.

[2] J. Gutmann, J. S., M. Fukuchi, and M. Fujita. "A Floor and Obstacle Height Map for 3D Navigation of a Humanoid Robot." IEEE International Conference on Robotics & Automation 2006.

[3] J. Xiong, Yalin, and L. Matthies. "Vision-guided autonomous stair climbing." IEEE International Conference on Robotics & Automation 2002.

[4] D. Zhong Qian. "Massivee Point Cloud Data Processing Technology based on CUDA". Zhejiang University of Technology, 2014.4.

[5] D. Li Tianqing. "Research on Autonomous Climbing Stairs of Mobile Robot Based on Multi-sensor Fusion".

[6] https://en.wikipedia.org/wiki/Flood_fill

[7] J. SHANG W Y, QIU F J. "Design and Optimization of the Suspension System on the Wheel-Tracked Exploration Robot". Applied Mechanics and Materials, 2010, 37(2)：94-98.

[8] M. YOO A C, GILBERT G R, BRODERICK T J. "Military robotic combat casualty extraction and care" . Surgical Robotics. Springer. 2011: 13-32.