J. COMBIN. MATH. COMBIN. COMPUT. 127a (2025) 4557-4576

Journal of Combinatorial Mathematics
and Combinatorial Computing
www.combinatorialpress.com/jcmcc

Combinatorial Press

A Safety Detection Model for Substation Operations Based on Combination of Spatial Context Reasoning Algorithm and Deep Learning Techniques

Hengjie Liu[1,✉], JieTang[2], We Li[1]

[1] Zhengzhou Electric Power Co., Ltd., Zhengzhou, Henan, 450064, China

[2,] Department of Mechanical Engineering, Henan University of Science and Technology, Zhengzhou, Henan, 450064, China

ABSTRACT

The environment near substations is complex, and electrocution accidents of operators occur from time to time during on-site operations, and the development of safety detection models for substation operations has received more and more attention. The article proposes a safety distance detection model for substation operation, which is mainly composed of binocular stereo matching perception model and safe area detection model. The binocular stereo matching perception is based on the PSMNet network model, combined with the parallax regression calculation to obtain the three-dimensional coordinates of the operation area in the process of substation operation, and the three-dimensional reconstruction of the substation operation process. The spatial context inference algorithm is utilized in the safe region detection model to detect the edge of the safe region, and the image segmentation of the safe region of the substation operation scene is performed by the improved OTSU algorithm. Then the three-dimensional coordinates obtained from binocular stereo matching perception and the three-dimensional coordinates of safe region detection are solved for the Euclidean distance, and then the safe distance detection of substation operation is realized. The EPE result accuracy of binocular stereo perception matching on the dataset is reduced by 0.71px compared with CRL, and the resulting mismatch pixel rate is between 0.83 and 1.48%. The average time-consuming image segmentation of the improved OTSU threshold segmentation method is 6.34ms, and the average relative error of the safety distance detection for substation operation is only 0.85%, and the maximum absolute error of the safety distance detection is only 0.13 m. Combining the spatial contextual reasoning algorithm with the deep learning technology can realize the effective detection of the safety distance for substation operation in multiple scenarios, and fully ensure the operation of the substation workers' safety.

Keywords: Binocular stereo, PSMNet network, spatial context reasoning, improved OTSU algorithm, safety detection model, substation operation

✉ Corresponding author.
E-mail address: L15333193316@126.com (H. Liu).
Received 25 February 2024; Revised 05 April 2024; Accepted 15 December 2024; Published Online 15 April 2025.
DOI: 10.61091/jcmcc127a-258

1. Introduction

With the growing global demand for electricity and the deepening of technological advances, substations and power grid systems are becoming more and more complex, substations as an important part of the power system, its operation site is directly related to the stable operation of the entire power system [1-2].

The existence of high voltage and strong electromagnetic field in the substation, even the slightest mistake may lead to serious electric shock or other injuries [3-4]. Life safety not only concerns the life rights and interests of every worker, but also directly affects the stable operation of the enterprise and social image [5]. Frequent occurrence of accidents will greatly reduce the credibility of the enterprise, which may lead to the loss of partners and the loss of potential business. In addition, substation operation work is to a large extent a key aspect of the economic benefits of the enterprise, and when equipment failure occurs, the immediate consequence is often damage to the equipment [6-7]. This requires urgent repair or replacement, which not only involves the cost of components, but also in the repair process, the equipment downtime will lead to production interruption or unstable power supply, all of which will bring high direct economic losses for the enterprise, and more seriously, such downtime or production interruption will further affect the upstream and downstream business [8-11]. Frequent accidents or prolonged equipment downtime will not only damage the financial problems of an enterprise, but also deal a blow to its reputation and market position. Long-term unstable production or power supply may lead to loss of customers, and new business opportunities will be affected by the instability of the enterprise [12-13]. In a highly competitive market environment, an organization may lose its competitive edge as a result, leading to a serious threat to its long-term economic prospects. However, the substation operation segment also often faces various potential risks and challenges, ranging from aging equipment, human errors to natural disasters, etc., which may pose serious threats to the substation operation work, which may not only lead to huge economic losses, but also pose a threat to public safety [14-16]. Therefore, how to ensure the safety and efficiency of substation operation has become an urgent problem in the industry.

Spatial reasoning is an important part of human intelligence, which allows us to solve problems by observation, inference and comparison.This spatial reasoning technique can help to remove the noise and supplement the missing parts in the image preprocessing stage, thus improving the accuracy and stability of the recognition [17]. Using spatial reasoning, we can accurately determine the position and relative order based on this structured contextual information [18]. In this way, more contextual features can be introduced during model training to improve the robustness and accuracy of safety recognition of substation operations.

