Research on Stereo Vision Technology Based on Improved Region Growing Method

Shilong Dai¹,²,³*, Xinghua Xia¹, and Hualiang Zhang²,³

¹School of Information and Control Engineering, Shenyang Jianzhu University, Shenyang 110168, China.
²Industrial Control Network and System Laboratory, Shenyang Institute of Automation, Chinese Academy of Sciences, Shenyang 110016, China.
³Institutes for Robotics and Intelligent Manufacturing, Chinese Academy of Sciences, Shenyang 110169, China.

*Corresponding author’s e-mail: xxh8787@sjzu.edu.cn

Abstract. Image segmentation is the cornerstone of image analysis and image processing, its main difficulty is the ill-posedness of image segmentation. The region growing method is the most commonly used method in image segmentation. Its advantages are fast calculation speed, lower algorithm difficulty, and easy understanding. This article uses the area growing algorithm and TOF combined with binocular fusion technology. Discussed how to select the seed points in the region growing method. The algorithm proposed in this paper has high accuracy and high matching quality in the boundary area of the object and the area with large difference in depth. It has both matching time and reliability. It has good results and overcomes the limitations of the active and passive distance methods in the vision system.

1. Introduction

In traditional measurement methods, the measurement tools used by people are specific, poor in versatility, and time-consuming. They are often only measured for two-dimensional plane geometric quantities. However, when directly using two-dimensional images for identification and size measurement, due to the influence of the principle of the camera's near-large and far-small [1], it is difficult to obtain scale information, requires a lot of calculations, and large errors. This article will start from the perspective of TOF and binocular vision fusion, the fusion image will be processed for region growth to improve the three-dimensional density accuracy.

Image segmentation is the technique and process of dividing an image into several specific areas with unique properties and proposing objects of interest [2]. At present, image segmentation technology is widely used in target recognition and tracking, image production and inversion [3]. This method can segment the regions with the same characteristics in the image, retain clear boundary contour information and segmentation results [4]. Therefore, it has been favored by many domestic and foreign scholars.

Wu Haibin et al. [5] proposed the use of two-dimensional OTSU to select seed points for regional growth. Zhang Le et al. [6] used the peak feature in the image histogram to determine the seed object. Wang Haoming et al. [7] used Harris angle detection to automatically find seed points. Wan and Higgins [8] proposed the theory of symmetrical region growth. The centroid growth method is used to
segment the reconstructed image, and the pixel gray value of the grown area is used to replace the initial seed point to obtain a better segmentation effect.

This paper proposes a new matching method for merging range data into a robust seed growth algorithm for stereo matching. A calibration system consisting of a TOF depth camera and a stereo color camera pair projects range data onto each of the two images, thereby providing an initial sparse set of pointing points. The point correspondence (seed) between the two images. The combination of the initial seed set and the region growing algorithm makes up for the shortcomings of the respective algorithms.

2. Brief description of basic principles

2.1. Experimental platform construction

The algorithm of TOF and binocular fusion is divided into first fusion and post fusion. The post fusion algorithm is to perform point cloud registration between the 3D data obtained by TOF technology and the 3D data obtained by the binocular camera, and set different weights for them. In order to adapt to different environments, three-dimensional data can be obtained, but this method only merges the two kinds of data. In this research, the first fusion method is selected, the TOF depth data is used as a priori parallax, and then mapped to the images taken by the left and right cameras, and then the known matching points in the image are used as the growth factor of the region growth, according to the growth area is matched. The hardware platform of the fusion system in this study is mainly composed of two identical industrial cameras and a TOF camera, and its build model is shown in Fig. 1.

![Figure 1. Experimental hardware platform](image)

2.2. Fusion algorithm flow

Firstly, the algorithm uses the point cloud data obtained by the TOF camera to calibrate and filter, the processed data is more accurate, then the data is mapped to the left and right images taken by the binocular camera, the mapped data is used as a seed. In this article, the maximum similarity calculation of the color is used to determine the corresponding relationship between the points in the 8 directions around the seed and its corresponding relationship. If found, use this point as the latest point to continue searching for similar points, if not, end this. In the second calculation.

2.3. TOF point cloud data noise reduction

First, use TOF to obtain the original image, as shown in Fig. 2(a), and then perform depth registration on the original data obtained by TOF, and then obtain the registered point cloud data, as shown in Fig. 2(b).

![Figure 2. Original image and point cloud image obtained by TOF](image)
After using TOF to obtain the point cloud data, it can be found that there are many noise points and outliers. The point cloud data at the boundary of the object that can be seen shows a gradient drop, which is greatly deviated from the real situation. Although these points will not have an excessive impact on the robustness of the growth of the region. Using the data obtained by the TOF camera to guide the binocular stereo vision matching, it is necessary to perform point cloud filtering processing on the three-dimensional point cloud data obtained by the TOF.

