Research on the Application of Binocular Vision-Based Pose Estimation Technology in Live-Line Working

Jianmin Wang, Guangnian Xu*, Baijun Wu, Jiangjun Dai
Zhejiang Huayun Information Technology Co., Ltd., Hangzhou, Zhejiang 310000
*yeiliang@hyit.com.cn
*Corresponding Author: e-mail:13706830777@139.com

Abstract. To engage robots in live-line working and reduce the cost of electric power companies, this study proposes a binocular vision-based pose estimation method that can automatically identify the target’s pose, guide the robotic arms to grasp the tools and fulfill live-line working tasks. First, this method matches and calculates the deep information of images via the binocular camera, and converts the images into the cloud of points; then, the advanced image segmentation technology is used to recognize the tools, and 3D modelling is conducted for the objects to accurately match the target image and the real target; the iterative closest point (ICP) method is used to solve the spatial pose of the target corresponding to the robot and thus guide the robots in tasks. Meanwhile, experiments are conducted to simulate the actual work process, and the result shows that the proposed method can accurately estimate the pose of targets and has good robustness and application value.

1. Introduction
As life improves, government institutions, mines and people’s daily life rely heavily on electricity power, and grid upgrading entails outage, which disturbs people’s constant needs of power and brings about pressure to power supply companies. To strengthen power grid maintenance and make timely remedies to power supply problems, establish a security defense system of power grids, and address all challenges, power companies need to “keep the power on whenever possible” during power grid maintenance tasks to ensure stability and reliability of power supply.

Live-line working generally includes wire connection, wire cutting, electrical grounding ring installation, lighting protection wire clamp installation, clearing of foreign matters in wires, etc. These tasks take up 90% of the workload in daily live-line working, and these tasks need different tools and have different procedures, so a live-line worker should receive different types of training and the errors are likely to occur in on-site operations. During live-line working, multiple workers need to cooperate, which is difficult, time- and labor-consuming. Also, the workers are constantly under threats of high voltage and high electric fields; and required to work at heights, they are constantly under stress and are susceptible to safety accidents. Among the 100 typical live-line working accidents recorded in 1995 in China, 55 were related to power distribution live-line working, taking up 55% of the total; and among the 39 deaths in these accidents, 19 were from power distribution live-line working, accounting for 61% of the total causalities.

Traditional tasks on power distribution networks rely purely on human labor, which is time- and resource-consuming, inefficient, and susceptible to safety hazards. Development of AI and computer vision technologies makes it possible for automatic and intelligent live-line working. For instance,
during inspection of the high-voltage cable lines, inspection of overhead lines has been an important measure to ensure normal operation of the power system. However, it has long suffered problems like wide distribution of networks, high workload and intensity, long period and harsh natural conditions. Therefore, UAV-assisted inspection [1], robot-assisted inspection [2], online monitoring [3] and other new technologies have been applied to inspection of electric power pylons, wires, metals, insulators. Zhai et al. used the method based on color and grey gradient features to detect and locate insulators, and assessed the defects based on morphological analysis, reaching a high accuracy of 92.4%; Zhai et al. proposed an F-PISA clustering positioning method based on the color and structural features [5], which realized detection of insulators and reached an accuracy of 78.8%. Hao et al. [6] used the Grabcut method to extract insulators and quantitatively analyzed the icing condition of insulators based on the raising defects of the cable. Huang et al. proposed a method for wire inspection robots to detect obstacles under weak illumination [7], which conducts adaptive homomorphic filtering on the images, reduces the impact of illumination, divides the image of the obstacles evenly into different areas, extracts the histogram vectors using the local orientation mode, and then classifies the whole histogram based on the Chi-square distance. Their experiment results showed that this method could effectively recognize suspension clamps, insulator chains and dampers. For work on insulation circuit-breakers in transformer substations, the computer vision technologies are applied, such as the method proposed by Zhang et al. [8], a method that uses computer vision technology to guide robot arms to connect and disconnect the insulation circuit-breakers and customizes the insulation circuit-breaker label points for image fitting and positioning. Zhu et al. [9] proposed a gesture detection method based on deep target detection to locate the gesture of the insulation circuit-breakers and guide the robots to move.

Image processing and AI technologies have been applied to power network distribution and other electricity-related work, and realized a certain degree of automation. In live-line working tasks, there are no relevant studies, and to realize automatic live-line working, robots are needed to connect and disconnect the wires, and computer vision-assisted guiding and positioning technology have become a major support. In the present study, we attempted to explore the application of binocular vision-based pose estimation in live-line working robots. Through deep estimation of the binocular camera images, the point cloud diagram is generated; image segmentation of the target tools is conducted to match the images in the existing target tool 3D model library and finally to realize target pose estimation by solving the iterative closest point.