As the core component of the power system, the substation undertakes important tasks such as urban current convergence, voltage level change, power distribution and voltage regulation, and plays a vital role in power grids at all levels. In this paper, a camera combined with machine vision technology is used to obtain on-site pictures of substation production operations, and the enhancement processing of the low-frequency components of the images is realized by guided filtering combined with edge-aware weighting. Then based on the binocular stereo matching perception model and the scene safe area detection model, the safe distance detection framework for substation operation is constructed. In the binocular stereo matching perception model, PSMNet network is mainly utilized to obtain the spatial coordinates in the substation operation scene and reconstruct it in three dimensions. The spatial context inference algorithm is used as the basis for the edge detection of the safe region of the substation operation scene, and then the improved OTSU threshold segmentation algorithm is introduced to obtain the spatial coordinates of the safe region of the scene. The 3D coordinates perceived by binocular stereo matching are combined with the coordinates of the safety region, and the safety distance of the substation operation is calculated by Euclidean distance. The performance of binocular stereo perception matching model is verified by using two datasets, and then taking the substation operation scene of a substation as an example, we analyze the execution efficiency of edge detection and threshold segmentation of the safe region, and compare and analyze the detection performance of the different methods on the safe distance of substation operation.

2. Framework for Safe Distance Detection at Substation Work Sites

In the process of electric power production and maintenance, in order to prevent the human body from contacting or excessively close to the charged body and personal injury accidents, between the human body and the charged body must maintain a certain safety distance. However, in the actual operation process, the substation in the industrial frequency electric field environment is complex, due to the high radiation on the human body to bring discomfort and high temperature or other environmental factors to bring the negative impact, is often easy to make the staff to produce paralysis, and then over the substation operation of the safety distance, to a certain extent, will lead to the staff to electrocution accidents. Based on this, this paper proposes a detection framework based on binocular vision stereo sensing and scene safety distance, which aims to better ensure the personal safety of the staff in the process of substation operation.

2.1. Image Acquisition and Processing of Transformer Operation Site

2.1.1. Acquiring images of the substation operation area.The safety detection method of substation operation designed in this paper mainly utilizes machine vision technology to collect monitoring images and videos of substation production operations, and extracts information data such as texture features of images through data information processing. The intelligent detection of safety risk of substation operation is realized by recognizing the moving target and classifying and calculating the extracted information data. The intelligent detection of the safety risk of substation operation is shown in Figure 1.

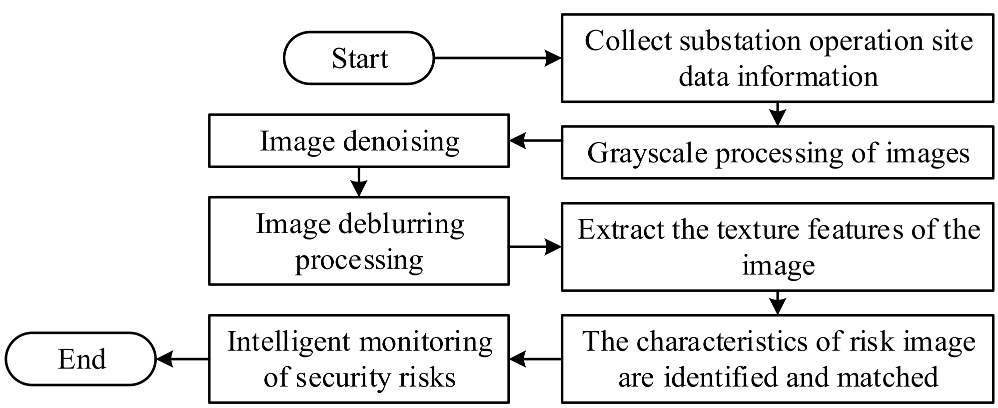


Fig. 1. Detection process of the safety distance of the substation

Data information is the basis and core of realizing all intelligent methods. The extraction of data related to the risk factors of substation production operations described above is a multi-source acquisition model. Much of the information can be obtained directly from the construction project management unit of the substation production operation and saved in the database of the substation production system. For the substation operation site data information, this paper utilizes camera equipment combined with machine vision technology to acquire the site situation of the substation production operation in the form of pictures. The core of machine vision technology is to replace the human eye for data information measurement and judgment, real-time access to the image information of the substation production site, the use of this technology for measurement and judgment, can provide basic data for the early warning of the safety risks of the substation production operation and the basis for judgment.

