Advances in Artificial Intelligence, Big Data and Algorithms
G. Grigoras and P. Lorenz (Eds.)

doi:10.3233/FAIA230896

Tracking and Positioning of Operation and Maintenance Personnel in Distribution Substations and 3D Reconstruction of Their Joint Points Based on Multiple-View Stereo

Wenzhong CHENG, Yuexing HU[1], Lizhen ZHANG, Lixun ZHAO and Zhihua LI
Skill Training Center of State Grid Shanxi Electric Power Company Limited

Abstract. It is fundamental to ensure the safety of distribution substations that the operation and maintenance personnel complete the standardized actions in the specified position. In this paper, a method of three-dimensional (3D) tracking and positioning the operation and maintenance personnel is proposed by collecting two-dimensional (2D) images, extracting 2D joint data, and fusing them to produce 3D joint data. Firstly, the calibration target is used to calibrate the multiple cameras installed in the distribution substation. Secondly, the YOLO-Pose model is adopted to extract the joint coordinates of multiple human 2D skeletons. Thirdly, the matching of different operation and maintenance personnel in different cameras is realized based on the epipolar constraint relationship of binocular cameras. The combination of the distance error obtained by the epipolar constraint relationship and the confidence level of 2D joint coordinates is used for the selection of two cameras to reconstruct 3D joint points of operation and maintenance personnel based on the principle of multiple-view stereo. The study will lay a foundation for the behavior identification and evaluation, behavior inversion of operation and maintenance personnel, and the establishment of electronic fences in distribution substations.

Keywords. Distribution substation operation and maintenance, Human joint coordinates, Multiocular cameras, YOLO-Pose model, Epipolar constraint relation

1. Introduction

As the power system changes in source, network, and load, the operation of the electric grid presents new characteristics. Instead of a human-based operation and maintenance mode, there is a lot of research and application of robotics and automation technology in distribution substations. Xia et al.[1] proposed a remote permitting system for 10 kV distribution substations based on face recognition technology, which improves the efficiency and digitalization of the grid by remote control with face recognition cameras. Cai et al. [2] and Bai[3] used inspection robots to complete the inspection tasks of indoor equipment in distribution substations. Wang et al.[4] proposed a cloud robot system

[1] Corresponding Author: Yuexing HU, Skill Training Center of State Grid Shanxi Electric Power Company Limited, Email: huyuexing99@163.com.

architecture and identified the components of the distribution cabinet with the YOLOv5s algorithm. Based on the special requirements of the robot on duty in the distribution room for the accuracy and stability of the visual program, Qin[5] proposed the recognition and location algorithm for square distribution switch and round handcart switch based on binocular vision. Almasoudi[6] implements a variety of artificial intelligence and deep learning approaches in order to foresee and predict the corrective measures that will be conducted in response to faults that occur inside the power distribution network of the Grid station in Tabuk city with regard to users. Pirbodaghi[7] presents a comprehensive solution to monitoring and inspection of power-line distribution system, including two robotic platforms for inspection of lines and transmission tower respectively. Considering the possible threat to life in the event of dangerous behaviors, operators are required to strictly adhere to the distribution rules. However, it is labor-intensive and inefficient to view videos through manual dispatching, so video-based behavior recognition automatically is very important. Li[8] detected the human body through the network structure of Fast-RCNN, established the skeleton model, located the joint points of limbs through the Alpha Pose model, and proposed a method for identifying dangerous behaviors based on the spatiotemporal characteristics of joint points. In distribution substations, behavior recognition based on human joint characteristics mainly relies on the obtained 2D joint data, which cannot achieve tracking and localization of the human body. Although 3D data of joints can be acquired by depth cameras, they are limited to close distance, and the problem of human body occlusion still cannot be solved. Therefore, the paper proposes a method for tracking and positioning operation and maintenance personnel and 3D reconstruction of joint points in distribution substations based on the multiple-view stereo.

2. System Function Modules

The function modules of the tracking and positioning of personnel and 3D joint data acquisition system based on the multi-view stereo are shown in Figure 1.

