Cone Sleeve Center Measurement Algorithm based on Multi-information Integration

Bowen Shi*, Jingyu Zhang and Shoubo Yang
School of Beihang University, Beijing 100191, China

*Corresponding author: shibowen1@126.com

Abstract. In order to meet the high-precision docking requirements of autonomous aerial refueling, this paper proposes a high-precision and fast positioning algorithm for the center of the cone sleeve based on two-dimensional and three-dimensional multi-information fusion. The algorithm performs target detection and tracking on the cone sleeve in the image after calibration and calibration, and adaptively selects the region of interest in the positioning area; the region of interest is calibrated based on the gray-scale constrained two-dimensional model and the three-dimensional model based on the depth gradient. The two-dimensional and three-dimensional information fusion result is back-projected and compensated to the region of interest to realize the high-precision center point position measurement of the center point. The algorithm is used in the semi-physical simulation experiment on the ABB mechanical arm to verify its feasibility, robustness, fast and high precision and other characteristics.

1. Introduction
At present, the common hose aerial refueling system is mainly divided into two categories: with and without cooperation mark. Cooperative logos are generally LED lights, fluorescent pigments, and coding patterns. For example, Li Borui and others proposed a monocular vision method based on feature tracking. A number of auxiliary identification points were set in the tanker body in advance. Feature points iteratively estimate the relative pose information [1]. Delin LUO et al. Proposed a binocular vision system (BVS) based on green LED and filter to estimate the pose of the refueling cone [2]. The non-cooperative mark is usually obtained by edge fitting to find the center or the center of the tracking frame. For example, Christopher Parsons proposed to process the composite image of the stereo camera system through computer vision algorithms and sense the navigation pose [3]. Mammarella, M. et al. Proposed EKF for combining GPS and MV-based position data to provide a reliable estimate of the relative position of the tanker-drone over the entire docking and refueling phase [4]. Most of the existing methods are affected by the measurement accuracy, which makes it impossible to meet the docking requirements between drones in practical applications.

This article focuses on the more convenient non-cooperative identification algorithms. In order to better meet the actual installation conditions of the aircraft, we use the parallax method with two cameras placed side by side in parallel. The main disadvantage of this type of algorithm is that the accuracy of the solution of the center point is low, and the installation environment is basically the receiving machine, which is very unfavorable for the monitoring of the tanker.
In response to the above problems, this article proposes the following solutions:
1) Adaptively adjust according to the pixel size of the target cone, and shrink to the ROI area of the fixed value pixels in the target tracking frame according to the level function.
2) In the ROI region, perform adaptive threshold segmentation on two-dimensional image information, and simultaneously extract edge information. Weighted and superimposed the dual information to obtain the proportion and distribution of the cone region and the hose region in the ROI region.
3) Within the ROI region, perform parallax stereo matching of the new ROI region based on the pixel coordinate information of the ROI region, and obtain the depth gradient information in the region. The adaptive tube segmentation and the cone sleeve part are extracted and obtained the proportion and distribution of the two in the new ROI area.
4) Based on the fusion of the proportion and distribution information in the two-dimensional and three-dimensional ROI regions, a calibration model is established to obtain the center point compensation vector of the dual camera image cone sleeve, and the center point is relocated.

The structure of this article is as follows: First, the principle of each module of the measurement system is described in detail. Secondly, the feasibility and accuracy of the algorithm are verified through physical simulation experiments. Finally, the summary and future outlook of the paper are given.

2. Methodology
The measurement system includes camera calibration, target tracking, center calibration and 3D reconstruction. Among them, the offline camera calibration uses Zhang's checkerboard calibration, and the 3D reconstruction is mainly based on the SGBM parallax method.

2.1. Detection and Tracking Stage
The KCF-Adaboost algorithm combined with the existing detection and tracking is used to meet the real-time and portability requirements of the embedded terminal. The target detection uses the Adaboost algorithm window scan to initialize the target tracking and positioning. In order to meet the real-time requirements of the detected image features, this article cancels the feature selection, and only uses the Haar features that are most suitable for the cone sleeve texture. Ada + KCF fusion algorithm is used to detect and track the target cone sleeve in the calibrated image of each frame with dual cameras.

2.2. Preprocessing Stage
a. Get the pixel coordinates and width and height of the upper left corner of the tracking frame on the left and right images of each frame (u, v, width, height), as shown in Figure 2, the red box is the target tracking box.