2. Binocular Stereo Vision Matching and Deep Estimation

2.1. Binocular Vision Model

Binocular vision is a technology that uses two cameras to simulate the imaging effect of human eyes, analyzes the differences in the images of the same target, and obtains the geometric relations through triangulation and computes the deep information of the images from the two cameras. Figure 1 shows a binocular vision model. As Figure 1 shows, \((O_L, O_R)\) are the optical centers of the two cameras, \(T\) is the horizontal distance between the two cameras, \(P\) is a givenpoint, with its coordinates being \((X_C, Y_C, Z_C)\), and its pixel coordinates of the two cameras are \((μ_L, ν_L)\) and \((μ_R, ν_R)\); \(P_L\) and \(P_R\) are the abscissa of the projected point of \(P\) on the left and right images, respectively.
According to the geometric relations obtained through trigonometric survey, the projecting location and depth of Point P in the pixel coordinates in the left and right images are obtained:

$$\frac{T}{Z_c} = \frac{Z_c + X_r - X_l}{Z_c - f}$$  \hspace{1cm} (1)

Then, the distance is:

$$Z_c = \frac{Tf}{d}$$  \hspace{1cm} (2)

where \(d\) is the parallax, \(f\) is the focal length of the camera, \(Z_c\) is the distance from Point P to the focus. According to Eq(2).

2.2. Binocular Calibration and Image Correction

In engineering practice, the focus lengths of the two cameras are hardly the same, and the installation and lenses of the cameras cannot meet the ideal conditions; the image is distorted and correction is needed to make remedy [10].

Binocular calibration is a process to initialize the parameters of the binocular measuring system, and a process to solve the camera’s parameters and correct image distortions. The checkboard calibration method [11] has good robustness, and the Zhengyou Zhang calibration method [12] do not need the calibration target to move the parameters, which can meet the reliance requirements of the calibrated objects.

We took several checkboard images from different angles using two cameras, and used the Zhengyou Zhang calibration method to obtain the camera’s parameters, as shown in Table 1.

| Parameters                                      | Value                        |
|-------------------------------------------------|------------------------------|
| Intrinsic factor of the left camera             | \[
|                                               | \begin{bmatrix}
|                                               | 574.8 & 0 & 314.1 \\
|                                               | 0 & 574.9 & 191.8 \\
|                                               | 0 & 0 & 1 \\
|                                               | \end{bmatrix} \] |
| Distortion factor of the left camera            | \[0.036, 0.293, -0.0001, -0.005\] |
| Intrinsic factor of the right camera            | \[
|                                               | \begin{bmatrix}
|                                               | 579.3 & 0 & 313.9 \\
|                                               | 0 & 579.5 & 176.9 \\
|                                               | 0 & 0 & 1 \\
|                                               | \end{bmatrix} \] |
| Distortion factor of the right camera            | \[0.095, -0.428, -0.0038, -0.0011\] |
| Rotation factor of the two cameras               | \[-0.015, 0.009, -0.001\] |
| Translation factor of the two cameras            | \[-64.741, 0.036, 0.522\] |

2.3. Stereo Matching and Deep Estimation

One point in space is projected in the two cameras, and stereo matching can be conducted to find the projected pixel points of the point in the two images. The algorithm is to establish an energy cost function, and by minimizing this function, we can estimate the parallax value of the pixel points, so it
is in essence an optimization problem. By the number of matching points, the stereo matching algorithm can be divided into sparse matching and dense matching. Sparse matching extracts Haar angular points [13], SIFT [14] and SURF [15] and other features in the two images; and obtains the sparse parallax image through feature matching, and the dense parallax image through interpolation. For dense matching, the child window of a given point in the image is selected, and similar child windows in another image are found based on similarity, and the pixel points in the matched child window are taken as the matching points of the pixel, thus to obtain the dense parallax image. By the method of optimization, stereo matching can be divided into universal stereo matching, semi-universal stereo matching, local stereo matching. Scharstein et al. divided stereo matching into the following steps: cost calculation, cost aggregation, parallax optimization and parallax correction [16]. By whether the universal optimization method is used, the stereo matching algorithms can be classified into universal stereo matching algorithms and local stereo matching algorithms. Semi-universal stereo matching adopts the target function that is consistent with the universal image, and it only converts the two-dimensional optimization problem into a one-dimensional optimization problem, and solves the problem through dynamic planning. In terms of the accuracy, universal optimization outperforms semi-universal optimization, and semi-universal optimization outperforms local optimization; in terms of the time cost, local optimization has the highest efficiency, followed by semi-universal optimization and then by universal optimization.

Given the requirements for accuracy and speed, and based on the analysis of the stereo matching optimization method, the stereo matching of live-line working tools used the semi-universal stereo matching algorithm. In the experiment, with the live-line working tool as the object, the algorithm is implemented, and the parallax images are obtained, as shown in Figure 2.