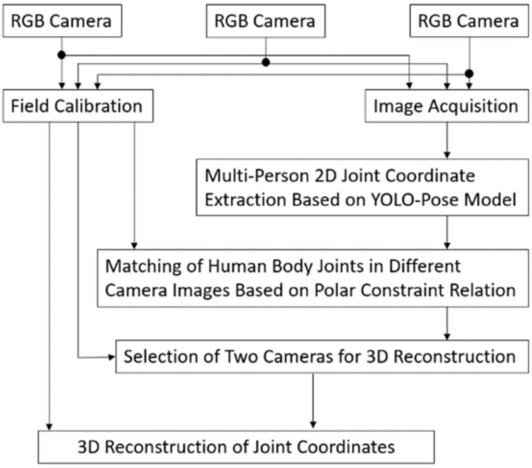

Figure 1. Function module of operation and maintenance personnel positioning and 3D joint point data acquisition system based on the multi-view stereo.

First, multiple RGB cameras are installed according to the area to be monitored in the distribution substation. Second, the calibration targets are used to calibrate the cameras on site and establish the conversion relationship between each camera coordinate system and the world coordinate system. Third, the 2D images of each camera are acquired at a rate of 25 frames per second. Fourth, the YOLO-Pose model is adopted to extract 2D joint coordinates of multiple human bodies in 2D images. Fifth, based on the epipolar constraint relationship, the identity and joint coordinates of different operation and maintenance personnel in different cameras are matched. Sixth, the two cameras for 3D reconstruction are preferentially selected according to the polar constraint relationship and the confidence value of 2D joint coordinates, and the calibration results are used to realize the 3D reconstruction of joint coordinates.

3. Principle of Multiocular Camera Calibration and 3D Reconstruction of Joint Coordinates

3.1. Multiocular Camera Calibration

A checkerboard calibration board with a grid size of 50*50mm and an array of 15*11 is selected as the calibration target, 2D images of each camera are acquired at 20 positions within the monitoring area. The multiocular camera system is calibrated by Zhengyou Zhang calibration[9] to obtain the internal parameters M, external parameters R|t, homography matrix H, and distortion coefficients $\boldsymbol{k_1}, \boldsymbol{k_2}$, $\boldsymbol{p_1}$, and $\boldsymbol{p_2}$ of each camera.

$$\mathbf{M} = \begin{bmatrix} \boldsymbol{f_x} & \boldsymbol{\gamma} & \boldsymbol{u_0} \\ \mathbf{0} & \boldsymbol{f_y} & \boldsymbol{v_0} \\ \mathbf{0} & \mathbf{0} & \mathbf{1} \end{bmatrix}, \mathbf{R} = \begin{bmatrix} \boldsymbol{r_{11}} & \boldsymbol{r_{12}} & \boldsymbol{r_{13}} \\ \boldsymbol{r_{21}} & \boldsymbol{r_{22}} & \boldsymbol{r_{23}} \\ \boldsymbol{r_{31}} & \boldsymbol{r_{32}} & \boldsymbol{r_{33}} \end{bmatrix}, \mathbf{t} = \begin{bmatrix} \boldsymbol{t_1} \\ \boldsymbol{t_2} \\ \boldsymbol{t_3} \end{bmatrix} \tag{1}$$

Where (u0, v0) is the coordinates of the principal point, $\boldsymbol{f_x}$and $\boldsymbol{f_y}$ the scale factors in image u and v axes, and γ the parameter describing the skewness of the two image axes, the extrinsic parameters (R, t) is the rotation and translation which relates the world coordinate system to the camera coordinate system. The camera intrinsic matrix H and the relationship between a 3D point (X, Y, Z) and its image projection (u, v) is given by:

$$\mathbf{H} = \mathbf{M}[\mathbf{R}|\mathbf{t}] = \begin{bmatrix} \boldsymbol{h_{11}} & \boldsymbol{h_{12}} & \boldsymbol{h_{13}} & \boldsymbol{h_{14}} \\ \boldsymbol{h_{21}} & \boldsymbol{h_{22}} & \boldsymbol{h_{23}} & \boldsymbol{h_{24}} \\ \boldsymbol{h_{31}} & \boldsymbol{h_{32}} & \boldsymbol{h_{33}} & \boldsymbol{h_{34}} \end{bmatrix} \qquad \mathbf{s}\begin{bmatrix} u \\ v \\ \mathbf{1} \end{bmatrix} = \mathrm{H}\begin{bmatrix} \boldsymbol{X} \\ \boldsymbol{Y} \\ \boldsymbol{Z} \\ \mathbf{1} \end{bmatrix} \tag{2}$$

where s is an arbitrary scale factor.