The sparse outlier removal method adopted this time is based on the statistical filtering of point clouds, which calculates the distance distribution from a point to a neighboring point in the input data. Assuming that the result obtained is a Gaussian distribution whose shape is determined by the mean and standard deviation, the points whose average distance is outside the standard range can be defined as outliers and can be obtained from the data set Removed, the specific process is as follows:

1. For each point \( p_i (i = 1, 2, \ldots, N) \), calculate its K nearest neighbors, and get the average distance \( d_i \) of this K nearest neighbors.
2. For all \( d_i \), count the mean \( \bar{d} = \frac{1}{N} \sum d_i \) and variance \( \sigma^2 = \frac{1}{N} \sum (d_i - \bar{d})^2 \).
3. Establish a threshold \( \theta = \bar{d} + \alpha \sigma \).
4. For each point \( p_i \), if its K neighbors have an average distance \( d_i > \theta \), delete the point.

In this article, the number of adjacent points analyzed for each point is set to 50, and the multiple of standard deviation is set to 1, the point is marked as an outlier Point and be cleared. In PCL, you can use the Statistical Outlier Removal filter to complete the filtering, and the filtering result is shown in Figure 3.

![Figure 3. TOF data noise reduction processing](image)

In addition to this method for accurate seed data, it is also possible to accurately remove the edge of the point cloud data obtained by the TOF. If the time is too long or located at the boundary, the distance result is inaccurate, which is not suitable as a seed point for matching.

2.4. Principles of Regional Growth

Traditional region growth is the process of selecting pixels with similar characteristics in the image and merging them into regions. The specific steps are: first select an initial seed point from the target area, and use the area growth criterion to determine whether the surrounding pixels are merged into the grown area. The new pixels in the grown area are used as seed points, and the judgment process is repeated. When no new pixels are included in the area, the growth is stopped, and a segmented area is formed.

3. Algorithm ideas

The idea of region growth is the process of combining pixels or sub-regions into a larger region according to a predefined growth criterion. The basic method is to start with a set of "seed" points, add neighboring pixels similar to the seed (grayscale or color in a specific range) to the seed stack, and iterate continuously to generate a large area. In this research, the specific process of the region growing algorithm improved for a specific fusion algorithm is as follows.
Step 1: Select the initial seed point $S_0$, and use $S_0$ as the 0th generation seed. In this research, first use the filtered and denoised point cloud data obtained by TOF, map them to the corrected left and right images, and use the points mapped to the image as the seed point set.

Step 2: Find the first-generation seed, use the following formula to determine the connected pixels of $S_0$, and put the pixels that meet the conditions into it. In the search, the basis of the determination is the similarity of colors, expressed as:

$$S(p, q) = \sum_{j}(d_j^p - d_j^q)^2$$

In the formula, $d_j^p$ and $d_j^q$ are the values of pixels $p$ and $q$ on the color channel $j$, respectively. If $S(p, q)$ is less than the predetermined threshold $T$, it is determined that $q$ belongs to the next generation seed.

Step 3: Take $i=1$, take each pixel $S_i$ from $r[i]$ in turn as the $i$-th generation seed. In its connected pixel $c_i$, start to calculate the eligible pixels and put it into $r[i+1]$. At the same time, it is necessary to determine whether the pixel $S_i$ has been determined to have appeared in the support area. If it has been determined, the pixel is no longer considered in the subsequent calculations, and $i$ is incremented by 1.

Step 4: After completing the area growth of the seed point, return to Step 1, and then extract a seed point for area growth. If laboratory conditions permit, all seed spots can also be grown in parallel at the same time.

Take the assumed point cloud data obtained from TOF and map it to a point in the right image, and set the thresholds to 5, 20, 30, 40, and 80 respectively. With this point as the initial point, the result of region growth is shown in Figure 4.

![Original right image](image1)
![5](image2)
![20](image3)
![30](image4)
![40](image5)
![80](image6)

Figure 4. Results of using the region growing algorithm at different thresholds

It can be seen from the figure that if different thresholds are selected, the area of the obtained growth area is different, and from a time comparison, $30 > 40 > 30 > 20 > 5$, this is because the larger the threshold, the larger the growth area, and the longer the calculation time. As can be seen from Figure 4, if you want to obtain a suitable area for matching, you need to select a suitable threshold. In the matching process, using the idea based on the SSD function, for the $m \times n$ block matching window in the image, the error energy function expression is constructed as:

$$E_d(x, y) = \frac{1}{3nm} \sum_{x=0}^{m} \sum_{y=1}^{n} \sum_{i=1}^{3} (L(x, y, c) - R(x-d, y, c))^2$$

(2)
The reliability function $R_d$ of the disparity map $D_p$ is evaluated by the error energy of the disparity map, and its expression is:

$$R_d = \frac{1}{mean(E_d - (ne))} = \frac{1}{S_d} \left( \sum_{(i,j) \in E_d(x,y) = ne} E_d(x,y) \right)^{-1} \tag{3}$$

It can be seen from the formula that when the average error energy of the disparity map is smaller and the value is larger, the disparity map is more reliable.

