Research on Robot Binocular Ranging Based on SIFT Feature Extraction Algorithm

Dong Haifeng¹, Yao Jun¹

¹School of Computer Science, Xi’an Shiyou University Xi’an, shannxi,710065, China
Email: 1324384654@qq.com

Abstract. Binocular stereo vision belongs to the reconstruction technology of images. It has the characteristics of non-contact measurement and easy implementation. It has broad application prospects in many fields. The study of its basic theory also has great theoretical and practical significance. As part of stereo vision, robot binocular distance measurement has an irreplaceable role. In this paper, a complete binocular ranging system is divided into the following five steps: principle analysis, camera calibration, image preprocessing, SIFT feature point extraction, distance measurement, a simple model in robot binocular ranging is realized, which is a robot Route planning and obstacle avoidance provide data support, which has the advantages of easy operation and simple calculation.

1. Introduction
If the robot needs to be positioned in action, distance measurement is essential[1]. There are many kinds of robot ranging methods, such as monocular ranging, distance measurement based on depth of field [2], and the binocular ranging of robots is the most accurate and the most widely used modern robot. The ranging method is also the main research object of the subject. The distance measurement method we use is the planar binocular stereo vision distance measurement method [2], that is, two identical cameras are fixed on the same horizontal plane and are parallel to each other. The optical centers of the left and right cameras are collinear. The main contents are as follows: According to the principle of planar stereo binocular vision ranging, the core ranging formula is derived; the camera is calibrated using Zhang Zhengyong’s checkerboard calibration method in the Matlab toolbox[3]; then several groups of pictures are preprocessed, and the image is first Grayscale, the smoothing function provided by opencv is used to process the image, so as to prepare for the following image feature extraction and matching [4]; finally, SIFT feature extraction algorithm is used to start stereo matching [5], and the obtained data is introduced into the distance calculation formula to obtain the distance measurement results and modify the results.

2. Principle of binocular vision ranging
The binocular vision ranging of this subject is based on plane stereo vision [6], that is, the two cameras are placed in parallel, the optical axis is parallel, the projection surface, etc. are all parallel, the focal length is equal, and it can be understood that the same camera is on the horizontal plane Pan, the imaging and ranging principle diagram is as follows:
Figure 1. Binocular distance measurement plan

T is the distance between the two cameras, called the baseline distance; the focal length of the camera is f, where the left camera is equal to the focal length of the right camera; P is the feature matching point formed by an object in space between the left and right cameras. Figure 1 A triangle is formed between a point to be measured and three points of the object image point in the space taken by the left and right cameras. The two left and right cameras form different visual images of the same point. Based on this parallax, it is the main basis of our planar stereo vision ranging. We use this parallax to obtain the location information of the unknown three-dimensional world. Let the camera obtain the image of the feature point P to form an image point, and form a triangle with the feature point P. Let the left image point coordinate of the camera be P_left (X_left, Y_left) and the right image point coordinate be P_right (X_right, Y_right). Z is the distance from P to the camera;

\[ D = X_{left} - X_{right} \]  
(1)

From similar triangles:

\[ \frac{Z-f}{Z} = \frac{X_{left} - X_{right}}{T} \]  
(2)

From (1) and (2):

\[ Z = \frac{fT}{D} \]  
(3)

After a series of transformations, we finally got the transformation from the left image plane to the world coordinate system with the optical center of the left camera lens of the binocular camera as the coordinate origin. It has been successfully matched in the imaging planes of the left and right cameras. The coordinates in the image coordinate system of the left camera are \((x, y)\). The offsets \(C_x\) and \(C_y\) of the origin in the coordinate system of the left and right image planes and the three-dimensional coordinate system can be obtained by camera calibration:

\[ \text{Distance}\ Z = \frac{\text{-}Tf}{D - (C_x - C_y)} \]  
(4)

### Experimental data

| Distance formula | \(k\) | \(b\) | \(Z\) | distance |
|------------------|------|------|------|----------|
| \(fx*T = k*\text{dis} + b\) | 44.97 | 14454.17732 | b/(d- k) | 144541.7732/(pt2.x - pt1.x -44.97) |

### 3. Image preprocessing

#### 3.1. Camera calibration

Based on Zhang Zhengyong’s checkerboard calibration method in the Matlab toolbox, we collected 12 sets of image pairs (the photos of the left and right cameras on the same attitude calibration board are one set). The image basename is read into the picture, and the program automatically selects the corner points (left in Figure 3), and then generates a spatial rendering of the binocular camera (right in Figure 3):
After many iterations of the program, the internal and external parameter matrix of the camera will be finally obtained.

\[
\begin{bmatrix}
823.338 & 0 & 355.581 & 0 \\
0 & 823.696 & 211.364 & 0 \\
0 & 0 & 1 & 0 \\
\end{bmatrix} \quad \begin{bmatrix}
825.887 & 0 & 355.581 & 0 \\
0 & 825.955 & 211.364 & 0 \\
0 & 0 & 1 & 0 \\
\end{bmatrix}
\]

(5)

3.2. Image preprocessing
When the camera collects objects in the real world, there are various errors. In computer graphics, we call this factor image noise. The pictures collected by the binocular camera used in the experiment are all color pictures, which contain too much information. When the calculation is processed, a lot of time and space will be wasted and the speed is relatively slow, so we need to replace the color picture with a grayscale picture, that is, grayscale, and use the Gaussian filtering method to eliminate Gaussian noise in the picture and retain the original picture information, especially the edge information of the picture.