Assuming γ = 0, The distortion correction model is as follows:

$$\begin{bmatrix} \boldsymbol{\bar{u} = u + u[k_1(u^2+v^2) + k_2(u^2+v^2)^2] + [p_1(3u^2+v^2) + 2p_2uv]} \\ \boldsymbol{\bar{v} = v + v[k_1(u^2+v^2) + k_2(u^2+v^2)^2] + [p_2(3u^2+v^2) + 2p_1uv]} \end{bmatrix} \tag{3}$$

Where (u, v) is the ideal image coordinates, and $(\boldsymbol{\bar{u}}, \boldsymbol{\bar{v}})$ is the corresponding real observed image coordinates, $\boldsymbol{k_1}, \boldsymbol{k_2}$ are the coefficients of the radial distortion, $\boldsymbol{p_1}$, and $\boldsymbol{p_2}$ are the coefficients of the tangential distortion.

3.2. The Principle of 3D Reconstruction of Joint Coordinates

Assuming that the points of the 3D human joint P in the images of the two cameras C1 and C2 are P1 and P2, from Equation (2), we have:

$$\mathbf{s1}\begin{bmatrix} u_1 \\ v_1 \\ 1 \end{bmatrix} = \begin{bmatrix} h_{11}^1 & h_{12}^1 & h_{13}^1 & h_{14}^1 \\ h_{21}^1 & h_{22}^1 & h_{23}^1 & h_{24}^1 \\ h_{31}^1 & h_{32}^1 & h_{33}^1 & h_{34}^1 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix}, \mathbf{s2}\begin{bmatrix} u_2 \\ v_2 \\ 1 \end{bmatrix} = \begin{bmatrix} h_{11}^2 & h_{12}^2 & h_{13}^2 & h_{14}^2 \\ h_{21}^2 & h_{22}^2 & h_{23}^2 & h_{24}^2 \\ h_{31}^2 & h_{32}^2 & h_{33}^2 & h_{34}^2 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} \quad (4)$$

Where $(u_1, v_1, 1)$ and $(u_2, v_2, 1)$ are the homogeneous coordinates of P1 and P2 in their respective images, and (X, Y, Z,1) is the homogeneous coordinate of P in the world coordinate system. Four linear equations about x, y, and z are obtained by eliminating s1 and s2:

$$\begin{aligned} (u_1 h_{31}^1 - h_{11}^1)X + u_1(h_{32}^1 - h_{12}^1)Y + u_1(h_{33}^1 - h_{13}^1)Z &= h_{14}^1 - u_1 h_{34}^1 \\ (v_1 h_{31}^1 - h_{21}^1)X + v_1(h_{32}^1 - h_{22}^1)Y + v_1(h_{33}^1 - h_{23}^1)Z &= h_{24}^1 - v_1 h_{34}^1 \\ (u_2 h_{31}^2 - h_{11}^2)X + u_2(h_{32}^2 - h_{12}^2)Y + u_2(h_{33}^2 - h_{13}^2)Z &= h_{14}^2 - u_2 h_{34}^2 \\ (v_2 h_{31}^2 - h_{21}^2)X + v_2(h_{32}^2 - h_{22}^2)Y + v_2(h_{33}^2 - h_{23}^2)Z &= h_{24}^2 - v_2 h_{34}^2 \end{aligned} \quad (5)$$

Since the spatial point P satisfies Equation (5), the 3D point coordinate (X, Y, Z) of P can be obtained by the least square method.

4. Coordinate Extraction of Human 2D Skeleton Joints Based on YOLO-Pose

In this paper, a YOLO-Pose human pose estimation model based on the YOLOv5 target detection framework [10] is selected to detect the pose of operation and maintenance personnel. YOLO-Pose is a new Heatmap-free, end-to-end trained method with both the advantages of top-down and bottom-up methods for human joint point detection, which can simultaneously detect the bounding boxes of multiple people and their corresponding 2D poses. YOLO-pose outperforms current methods when tested on the COCO dataset. YOLOv5 is the leading performance target detector for human pose estimation, in which human pose estimation can be viewed as a single-class human detection problem with 17 joint points per person, and the coordinates and confidence of each joint point $\{\mathbf{x}, \mathbf{y}, \mathbf{conf}\}$ is provided.

5. 3D Reconstruction of Joint Points of Operation and Maintenance Personnel

5.1. Multi-person Target Match Based on Epipolar Constraint Relation of Binocular Camera

The joint coordinates and confidence of different operation and maintenance personnel detected from different camera images are respectively expressed as follows: $x(i)_n^m, y(i)_n^m, conf(i)_n^m$, where $i = 1 : 17$ is the serial number of 17 detected joint points, $n = 1 : \mathbf{N}$ indicates the serial number of the operation and maintenance personnel in the substation; $\boldsymbol{m} = \mathbf{1}$: $\mathbf{M}$ expresses the serial number of the camera installed in the substation.