Figure 1. Schematic diagram of parallel structure.
b. Take the union of the vertex coordinates in the v direction of the left and right tracking boxes so that the double tracking boxes are on the same horizontal line \((U (v_L, v_R), U (v_L + \text{height}_L, v_R + \text{height}_R))\).

c. Establish a hierarchical function to scale the tracking frame to form a local ROI area and obtain the ROI center point coordinates, that is, the center of the tracking frame.

\[
center(U_c, V_c) = \left( u + \frac{\text{width}}{2}, \left( \bigcup (v_L, v_R) + \bigcup (v_L + \text{height}_L, v_R + \text{height}_R) / 2 \right) \right)
\]

\[
ROI(U, V, \text{Height}, \text{Width}) = (U_c - 50, V_c - 50, 100, 100)
\]

The actual center point of the taper sleeve should be the center of the interface between the taper sleeve and the pipeline, and the approximate points solved by the existing algorithms are not the center point of the taper sleeve. The two-dimensional image points are misaligned, and the three-dimensional coordinates are even more misaligned. Observing the taper sleeve from the rear perspective can guide the taper sleeve center points of the left and right images through the extension direction of the oil pipeline, and combining three-dimensional information is more robust and accurate than relying only on two-dimensional information. According to prior knowledge, the direction in which the oil pipeline is deflected in the ROI rectangle should be the vector direction of the center point calibration compensation. Assuming that the actual center point coincides with the center of the tracking frame, the proportion of the pipeline in the ROI rectangle should be about 1/4. If the proportion of the pipeline is small, the compensation amount is negative; if the proportion of the pipeline is large, the compensation amount is positive.

2.3. 2D Calibration Stage

In the pre-processing stage, the new ROI area generated in step c is mainly divided into two parts, the oil pipeline area and the taper sleeve metal joint area. Due to the obvious difference between the two gray levels, the gray value is different in different environments.

a. Separate the two into binary states of 0 and 1 by means of adaptive threshold.

\[
k = \omega_1 \times \omega_2 \times \left( \mu_1 - \mu_2 \right)^2
\]

\[
\text{rate} = \frac{\text{area}_{f>k}}{\text{area}_{f=k}}
\]

The ratio of the number of pixels belonging to the foreground to the entire image is recorded as \(\omega_1\), and its average gray scale \(\mu_1\); the ratio of the number of background pixels to the entire image is \(\omega_2\), and its average grayscale is \(\mu_2\); Inter-class variance is recorded as \(g\).

b. At the same time, the arc-shaped edge of the pipeline at the conical sleeve connection is extracted by the canny operator, and the pipeline area and the cone area are divided according to the edges.
c. Determine the direction and magnitude of the center point compensation vector by establishing a vector model.

\[
\bar{C} = \sum_{i=1}^{4} (m_i, n_i) \times rate
\]

(5)

where \(X_v\) and \(Y_v\) 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. \(\bar{C}\) is compensation vector.

2.4. 3D Calibration Stage

A. Extract the \((u, v)\) coordinates of the upper left and lower right corners of the new ROI region in the new ROI region generated in step c in the preprocessing stage. Since the depth and the parallax are inversely proportional, the minimum parallax step size can be set according to the maximum depth \(d_{max} = V_{max} - V_{min}\). It can be known from prior knowledge that the largest depth in the figure is the edge point of the cone sleeve, and the target is to the right in the left image and to the left in the right image.

B. Transform the matching area and search area of the SGBM algorithm according to the disparity search step size and the new ROI area coordinates, and convert the semi-local matching into local matching under dynamic relative coordinates [6]. Double-target correction is performed on the cone image in the ROI region, so that the horizontal direction of the left and right images is parallel. The semi-global matching algorithm of the ROI region mainly includes the following four core steps: selecting matching primitives (preprocessing); gradient cost calculation and SAD cost calculation; constructing the cost energy and function of the multi-directional scan line and obtaining the optimal solution; parallax Calculation and rejection of mismatches (post-processing) [7,8]:

1) Preprocessing: The image is processed by the horizontal Sobel operator to obtain the gray value of the first-order discrete difference result.

\[
Sobel(x, y) = 2[P(x+1, y) - P(x-1, y)] + P(x+1, y-1) - P(x-1, y-1) + P(x+1, y+1) - P(x-1, y+1)
\]

(6)

2) Cost calculation: The following two costs are calculated in the SAD window:

The sampling gradient cost is to select the maximum point of the bell curve matching probability through the minimum difference. The cost function is as follows

\[
C(x, y, d) = \sum_{i=-n}^{n} \sum_{j=-n}^{n} |L(x+i, y+j) - R(x+d+i, y+j)|
\]

