

High-precision Cone Center Point Extraction Method Based on Stereo Vision Feature Matching

Bowen Shi^{1, a}, Zhen Liu^{2, b}, Jingyu Zhang^{2, c}, and Shoubo Yang^{2, d}

¹*School of Beihang University, Beijing 100191, China*

²*School of Beihang University, Beijing 100191, China*

^a*shibowen1@126.com*, ^b*liuzhen008@buaa.edu.cn*, ^c*ZhangJingY@buaa.edu.cn*, ^d*sbyang@buaa.edu.cn*

Keywords: feature matching, autonomous aerial refueling, double Gaussian function, center point.

Abstract: In order to achieve high-precision docking of airborne refueling in close proximity, this paper proposes a high-precision cone center point extraction method based on stereo vision feature matching. Based on the detection and tracking results of the target cone, the method performs extended compensation on the tracking frame of the cone to form a region of interest (ROI). With camera calibration results and feature matching which is performed on the ROI, a depth map of the cone can be obtained. With the ideology of double Gaussian function, the center offset is solved according to the distortion ratio of the cone ring width to locate the cone center point. The method does not need to add any cooperation logo to the cone, and is easier to implement in the engineering application; it is not necessary to fit the cone-shaped circle to solve the center point, which can be directly solved; compared with other methods, the precision is greatly improved. This method has been implemented in ABB robot simulation system. The experimental results successfully verify the feasibility and effectiveness of the method.

1. Introduction

Automated aerial refueling (AAR) technology has become one of the hotspots in the field of UAV. The application of drones not only reduces the casualties of drivers, but also greatly improves the feasibility of military strikes. The UAV's self-propelled aerial refueling system is divided into two types: hose type and hard tube type [1]. This paper mainly uses hose cone type refueling tube as the research object. The difficult breakthrough of autonomous aerial refueling technology is the accurate measurement of the position and attitude of the cone. In the middle and long distance part of the docking phase, the position measurement of the cone is mainly completed by the multi-information navigation mode such as binocular stereo vision navigation. However, in the close-up docking phase, most of the visual methods are to ensure the spatial positioning accuracy of the target by adding an optical cooperation logo on the oil pipe or the cone. This kind of method is not only complicated to install, but is seriously affected by optics. Once the marker is occluded or blurred, it will cause the cone to be lost or even misdetected.

Therefore, unmarked visual measurement systems are particularly important. Most methods without optical cooperation mark replace the center point with the cone tracking frame point or the fitting cone center, but the theoretical center point in the actual docking process should be the oil pipe nozzle in the cone. The point accuracy is much lower than the method with optical cooperation mark[2,3], which results in the docking is limited to the oil pipe and needs to slide into the cone at the right angle to complete the docking. Therefore, it is necessary to propose a method that does not rely on the optical cooperation logo and improve the accuracy of the cone center point, which lays a higher foundation for the subsequent development of the autonomous air refueling technology.

This paper discusses the coordinate solution method of the center point of the target cone. The main contributions are as follows:

1) We propose a compensation mechanism based on cone detection and tracking, and improve the processing speed through image segmentation technology.

2) We propose a feature matching technique based on the ROI, and extract the coordinates of the center point from the double Gaussian function ideology by extracting the depth values of the matching points in the horizontal and vertical directions of the depth map from the generated dense 3D point cloud.

3) We propose to re-model the problem that the partial matching point is too sparse due to the limitation of the feature matching algorithm, and optimize the center point coordinate result according to the neighborhood matching point and the ring prior knowledge to achieve high precision requirements.

The structure of this paper is as follows: First, the proposed method is described in detail. Secondly, the measurement results of the whole method are given through experiments. Finally, the summary and follow-up of the paper are given.

2. Methodology

The main purpose of this paper is to accurately measure the position of the cone center point during the close-up docking of the autonomous air refueling. Our approach is primarily based on a two-camera stereo vision system: the first phase is target detection, tracking and segmentation. The second stage is to solve the cone center point of the ROI area depth map. The third stage is model optimization.

2.1 The First Stage.

Adaboost target detection is an iterative algorithm, which combines multiple weak classifiers together to filter as a strong classifier [4]. Although it is not as accurate as deep learning, its speed is fast enough to guarantees the real-time performance of KCF tracking algorithm, and the framework is more easy to carry out hardware porting[5]. In order to improve the image processing speed, we generally perform different degrees of preprocessing according to the image size, and compress the large image by adaptive coefficients. The compression ratio is defined as

$$resize_scale = \begin{cases} 1, & width < 200; \\ 2, & 200 \leq width < 400; \\ 3, & 400 \leq width < 600; \\ 4, & width \geq 600. \end{cases} \quad (1)$$

Due to the compression, some of the feature information is affected to different extents, which makes the cone recognition process so difficult that the tracking frame area does not cover the

complete cone. This can have a serious impact on subsequent feature matching results. To solve the problem, we compensate and expand the tracking frame area according to the two values: compression coefficient and mesh size to form a ROI that covers the complete cone. According to the gray distribution of the actual ROI, the approximate region of the image segmentation is determined by the apparent difference gray value of the pixel on the step edge, and the depth values of the useless regions other than the ring cone are all set to zero. A suitable threshold t divides it out.