4. SIFT feature matching algorithm to extract feature points
SIFT, that is, feature-invariant feature conversion, it finds extreme points in spatial scale, and extracts its position, scale, and rotation invariants. It has a wide range of applications. SIFT features are based on some local appearance points of interest on objects. It has nothing to do with the size and rotation of the image. The SIFT feature has a large amount of information and is suitable for fast and accurate matching in massive databases. This paper decomposes the SIFT algorithm into the following four steps: scale space extreme value detection, key point location, direction determination, and key point description.

4.1. Experimental procedure
In the experiment, the SIFT matching algorithm is used. After the image is Gaussian blurred, the Gaussian pyramid is established, and then the feature point extraction step using the SIFT algorithm is as follows:

4.1.1 Extremum point detection:
Gaussian scale space is represented by a Gaussian pyramid in the code;
- Gaussian pyramid: gauss_pyr = build_gauss_pyr (init_img, octvs, intvls, sigma);
- Gaussian difference pyramid: dog_pyr = build_gog_pyr (gauss_pyr, octvs, intvls);

The pyramid model of the image refers to: continuously reducing the order of the original image to obtain a series of images of different sizes, from large to small, from the bottom to the top of the tower model. The original image is the first layer of the Golden Tower, and the new image obtained...
by each downsampling is one layer of the pyramid (one image per layer), and each pyramid has a total of n layers. The number of pyramid layers is determined by the original size of the image and the size of the image on the top of the tower. The calculation formula is as follows: Number of groups of Gaussian pyramid: octvs = log( MIN( init_img->width, init_img->height)) / log(2); Look for the extreme point in the scale space of the Gaussian difference pyramid. The search criterion for the extreme point is: center on the detection point, compare between 3*3*3 adjacent pixels, and compare the brightest point, namely the maximum point and the darkest point, i.e., the minimum point, are recorded in an array;

4.1.2 Determining the orientation of key points
In the previous step, we can further locate on the basis of the extreme points we can find, eliminate some edge effects, and find the key points.

4.2. Feature point matching
Feature point matching is the stereo matching we mentioned earlier. His main purpose is to find the points in the same position in the real space on the left and right pictures, and match them, and connect them with epipolar lines. A parallax parallax, which is the most important step for us to perform binocular distance measurement; after extracting the feature points of the left and right pictures through the SIFT algorithm, matching is performed:

![Figure 3 Photos collected by the left and right cameras](image)

Read the two pictures above into the program and run the program to get the following matching diagram:

![Figure 4 Matching results](image)

5. Distance measurement and result analysis

5.1. Ranging results
In the above discussion, we have known all the steps in the ranging process, and then bring the relevant data into the distance calculation formula: Where T is the baseline distance (mm) of the two cameras, the focal length f (pixel value) is available for camera calibration, the parallax value D (pixel value) is also calibrated by the camera, and Cx-Cy (pixel) is read by the computer, so the distance The distance unit is consistent with T, that is, the measured distance unit is also millimeters, without performing related conversions. After running the main function, the following distance measurement results are obtained:
5.2. Results analysis
From the distance measurement results in the above figure, we can see that the distance between the object to be measured and the camera is approximately 840.2mm. For easy observation and analysis, the matching points on the two pictures are simultaneously output. The coordinates and the parallax of the two points are output. The parallax of the matching feature points for a certain point to be measured in the two pictures is basically stable at about 218 pixels. After analysis, we can see that the experiment is basically completed on the above, and the distance is measured more accurately.

There are still errors in the results. Through analysis, it is found that the focal length of the camera lens is not fixed. As the distance of the distance-measuring object from the camera increases, the focal length of the camera will gradually change. For this relationship, we tried to fit. Finally, the virtual fitting relationship of the focal length of the camera is obtained, and the results are corrected.

6. Conclusion
This article is based on SIFT feature extraction algorithm for robot binocular ranging. It is a kind of ranging method using OPENCV function library. The model we are studying is the most simple planar stereo binocular vision ranging model. After analysis, We know how to reverse the three-dimensional coordinates of the real world from the two-dimensional pixel coordinates, and the reconstruction of the three-dimensional world, which includes a series of problems such as the internal and external parameters of the camera, the distortion factor, the principle of binocular visual distance measurement. Within the allowable range, comprehensive calibration results have obtained ranging data.

Acknowledgments
I am grateful to my school for giving me the opportunity to learn and become talented, and thank the teacher for giving me patient guidance and criticism and correction when I have no direction, so that I can improve the deficiencies of the thesis in time.

References
[1] Guo Pan, Du Hong. Research on robot binocular ranging algorithm[J]. Wireless Internet Technology, 2018,15(05):99-101.
[2] Shao Nuan, Li Huianguang, Liu Le. Binocular vision ranging based on feature matching algorithm [J]. Journal of Yanshan University, 2012, 36(01): 57-61.
[3] Wang Lianzeng. Binocular vision positioning and ranging system based on VS+OpenCV[J]. Modern Computer, 2020(14):58-62.
[4] Sun Sirui, Tao Yuesong, Chen Bin, Guo Dong, Zhang Chi. Binocular vision intelligent anti-collision system based on Opencv[J]. Automation Technology and Application, 2020, 39(05): 11-14.
[5] Sun Ying, Liu Zhongli. Research on Multimedia Image Information Retrieval Method Based on SIFT Features [J]. Information and Computer (Theoretical Edition), 2020, 32(07): 179-180.

[6] Yang Chenxi, Hua Yunsong. Research on target distance measurement based on binocular stereo vision [J]. Software, 2020, 41(01): 128-132.