(7)

3) Multi-directional constraint matching: Perform the least-cost path calculation for the eight directions connecting the other eight points in the eight neighborhood centered on point M, as follows:

\[
L_c(p, d) = \min \{L_c(p-r, d), L_c(p-r, d-1) + P_r L_c(p-r, d+1) + P_r \min_i L_i(p-r, i) + P_r \}
\]

(8)

\[
P_r = 8 \times \text{cn} \times \text{WindowSize}^2 \quad \quad P_r = 32 \times \text{cn} \times \text{WindowSize}^2
\]
cn is the number of channels of the image, WindowSize is the size of the SAD window, and the values are odd and all are constants.

4) Post-processing: Iteratively calculates parallax and removes mismatch points through uniqueness and consistency detection to form a parallax map.

C. Converting the parallax value after stereo matching into a depth value

\[
\begin{align*}
    x_i &= f \frac{X}{Z} \\
    y_i &= f \frac{Y}{Z} \\
    x_r &= f \frac{X - T_x}{Z} \\
    y_r &= f \frac{Y}{Z}
\end{align*}
\]

D. Adaptively select a suitable depth cutoff value according to the depth value in the new ROI region. Since the depth value of the pipeline to the taper sleeve increases slowly and non-linearly, this has a serious impact on the adaptive depth threshold. Will cause the selected threshold to be a global median value, and the median value cannot represent the dividing line between the pipeline and the taper sleeve. The most significant difference in the depth information between the taper sleeve and the oil pipeline is that in the case of the U-horizontal two-dimensional image being equidistant, the change in the depth of the oil pipe is much larger than the depth change of the taper sleeve. The depth gradient is set, so the adaptive threshold is selected according to the depth gradient to make the threshold more consistent with the actual situation. The boundary between the oil pipeline and the cone sleeve is segmented, and a three-dimensional information compensation vector is obtained.

\[
\begin{align*}
    K_{3D} &= \text{grad}(depth(x, y, z)) \ast k \\
    \text{Rate}_{3D} &= \frac{\text{space}_{t \leq K_{3D}}}{\text{space}_{t < K_{3D}}}
\end{align*}
\]

Where I representing the depth gradient.

E. According to the two- and three-dimensional information fusion compensation vectors, the center points of the left and right image tracking frames are compensated and calibrated respectively, and the new center points of the left and right images after calibration are three-dimensionally reconstructed to obtain the high-precision positioning of the cone center point. Among them, the segmentation of the new ROI region and the acquisition of compensation vectors at the two-dimensional image level all take milliseconds. The only time-consuming ones are three-dimensional information-level SGBM stereo matching, depth gradient segmentation, and compensation vector acquisition. The average time is About 21ms, the overall time is less than 33ms, which can meet the technical parameters of autonomous aerial refueling for more than 30 frames.

3. Physical Experimental

The camera uses AVT GT1920c, with a resolution of 1936*1456, a focal length of 12cm, a baseline distance between the two cameras of 24cm, an adaptive exposure. The double target results are as follows [9].

Focal Length: \( fc \_\text{left} = [2788.27261, 2788.50299] \pm [1.30878, 1.31786], \text{fc} \_\text{right} = [2806.12065, 2805.90578] \pm [1.34215, 1.36266] \)

Rotation vector: \( \text{om} = [-0.00459, -0.00033, 0.00418] \pm [0.00038, 0.00046, 0.00003] \)

Translation vector: \( T = [-234.21569, 0.56396, 2.83945] \pm [0.05121, 0.04328, 0.32557] \)
The simulation robot selects dual ABB robots to work together. The models are IRB1410, the tracking accuracy is ± 0.05mm, the movement range is up to 1.44m, the load capacity is 5kg, the simulation cone diameter is 20cm, and the weight is 3.74kg, which meets the experimental conditions, as shown in Figure 3. This chapter will expand from three aspects: experimental environmental conditions, experimental level renderings and contrast accuracy evaluation.

![Simulation environment](image)

**Figure 3.** Simulation environment.

### 3.1. Experimental Level Renderings
The algorithm in this paper performs initial positioning of the cone sleeve based on the target detection and tracking results, and then obtains the local ROI region, and calibrates the center point of the ROI with high precision through the fusion of two- and three-dimensional information. The uncalibrated cone sleeve positioning algorithm replaces the cone sleeve center with the center of the tracking frame, the deviation is too large, and generates random errors with the drift of the positioning area, as shown in Figure 4(a) below.