![Fig. 2 Images obtained by stereo matching](image)

The cost matching in the semi-universal stereo matching uses the $C_{BF}$ algorithm proposed by Birchfield et al. [17]; with the left image as the reference, we obtain parallax value close to the matching point $I_R(x_R, y)$ in the right image:

$$I_R^1 = \frac{1}{2} (I_R(x_R) + I_R(x_R - 1))$$

$$I_R^2 = \frac{1}{2} (I_R(x_R) + I_R(x_R + 1))$$

The maximum and minimum of the matching point and the interpolation point are as follows:

$$I_{\text{max}} = \max \left(I_R^1, I_R^2, I_R(x_R)\right)$$

$$I_{\text{min}} = \min \left(I_R^1, I_R^2, I_R(x_R)\right)$$
The calculation equation for the BT cost matching of the pixel P is as follows:

\[ C_{BT}(p, d_p) = \max \{0, I_L(x) - I_{\text{max}}, I_{\text{min}} - I_L(x)\} \]  

(7)

The energy cost function of the semi-universal matching method is as follows (an extra parallax smoothness constraint is added):

\[ E(d) = \sum_p C_{BT}(p, d_p) + \sum_{q \in N_p} P_1 T[|d_p - d_q| = 1] + \sum_{q \in N_p} P_2 T[|d_p - d_q| > 1] \]  

(8)

Where the first item represents the cost of all pixel matching, the second and the third represent the penalty constraints; \( P_1 \) and \( P_2 \) represent the penalty factors of the matching point of \( N_p \) in the neighborhood.

3. **Object Segmentation in Images**

Live-working tools image segmentation is to separate the live working tools from the image, extract corresponding deep information, and generate point cloud for pose estimation. The U-Net [18] network is used in the present study to segment metal images. The segmentation effect is shown in Figure 3. The left is the working apparatus, and the red part on the right is the segmentation result.

As Figure 4 shows, the U-Net model consists of up-sampling and down-sampling; meanwhile, up-sampling and down-sampling channels are merged to extract multi-scale information; the input of the network is \( 572 \times 572 \times 1 \), and the output is \( 388 \times 388 \times 2 \). The training dataset consists of live-working tools images collected through the cameras; and the images are augmented through rotation, translation, extension, color and space transfer and other technologies to enrich the training dataset; the relations between the training loss, validation loss and the epochs are shown in Figure 5.
Figure 5 shows that the model began to converge after 800 times of training, and achieved good effect in the training dataset and the validated data.

4. ICP Pose Estimation

The iterative closest point (ICP) matching method is the most widely used matching method [19]. It has been used for 3D-3D pose estimation problems. We assume that there is a set of matching points: \( P = \{p_1, \ldots, p_n\} \), \( P' = \{p'_1, \ldots, p'_n\} \). According to a given constraint, we calculate the optimal matching parameters \( R \) and \( t \) to minimize the error function, i.e.

\[
E(R, t) = \frac{1}{n} \sum_{i=1}^{n} ||p_i - (R p_i' + t)||^2
\]  

In live-line working, the binocular vision stereo matching is used to compute the point cloud of the live-line working tools; by feature point matching with the 3D model in the database, calibration and pose computation is realized through ICP computing; the ICP problem is then solved by the singular value decomposition method. The procedures are shown as follows:

1) Extract \( p_i \in P \) in the point cloud \( P \);
2) Extract \( p'_i \in P' \) in the 3D model point cloud \( P' \) to minimize \( ||p_i - p'_i|| \);
3) Calculate the parameters \( R \) and \( t \), to minimize the error function;
4) Obtain \( R \) and \( t \) according to the previous step; and convert \( p_i \) into \( p'_i'' \);
5) Calculate the average distance of the corresponding point sets of \( p'_i \) and \( p'_i'' \):

\[
d = \frac{1}{n} \sum_{i=1}^{n} ||p'_i - p'_i''||^2
\]

6) If \( d \) is smaller than the given threshold or larger than the epoch, then the iteration stops and return to Step 2 until the requirements for convergence are met.

The solving process of ICP reveals that the key to the algorithm is the collection of the original point sets, the identification of corresponding sets, computing of the transformation matrix. The original point sets are obtained usually by random sampling and average sampling; identification of corresponding point sets is realized by point-to-point or point-to-plane projection; the transformation matrix is usually solved by the SVD method.