2.2 The Second Stage

Feature matching is to calculate the similarity between image features to complete the matching[6]. The algorithm architecture is mainly to establish a grid of nine squares, the center point is the matching point to be judged, the grid pattern is rotated in eight directions with an angle $\theta = 45^\circ$, and the best matching point is selected by the maximum value of the numerical distribution curve[7].

Multiply the pixel coordinates of the matching point with the matrix of the calibration result to obtain a dense three-dimensional point cloud[8]. The schematic diagram is as follows

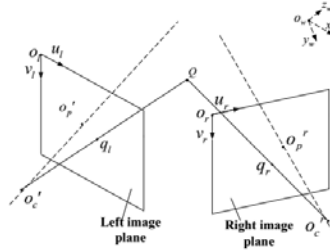


Fig. 1 binocular stereo vision measurement model

We extract the depth value of the dense three-dimensional point cloud to form a depth map, and extract all the depth values of the matching points on the midline in the horizontal direction and the vertical direction, and obtain the following point sets and perform double Gaussian function fitting[9].

Because the cone will be distorted during the image acquisition process, such as the right offset of the cone relative to the double camera, the ring pitch of the left cone is thicker than that of the right cone, and the center point is also closer to the left side, the closer the ratio of the left and right peaks of the double Gaussian function is inversely proportional to the distance between the center point and the left and right side rings. According to the span ratio, the ring pixel diameter and the pixel coordinates of the upper-left corner of the ROI, the coordinates of the center point can be solved by the following formula.

$$\begin{cases} X_{center} = X_{box} + [X_2 + (X_3 - X_2 - 1) \times (X_4 - X_3 + 1) / (X_2 - X_1 + X_4 - X_3 + 2)] \times 2 \\ Y_{center} = Y_{box} + [Y_2 + (Y_3 - Y_2 - 1) \times (Y_4 - Y_3 + 1) / (Y_2 - Y_1 + Y_4 - Y_3 + 2)] \times 2 \end{cases} \quad (2)$$

where X_n and Y_n are the inner and outer ring scatter coordinate values at both ends of the horizontal and vertical , X_{center} and Y_{center} are cone center point coordinate value, X_{box} and Y_{box} are ROI box upper-left coordinate value.

2.3 The Third Stage

Since feature matching is easily affected by the background and texture. Matching points are sparse in some place. This situation will result in a decrease in the accuracy of the center point position, or even an error. We can extract the matching points of the dense cone near the center line.

We know that the edge gradient change of the elliptical ring is regular. The concave function, although the gradient is a variation, the derivative of the absolute value of the gradient is a fixed value. From this we can find the gradient value near the midline, and perform the region fitting to compensate the sparse point set. According to the image sequence of the cone in the simulation environment, we select the horizontal direction neighborhood value $X_{range} = 3$, and the vertical direction domain value $Y_{range} = 9$. At the same time, we can select the matching point closest to the outer side of the inner and outer ring edges as the Gaussian peak and valley junction point, and continue to solve the center point coordinates according to the second stage algorithm. Because the third stage uses the combination of gradient feature constraints and geometric feature constraints to reconstruct the region matching points, the non-mean sub-pixel points are inserted between adjacent pixels. The new model can improve the accuracy to the sub-pixel level through the second-stage algorithm. The results of the accuracy verification will be shown in the experiment below.

3. Physical Experiment

Based on the ABB robot IRB1410 dual-arm simulation environment, the robot arm repositioning accuracy can reach 0.02mm. This chapter will expand from three aspects: experimental environmental conditions, experimental level renderings and contrast accuracy evaluation.