In the fourth section, the YOLO-Pose model is adopted to identify the human body in each camera image and the corresponding seventeen 2D joint coordinates. However, before 3D reconstruction, it is necessary to identify the joint points corresponding to each operation and maintenance personnel in different cameras. Therefore, this paper proposes a method to match the joint points of different personnel in each camera based on the epipolar constraint relationship of binocular cameras.

As shown in Figure 2, the centers of the two cameras are $\boldsymbol{O_1}$and$\boldsymbol{O_2}$, respectively, and the corresponding pixels of point P in 3D space are $\boldsymbol{p_1}$ and $\boldsymbol{p_2}$ in two camera images $\boldsymbol{I_1}$ and $\boldsymbol{I_2}$, respectively, and the connecting lines $\boldsymbol{O_1P}$ and $\boldsymbol{O_2P}$ will intersect at the space point $\mathbf{P}$ in three-dimensional space. The three points $\boldsymbol{O_1}$ $\boldsymbol{O_2}$, and $\mathbf{P}$ can determine a plane called the polar plane. The connection of $\boldsymbol{O_1}$ $\boldsymbol{O_2}$ is called the baseline and its intersections with image planes $\boldsymbol{I_1}$ and $\boldsymbol{I_2}$ are $\boldsymbol{e_1}$ and $\boldsymbol{e_2}$, respectively, which are called poles. The intersection lines $\boldsymbol{l_1}$、 $\boldsymbol{l_2}$ between the polar plane and the two image planes $\boldsymbol{I_1}$ and $\boldsymbol{I_2}$ are called polar lines.

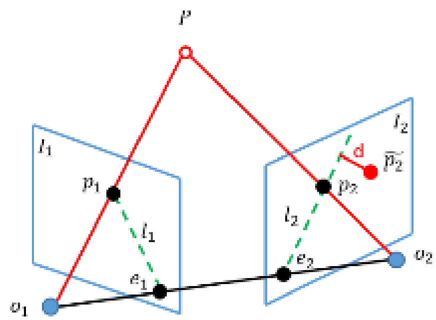

Figure 2. Schematic diagram of joint point matching based on epipolar constraint relation

If we know that the corresponding pixel of the 3D joint point P in the first camera image $\boldsymbol{I}_{\mathbf{1}}$is $\boldsymbol{p_1}$, the specific position of the $\mathbf{P}$ in the space cannot be determined, and the corresponding position of $\mathbf{P}$ in the second camera image $\boldsymbol{I_2}$ is unknown. But because the point $\mathbf{P}$ is on the ray $\boldsymbol{O_1P}$, it can be known that $\boldsymbol{p_2}$ must be on the straight line $\boldsymbol{e_2p_2}$. Therefore, the matching degree between the detected point $\widetilde{\boldsymbol{p_2}}$ in the second camera and the detected point $\boldsymbol{p_1}$ in the first camera image $\boldsymbol{I_1}$ can be measured by finding the distance from $\widetilde{\boldsymbol{p_2}}$ to the straight line $\boldsymbol{e_2p_2}$.

The matching steps for different personnel in the two cameras based on the epipolar constraint relationship of the joint points in binocular cameras are as follows:

(1) Calculating the parameters of the equation for the polar line located in the second camera image $\boldsymbol{I_2}$ corresponding to the first joint point ($\boldsymbol{x}(\mathbf{1})_{\mathbf{1}}^{\mathbf{1}}$, $\boldsymbol{y}(\mathbf{1})_{\mathbf{1}}^{\mathbf{1}}$) of the first person in the first camera image $\boldsymbol{I_1}$.

(2) Calculating the distance $\boldsymbol{d}(\mathbf{1})_{\mathbf{1}}$ from the first joint point of the first person in the second camera image $\boldsymbol{I_2}$ to the epipolar line obtained in step 1.

(3) Repeat steps 1 and 2 to find the average distance $\overline{\boldsymbol{d_1}} = \frac{\mathbf{1}}{\mathbf{17}}\sum_{i=1}^{17}\boldsymbol{d}(\boldsymbol{i})_{\mathbf{1}}$ obtained by matching all 17 joint points of the first person in the image $\boldsymbol{I_2}$ with all 17 joint points of the first person in the image $\boldsymbol{I_1}$.