![Uncalibrated center point](image)

![ROI adaptive threshold segmentation](image)

![2D calibration results](image)

**Figure 4.** (a) Uncalibrated center point. (b) ROI adaptive threshold segmentation. (c) 2D calibration results

Among them, the red cross point is the center point of the tracking frame, the yellow point is the center point of the cone sleeve, the green point is the two-dimensional calibration point, and the green frame is the calibration reference range. Obviously, the center point after only two-dimensional calibration is very close to the actual center point of the cone sleeve. Since the above video stream frame image is only taken by a single camera, three-dimensional calibration cannot be performed, so the three-dimensional calibration effect is only based on semi-physical simulation experiments.

According to the semi-physical simulation experiment, the center point of the two-dimensional calibration has a horizontal error RMS = 6.3pixel and a vertical error RMS = 3.4pixel, as shown in Figure 5 (a) below. The disparity map formed after stereo matching through the SGBM algorithm is adaptively thresholded and brought into the calibration model to further correct the two-dimensional calibration. The RMS error in each direction can reach the sub-pixel level. The effect is shown in Figure 5 (b).
Figure 5. (a) 2D calibration effect. (b) Two-dimensional and three-dimensional combined calibration effect.

3.2. Accuracy Evaluation
In order to verify the accuracy and robustness of the algorithm proposed in this paper, we took the form of a semi-physical simulation experiment to verify, control the robot arm to move in the horizontal and vertical directions, and conduct ten sets of experiments, each group of video stream length is about 100 seconds. The number of sampling frames is 30, and the center of the joint between the simulated oil pipe of the robot arm and the cone sleeve is taken as the center point of the cone sleeve, which is the origin of the coordinate system of the robot arm workpiece. From the experimental data, it can be seen that the positioning accuracy of the center point of the cone sleeve is very high, and the precision of the displacement of the center point is very high, which effectively solves the problem of the low accuracy of the existing algorithms for positioning the center point of the back of the cone sleeve.

| Table 1. Center point positioning accuracy. |
|--------------------------------------------|
| Vertical | Horizontal |
| RMS      | 0.97pixel  | 0.59pixel  |

| Table 2. Center point displacement accuracy. (RMS=1.1%) |
|--------------------------------------------------------|
| Numble | standard value | Measurements | error |
|--------|----------------|--------------|-------|
| 1      | 100mm          | 99mm         | 1%    |
| 2      | 100mm          | 98mm         | 2%    |
| …      | …              | …            | …     |

4. Conclusion
This paper proposes a two-dimensional and three-dimensional multi-information fusion cone sleeve center point fast and high-precision positioning algorithm. First, based on the calibration and calibration image, the ROI area is compensated for two-dimensional and three-dimensional adaptive models, and the compensation vectors for the left and right two-dimensional image levels and the three-dimensional depth gradient information level are obtained respectively. Then, the left and right two-dimensional compensation vectors are respectively fused and superimposed on the three-dimensional compensation vectors to calibrate the center points of the two-dimensional cone sleeves of the left and right images. Finally, the three-dimensional position coordinate information of the center point is obtained through three-dimensional reconstruction. The physical simulation experiment has verified that the algorithm has improved measurement accuracy and point-up accuracy compared with other methods, and at the same time, it also meets the practical application requirements of autonomous air refueling in terms of speed. In the future, we will focus our research on high-precision measurement of the position and velocity of the cone sleeve.
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] Li Borui, Mu Chundi, Wu Botao. Vision-based automatic near-air refueling relative position and pose estimation [J]. Journal of Tsinghua University (Natural Science Edition), 2012 (12): 10-15.
[2] W. Luo D , Shao J , Yang X U , et al. Docking navigation method for UAV autonomous aerial refueling[J]. Science China(Information Sciences), 2019, 62(01):29-44.
[3] Parsons C , Nykl S . Real-Time Automated Aerial Refueling Using Stereo Vision[C]// International Symposium on Visual Computing. Springer International Publishing, 2016.
[4] Mammarella M , Campa G , Napolitano M R , et al. Machine Vision/GPS Integration Using EKF for the UAV Aerial Refueling Problem[J]. IEEE Transactions on Systems, Man and Cybernetics, Part C (Applications and Reviews), 2008, 38(6):791-801.
[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] Bryll R K. Global calibration for stereo vision probe: US, US 8055466 B2[P]. 2011.