3.1 Experimental Environmental Conditions

The range of motion of the IRB1410 robot arm in the simulation system is $\pm 20\text{cm}$ in the X direction, $\pm 40\text{cm}$ in the Y direction, and $\pm 10\text{cm}$ in the Z direction. The cone model has a diameter $D = 20\text{cm}$ and a depth $Z = 12\text{cm}$. The software platform configures OpenCV 3.2.0 for Matlab 2017a and VS2015. The measuring distance between the dual camera and the target cone is $L = 1.85 \sim 2.15\text{cm}$. The dual camera is a GT1920 model with a resolution of 1936×1456 and a baseline distance $B = 24\text{cm}$. The double target results are as follows.

$$\begin{aligned} \text{Focal Length:} \quad & \text{fc_left} = [2788.27261 \quad 2788.50299] \pm [1.30878 \quad 1.31786] \\ & \text{fc_right} = [2806.12065 \quad 2805.90578] \pm [1.34215 \quad 1.36266] \\ \text{Rotation vector:} \quad & \text{om} = [-0.00459 \quad -0.00033 \quad 0.00418] \pm [0.00038 \quad 0.00046 \quad 0.00003] \\ \text{Translation vector:} \quad & \text{T} = [-234.21569 \quad 0.56396 \quad 2.83945] \pm [0.05121 \quad 0.04328 \quad 0.32557] \end{aligned}$$

3.2. Experimental Level Renderings

The first level of the algorithm in this paper is to detect and track each frame of the video stream collected by the dual camera module, and display the position of the tracking frame in real time on the image. In the original tracking frame, the partial cone ring will appear outside the tracking frame. According to the displacement direction and displacement of the cone, the position and overflow of the overflow portion of the cone will be different. After the compensation mechanism algorithm is modified, the general extended pixel range: $range = \text{resize_scale} \times \text{grid} \approx 10 \text{ pixels}$, and the speed can be increased according to the constant value of the specific measured object. The area of the tracking frame after local expansion is the ROI area effect as follows

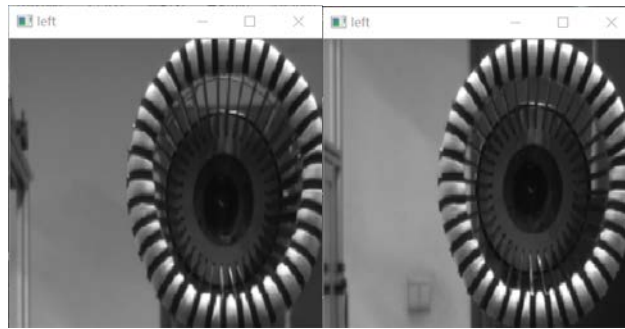


Fig. 2 uncompensated and compensated image

The second level of the algorithm is to match the ROI area of the left and right camera images. In the matching process, we can adjust the number of matching points to 1000 according to the characteristic information such as the center symmetry of the cone and the black and white phase. We will get the 2D matching point map into a dense 3D point cloud through the calibration result, and extract the depth information to generate the depth map, as shown below.



Fig. 3 Depth map

We extract the matching point depth values in the horizontal direction and the vertical direction of the depth map separately. The valley point offset is solved according to the double Gaussian distribution function, and the coordinates of the center point are solved. The effect is as follows:

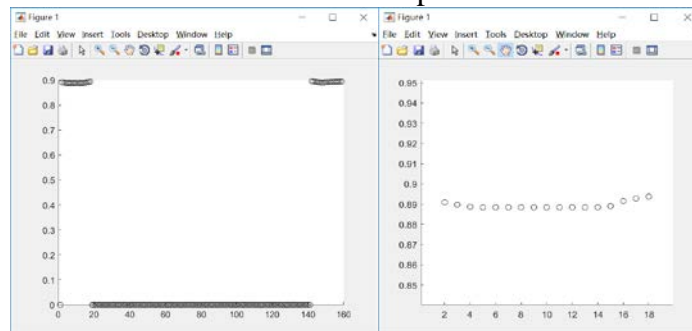


Fig. 4 scatter plot of midline depth values

The third level of the algorithm is to establish a new neighborhood fitting model for the sparse problem of partial position matching points. The center point coordinates are solved according to the new model. The effect is as follows:

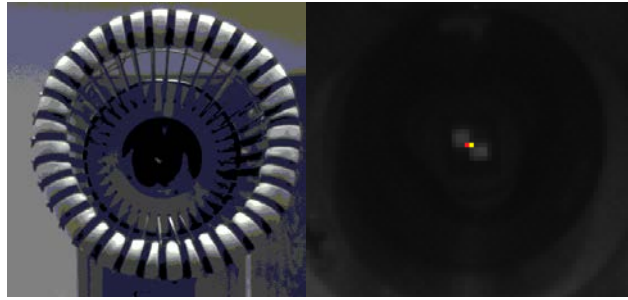


Fig. 5 true and measured value

Red dot is mark point, which represents true value :(1108, 246).Yellow dot is measured point, which represents measured value:(1108.9,246.4)