(4) Repeat steps 1-3 to find the average distance $\overline{d_n}$ $(\mathbf{n} = \mathbf{1}{:}\,\mathbf{N})$ obtained by matching all the people in the image $\boldsymbol{I_2}$ with the first person in the image $\boldsymbol{I_1}$, and the person in the image $\boldsymbol{I_2}$ corresponding to the minimum value of $\overline{d_n}$ matches the first person in the image $\boldsymbol{I_1}$.

(5) Repeat steps 1-4 to match all the peoples in the image $\boldsymbol{I_2}$ with the peoples in the image $\boldsymbol{I_1}$.

The schematic diagram of people matching principle is shown in Figure 3.

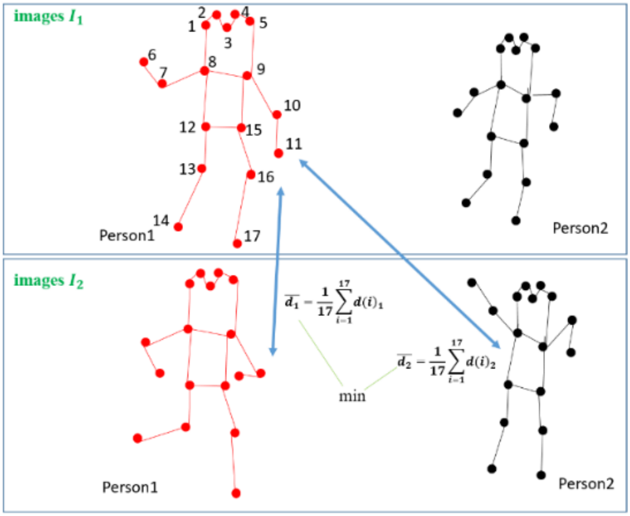

Figure 3. The schematic diagram of people matching principle

5.2. Camera Preference and 3D Reconstruction

Due to different angles taken by different cameras and different degrees of occlusion, the confidence of each joint coordinate varies greatly. To ensure the accuracy of the 3D reconstruction of joint coordinates, 2D data of the optimal two cameras should be selected. Therefore, the following camera preference objective function is established as follows:

$$\min_{j,k}\left\{\boldsymbol{a} * \frac{1}{17}\sum_{i=1}^{17} \boldsymbol{d(i)_n^{j,k}} - \boldsymbol{b} * \left(\frac{1}{17}\sum_{i=1}^{17} \boldsymbol{conf(i)_n^j} + \frac{1}{17}\sum_{i=1}^{17} \boldsymbol{conf(i)_n^k}\right)\right\} \tag{6}$$

Where $\frac{1}{17}\sum_{i=1}^{17} \boldsymbol{d(i)_n^{j,k}}$ represents the average distance to the polar line calculated from the 17 joint coordinates of the nth personnel in the jth camera and the kth camera according to the principle of section 5.1. $\frac{1}{17}\sum_{i=1}^{17} \boldsymbol{conf(i)_n^j}$ denotes the mean value of 17 joint coordinates confidence of the nth personnel in the jth camera. $\frac{1}{17}\sum_{i=1}^{17} \boldsymbol{conf(i)_n^k}$ is the mean value of 17 joint coordinates confidence of the nth personnel in the kth camera. $\boldsymbol{a}$ and $\boldsymbol{b}$ are the weight coefficients. When the value of Equation (6) is minimal, the corresponding values of j and k are the serial numbers of the two cameras selected for the 3D reconstruction of the joint coordinates of the nth personnel.

To ensure joint 3D reconstruction quality for each operation and maintenance personnel, the two cameras used for the reconstruction of each person are selected separately. Equation (5) is applied to achieve the 3D reconstruction of joint points.

6. Experimental Results and Discussion

Figure 4 shows the 3D reconstruction results of two people. As you can see in the picture, each person's 17 joints were identified and three-dimensional reconstruction was achieved. However, the reconstruction effect is not satisfactory in all cases, so the following factors affecting the reconstruction effect are analyzed

(1) The accurate identification of personnel in two 2D cameras is a prerequisite for three-dimensional reconstruction. In this paper, the binocular camera epipolar constraint relationship is used to realize the reliable matching of personnel in different cameras.

(2) Occlusion is the main factor that affects the 3D reconstruction effect of skeleton. In this paper, by screening two cameras based on the epipolar constraint relationship and the confidence level of 2D joint coordinates, and using the 3D reconstruction algorithm of binocular camera to realize the 3D reconstruction of skeleton joint coordinates to solve the occlusion problem. Therefore, the installation position and orientation of each camera should be designed according to the working location and orientation of the operation and maintenance personnel in the distribution substation.