5. Experiment Results and Analysis

The proposed binocular vision-based pose estimation method can be applied to live-lineworking to estimate the spatial coordinates and pose of the live-line working tools, guide the robots to grasp. The system is shown in Figure 6. The binocular cameras are installed in a fixed location, and the binocular camera pose estimation system can estimate the pose of the live-line working tools to realize automatic grasp of the target objects.
As Figure 7 shows, experiments are conducted for different initial poses and positions. The experiment shows that: when the camera is above the tools (0°), the distance and angle deviation along x, y and z directions simulation are the minimum; as the angle of the camera and the plane of the tool increases, the distance and angle deviation along the three directions increase as well, as shown in Figure 7a. As Figure 7b shows, as the distance between the camera and the tool increases, the distance and the angle deviation along the three directions increase as well. Through experiments under different angles and different distances, the angle error is controlled within 2°, the positional deviation is controlled within 2 mm, which meets the work requirements and confirms the robustness of the proposed method in uncertain contexts.

6. Conclusion
To meet the needs of actual live-line working and intelligent power grid planning, the present study propose a binocular vision-based live working method, which uses binocular camera stereo matching method for deep estimation, employs advanced image segmentation technology to segment the tools; through data matching, the tool 3D model is searched, the feature points are matched to conduct ICP pose estimation. The experiment result shows that the proposed method not only meets the actual needs of live-line working, but has good adaptability and practical value.
References

[1] Zhou, G., Yuan, J., Yen, I. L., & Bastani, F. (2016) Robust real-time UAV based power line detection and tracking. 2016 IEEE International Conference on Image Processing. IEEE, 744-748.

[2] Peng, X. Y., Q. J. J., Wu, G. P., et al. (2017) Full automatic inspection system and its demonstration application based on robot for overhead transmission lines. J. High Voltage Engineering, 08:2582-2591.

[3] Jiang, T., Cai, F. D., Yang, X. J., et al. (2016) Discussion and practice of visual remote inspection of transmission line. J. Electric Power, 11:42-45.

[4] Zhai, Y., Wang, D., Zhang, M., et al. (2017) Fault detection of insulator based on saliency and adaptive morphology. J. Multimedia Tools and Applications, 76:12051-12064.

[5] Zhai, Y., Cheng, H., Chen, R., et al. (2018) Multi-saliency aggregation-based approach for insulator flashover fault detection using aerial images. J. Energies, 11:1-12.

[6] Hao, Y., Wei, J., Jiang, X., et al. (2018) Icing condition assessment of in-service glass insulators based on graphical shed spacing and graphical shed overhang. J. Energies, 11:318.

[7] Huang, L., Wu, G. P., Ye Xuhui. (2018) Obstacle identification under low-light conditions of transmission line inspection robot. J. Acta Optica Sinica, 38:291-299.

[8] Zhang, Y. J., Li, S. C., Wang, J. H., et al. (2017) Research of positioning method for assembling and de-assembling the hardware connected on isolation circuit breaker. J. Mechatronics, 23:3-10+60.

[9] Zhu, S. L., Li, B. B., Zhang, C. C., et al. (2020) Study on visual guidance and positioning method of automatic make-up and disassembly robot of isolated circuit breaker. J. High Voltage Apparatus, 07:205-211.

[10] Goldberg, S. B., Maimone, M. W., Matthies, L. (2002) Stereo vision and rover navigation software for planetary exploration. Aerospace Conference. IEEE.

[11] Tsai, R. (1987) A versatile camera calibration technique for high-accuracy 3d machine vision metrology using off-the-shelf tv cameras and lenses. J. Robotics and Automation, IEEE Journal of, 3:323-344.

[12] ZHANG, Z. (2000) A flexible new technique for camera calibration. J. IEEE Transactions on Pattern Analysis and Machine Intelligence, 22:1330-1334.

[13] Mikolajczyk, K., Schmid, C. (2004) Scale & affine invariant interest point detectors. J. International Journal of Computer Vision, 60(1), 63-86.

[14] D. Lowe. (2004) Distinctive image features from scale-invariant keypoints. J. International Journal of Computer Vision, 60(2): 91-110.

[15] Bay, H., Tuytelaars, T., Gool, L. V. (2006) SURF: Speeded up robust features. Proceedings of the 9th European conference on Computer Vision-Volume Part I. Springer-Verlag, 404-417.

[16] Scharstein, D., Szeliski, R. (2002) A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. J. International Journal of Computer Vision, 47(1-3): 7-42.

[17] Birchfield, S., Tomasi, C. (1999) Depth discontinuities by pixel-to-pixel stereo. J. International Journal of Computer Vision, 35(3), 269-293.

[18] Ronneberger, O., Fischer, P., Brox, T. (2015) U-Net: Convolutional Networks for Biomedical Image Segmentation. J.

[19] Besl, P. J., McKay, N. D. (1992) A method for registration of 3-d shapes. Proceedings of SPIE the International Society for Optical Engineering, 14(3), 239-256.