4. Analysis of results
For the algorithm proposed in this article, three pairs of images are selected to show the matching effect, the time taken and the reliability.

4.1. Under disparity maps with different thresholds
The matching effect is as follows:

Fig. 5(a) is a disparity map with a threshold value $D=5$ based on the region growing method based on the fusion of binocular and TOF, and Fig. 5(b) is a disparity map with a threshold value $D=20$. The experimental parameters are $n=1$ and $m=5$.

![Figure 5. Comparison of matching results with a threshold of 5 and 20](image)

| Threshold | TIME(s) | Reliability |
|-----------|---------|-------------|
| 5         | 14.59   | 0.00437     |
| 20        | 20.40   | 0.00572     |

From the comparison between Figure 6 and Table 1, it can be seen that under different thresholds, the generated disparity maps are very different, and they are not all the same in time. It can be seen that the quality of the disparity maps is more reliable when the threshold is 20. Larger, the effect is better, but from the time point of view, the effect of the threshold value is 5 is better.

4.2 For three different scenarios, set the threshold to 5 hours
The matching effect is as follows: Figures 6(a)-(c) are disparity maps of the region growing method based on the fusion of binocular and TOF. The disparity maps of the algorithm used in this paper are displayed in three scenes respectively. The experimental parameters respectively are $n = 1, m = 5$. 

![Images of disparity maps](image)
Figure 6. Matching results of three scenarios

It can be seen from Figure 6 that the algorithm proposed in this paper has high accuracy and high matching quality in the boundary area of the object and the area with large depth difference. It has good results in matching time and reliability. The experimental results also show that the method can obtain a effect under a suitable threshold, and the time it takes is longer, as shown in Table 2.

| Algorithm name                  | Group | TIME  | Reliability |
|---------------------------------|-------|-------|-------------|
| The algorithm proposed in this   | 1     | 20.79 | 0.004377    |
| article                         | 2     | 20.91 | 0.005398    |
|                                 | 3     | 20.40 | 0.004371    |

The algorithm proposed in this paper is far superior to the depth data obtained by other TOF in terms of quality and point density, and the time used is far less than that required by the monocular and binocular stereo vision system. The algorithm in this paper has strong advantages. One of the main advantages of TOF seeds over points of interest is that they are regularly distributed in the image, regardless of the presence of texture. The distribution of points of interest largely depends on the texture and lighting conditions, but the actual situation is not the case. In a hybrid system, combining the active vision method and the passive parallax method can overcome the limitations of the active and passive distance methods.

5. Conclusion
In this paper, the fusion method of the binocular stereo vision measurement system and the TOF measurement system is improved. In order to make the results accurate, the binocular and TOF are calibrated before starting the measurement, and then the data obtained by the TOF camera is filtered and denoised. Use the processed information as a factor of regional growth to guide binoculars to match. Experiments have proved that the method can complete fast dense reconstruction on the basis of ensuring accuracy, and has a certain degree of robustness, and can have good effects in scenes such as texture and non-texture areas and repeated texture areas. The algorithm combines the advantages of TOF technology and binocular stereo vision, and makes up for the shortcomings of the respective algorithms, and the data shows that the algorithm has good real-time performance.
Acknowledgement

This work was partially supported by National Key Research and Development Program of China under Grant(NO. 2018YFE0205803)

References

[1] Banks J, Bennamoun M. Reliability Analysis of the Rank Transform for Stereo Matching. IEEE Transactions on Systems Man and Cybernetics Part B-Cybernetics, Vol.31, NO.6,870-880,2001.

[2] Ge L, Liang H, Yuan J. Robust 3D Hand Pose Estimation from Single Depth Images using Multi-View CNNs[J]. IEEE Transactions on Image Processing. Vol.99, No.4, 1,2018.

[3] Abdel-Aziz Y I, Karara H M. Direct Linear Transformation from Comparator Coordinates into Object Space Coordinates in Close-Range Photogrammetry[J]. Photogrammetric Engineering & Remote Sensing. 2015

[4] Chen B, Jung C, Zhang Z. Variational Fusion of Time-of-Flight and Stereo Data for Depth Estimation Using Edge-Selective Joint Filtering[J]. IEEE Transactions on Multimedia, Vol.20, No.11,2882-2890,2018.

[5] Wu Habin,Zhou Yunjun,Zhou Yingwei,et al Image segmentation using region growing base on 2D OTSU to selected seed points[J].Journal of Atmospheric and Environment Optics, 2013,8(6):448-453

[6] Zhang Le, Xiang An.An improved region growing algorithm and its application of foreign substances inspection[J].Nondestructive Testing,2010,32(7). 490-492.

[7] Wang Haoming,Li Delong,Yang Xiaoyuan,et al.A detection method of image forgery based on growing triangle areas[J].Journal of Wuhan University,Natural Science Edition,2011,57(5):455-460.

[8] Wan SY,Higgins WE.Symmetric region growing[J].IEEE Transaction on Image Processing,2003,12(9):1007-1015