3.3 Accuracy Evaluation

In order to evaluate the degree of accuracy improvement of the proposed method in this paper, we use the robotic arm to perform verification experiments in the horizontal and vertical directions. The number of experimental images in each group is 60-150 frames, and the frame is divided according to the number of frames. Sampling calculation, about 10 sets of images are sampled in a group of experiments, a total of 10 sets of experiments and 100 pictures. Through the control variable method, the other variable factors of the fixed cone, such as the cone without rotation, the plane of the cone (ie, the posture) are consistent. It can be seen from the experimental data that the larger the displacement, the greater the accuracy improvement; the more positive the camera, the smaller the accuracy improvement. According to the results of the 100 sets of experimental data, the $RMS = 0.47\text{ pixels}$ of the vertical direction error and the $RMS = 0.78\text{ pixels}$ of the horizontal direction error are orders of magnitude higher than the accuracy of the tracking frame center method and the fitted center method.

Table 1 The RMS of experimental results

	Vertical	Horizontal
RMS	0.47pixels	0.78pixels

4. Conclusion

In this paper, we propose a high-precision cone center point extraction method based on stereo vision feature matching. Firstly, according to the traditional target detection and tracking result, the compensation algorithm is used to obtain the region of interest and perform image segmentation. Then, feature matching is performed on the region of interest, and the conversion of the dense three-dimensional point cloud to the depth map is completed. Then, the two-way midline neighborhood point fitting is performed according to the defect of the depth map and the prior knowledge of the geometric characteristics of the cone ring. Finally, according to the new fitting model, the double-Gaussian function is applied to the depth value distribution of the two-way midline matching point to solve the central point offset, and the coordinates of the center point position are calculated. The experimental results have proved the improvement of the accuracy of the method.

At the same time, the system built by the method can be installed on the oil receiving machine of the self-propelled aerial refueling, and can accurately position the oil pipe nozzle in the cone in the short-distance docking stage, instead of merely staying in the positioning cone ring, which is more reliable and safe. Complete the docking process. In addition, it can be applied to the tank to the

receiving port in the process of rigid pipe autonomous aerial refueling. In the future, we can further improve the speed of the software side in the detection and matching process, and at the same time realize parallelization on the hardware side to improve the efficiency and improve the application of the algorithm.

Acknowledgments

This work was supported by the National Natural Science Foundation of China under Grant No. 51575033. This work was also supported by the Co-Construction Project of Scientific Research and Postgraduate Cultivation-Achievements Transformation and Industrialization Program-On-line dynamic detection system of train pantograph-catenary operating conditions provided by the Beijing Municipal Education Commission.

References

- [1] Nalepka J, Hinchman J. *Automated Aerial Refueling: Extending the Effectiveness of UAVs*[C]// *AIAA Modeling and Simulation Technologies Conference and Exhibit*. 2013.
- [2] Duan H, Zhang Q. "Visual Measurement in Simulation Environment for Vision-Based UAV Autonomous Aerial Refueling". *IEEE Transactions on Instrumentation & Measurement*, vol 64, Sept. 2015, pp. 2468-2480. doi: 10.1109/TIM.2014.2343392.
- [3] Wilson D B, Göktoğan A H, Sukkarieh S. "Guidance and Navigation for UAV Airborne Docking", *Robotics: Science and Systems*. 2015. Doi: 10.15607/RSS.2015.XI.017.
- [4] Wang J, Gao L, Zhang H, et al. *Adaboost with SVM-Based Classifier for the Classification of Brain Motor Imagery Tasks*[M]// *Universal Access in Human-Computer Interaction. Users Diversity*. Springer Berlin Heidelberg, 2011:629-634.
- [5] Hannuna S, Camplani M, Hall J, et al. *DS-KCF: a real-time tracker for RGB-D data*[J]. *Journal of Real-Time Image Processing*, 2016:1-20.
- [6] Li Y, Snavely N, Huttenlocher D P. *Location recognition using prioritized feature matching*[C]// *European Conference on Computer Vision*. Springer-Verlag, 2010:791-804.
- [7] Bian J, Lin W Y, Matsushita Y, et al. *GMS: Grid-Based Motion Statistics for Fast, Ultra-Robust Feature Correspondence*[C]// *IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, 2017:2828-2837.
- [8] Long H M, Guo H Y, Liang F, et al. *Distance Measurement Algorithm Based on Binocular Stereo Vision*[J]. *Applied Mechanics & Materials*, 2014, 635-637:948-952.
- [9] Rasmussen C E, Nickisch H. *Gaussian Processes for Machine Learning (GPML) Toolbox*[J]. *Journal of Machine Learning Research*, 2010, 11(6):3011-3015.
- [10] Bryll R K. *Global calibration for stereo vision probe: US, US 8055466 B2*[P]. 2011.