(3) The 3D reconstruction accuracy of skeleton joint coordinates depends on the extraction accuracy of 2D skeleton joint coordinates of YOLO-Pose model, camera calibration accuracy and structural parameters of binocular camera. Among them, the base line distance, structural symmetry and focal length of the binocular camera have great influence on the reconstruction accuracy, so reasonable selection should be made according to the size of the field of view and object distance.

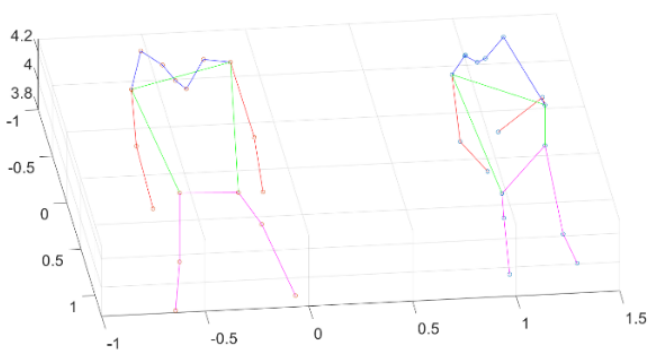

Figure 4. Joints 3D reconstruction of two people

7. Conclusion

Aiming at keeping the safety of equipment and operation and maintenance personnel in distribution substations, the paper puts forward a method of tracking and positioning the operation and maintenance personnel and 3D reconstruction of joint points based on the multiple-view stereo. The results of this paper can be used for evaluation of operation and maintenance personnel's behavior, the judgment of whether operation and maintenance personnel enter the energized area and the inversion of operation and maintenance personnel actions. Future research will focus on comparing the advantages and disadvantages of the method proposed in the paper with the method of 3D reconstruction of human joints based on deep learning with multiocular cameras, improving the generalization performance and accuracy of the model, and developing application modules such as behavior identification, assessment and evaluation related to distribution operation and maintenance.

Acknowledgements

The paper Supported by Science and Technology Project of State Grid Shanxi Electric Power Company Limited "Research and application of key technologies for intelligent analysis and evaluation of power distribution equipment operation and maintenance safety" (No. 520570220002).

References

[1] Xia Z. Y., Wang J, Guan Y. L., Chen K, Application of the remote permit system for 10kV distribution station based on face recognition technology[J], *Electric Power Safety Technology* **25(**2023),33-35.
[2] Cai C. B., Chen T., Zhao R. F, Analysis of intelligent inspection robot operation and maintenance in unattended distribution stations[J], *Zhongguancun*, **03(**2023):96-97.
[3] Bai L. J., Yan B. F., Guo D. D., Wang X. Y., Zhang T. J,Application of indoor power inspection robots in intelligent distribution stations[J], *Inner Mongolia Electric Power*, **38**(2020):67-69.
[4] Wang B., Guo Z. H., Zhuang Y. N., Han T. X., Zhang W., Liang Q. H, Design of intelligent identification system for power distribution components based on cloud architecture[J], *Mechanical Design and Research*, **38**(2022):60-64.
[5] Qin C, Research on visual positioning technology for the operation and maintenance of robots on duty in distribution stations[D]. *Shanghai Jiaotong University*, 2020.
[6] Almasoudi, Fahad M, Grid Distribution Fault Occurrence and Remedial Measures Prediction/Forecasting through Different Deep Learning Neural Networks by Using Real-Time Data from Tabuk City Power Grid, *Energies,* **16**(2023)：1026.
[7] Pirbodaghi Saeed, Thangarajan Dinesh, Liang Teo Hung, Shanmugavel, Madhavan. Ragavan, Veera. Sequeira, Joao Silva. A cooperative heterogeneous Unmanned Autonomous Systems solution for monitoring and inspecting power distribution systems. *ICCICCT 2015*, Pages 495-502.
[8] Li Z. Y, Design and implementation of hazardous behavior identification system for substation monitoring[D], *Huazhong University of Science and Technology*, 2020.
[9] Zhang Z. Y, A flexible new technique for camera calibration[J], *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(2000): 1330-1334.
[10] Debapriya Maji, Soyeb Nagori, Manu Mathew, Deepak Poddar. YOLO-Pose: Enhancing YOLO for Multi-Person Pose Estimation Using Object Keypoint Similarity Loss, *https://doi.org/10.48550/arXiv.2204.06806.*