2.1.2. Image low-frequency component enhancement processing.In order to better realize the safety distance detection at the substation operation site, this paper aims at the problem that the image is prone to the existence of fuzzy edges or false edges, and utilizes the bootstrap filter combined with the edge-aware weights to form a weighted bootstrap filter, which realizes the enhancement of the image's low-frequency components and avoids the appearance of false or fuzzy edges in the image. The edge-aware weights can be expressed as:

$$\Gamma(q') = \frac{1}{M}\sum_{q=1}^{M}\frac{\delta^2(q')+\lambda}{\delta^2(q)+\lambda} \tag{1}$$

Where, M is the total number of pixels within the image, q' is the image cord point, $\delta^2(q')$ is the variance of the guided image within the neighborhood of the pixel point q' as the center pixel, and λ is a smaller constant $\lambda = (0.001K)^2$, where K is the dynamic interval of the input image q [19].

The pixel $\Gamma(q')$ values at the image smoothing location and edge location are different, where $\Gamma(q') < 1$ is at the smoothing and $\Gamma(q') > 1$ at the edge location, and the window ε_l has a cost of:

$$F(c_l, d_l) = \sum_{i \in R_i}\left[(c_l I_i + d_l - q_i)^2 + \frac{\omega}{\Gamma(i)}c_l^2\right] \tag{2}$$

Where c_i, d_l is the coefficient of the linear function when the center of the window ε_l is l, I_i is the position of the image, q_i is the pixel of the input image at pixel point i, ω is the regularization factor, and $\Gamma(i)$ is the perceived weight at pixel point i.

The optimal solution of the coefficients c_t and d_t can be obtained by this equation, based on which the minimum input and output image difference is obtained to maximize the preservation of the detail information of the processed image. Then:

$$\begin{cases} c_i = \dfrac{(1/|\varepsilon|)\sum_{i\in\varepsilon_i} I_i q_i - \eta\overline{q}_i}{\delta_i^2 + \omega} \\ d_i = \overline{q}_i - c_i\eta_i \\ \overline{q}_i = (1/|\varepsilon|)\sum_{i\in\varepsilon_i} q_i \end{cases} \tag{3}$$

Where, δ_l^2 is the variance of the guided image in the window, η_l is the mean value, $\overline{q}_l$ is the mean value of the input image in this window, and $|\varepsilon|$ is the total number of cords in the window.

The reflection image acquired after illumination estimation is:

$$u(x,y) = \lg[Q(x,y)] - \lg[Q(x,y)\hat{p}_i(x,y)] \tag{4}$$

where the weighted bootstrap filtering is:

$$\hat{p}_i(x,y) = \overline{c}_i H_i(x,y) + \overline{d}_i \tag{5}$$

where $H_i(x,y)$ is the position of the pixel point within the image, $\overline{c}_i$ is the positional calibration function of the original image, $\overline{d}_i$ is the error compensation parameter, and $Q(x,y)$ is the original image captured by the monitor.

2.2. Framework for detecting safety distances for transformer operations

2.2.1. Safe distance detection for transformer operations.In the field of safety distance control for substation operations in substations, image-based distance measurement methods face important challenges. First, the complex environment of substations makes it exceptionally difficult to accurately delineate safety zones. Secondly, the traditional distance calculation system is unable to effectively calculate the actual distance between it and the energized body, and thus cannot provide exact safety guarantee.

In this paper, relying on the binocular stereo sensing model, area element recognition and edge segmentation model, we propose an intelligent early warning method based on binocular stereo sensing and safety area segmentation for the safety distance of substation operation. The specific framework of this method is shown in Fig. 2. Firstly, the binocular stereo sensing model is used to obtain the horizontal displacement difference (parallax value) of the corresponding pixels between the left and right eye pictures of the substation operation process captured by the ZED camera, and the 3D coordinates of the pixels of the left eye pictures are obtained with the help of the inner and outer parameter matrices, parallax maps, and the 3D reconstruction principle of the ZED camera. Next, the edge position of the substation operation area in the left-eye picture is detected by the spatial context inference model, and the position where the substation operation staff is located is taken as the safe area. Then combine the binocular stereo perception model and the edge segmentation model to obtain the 3D coordinates of the edge contour of the staff and the safe region. Finally, the Euclidean distance between the edge contour extreme points of the substation worker in the horizontal and vertical directions and the boundary of the safe area is calculated to dynamically perceive the operation and relative positional relationship of the worker in the safe area.

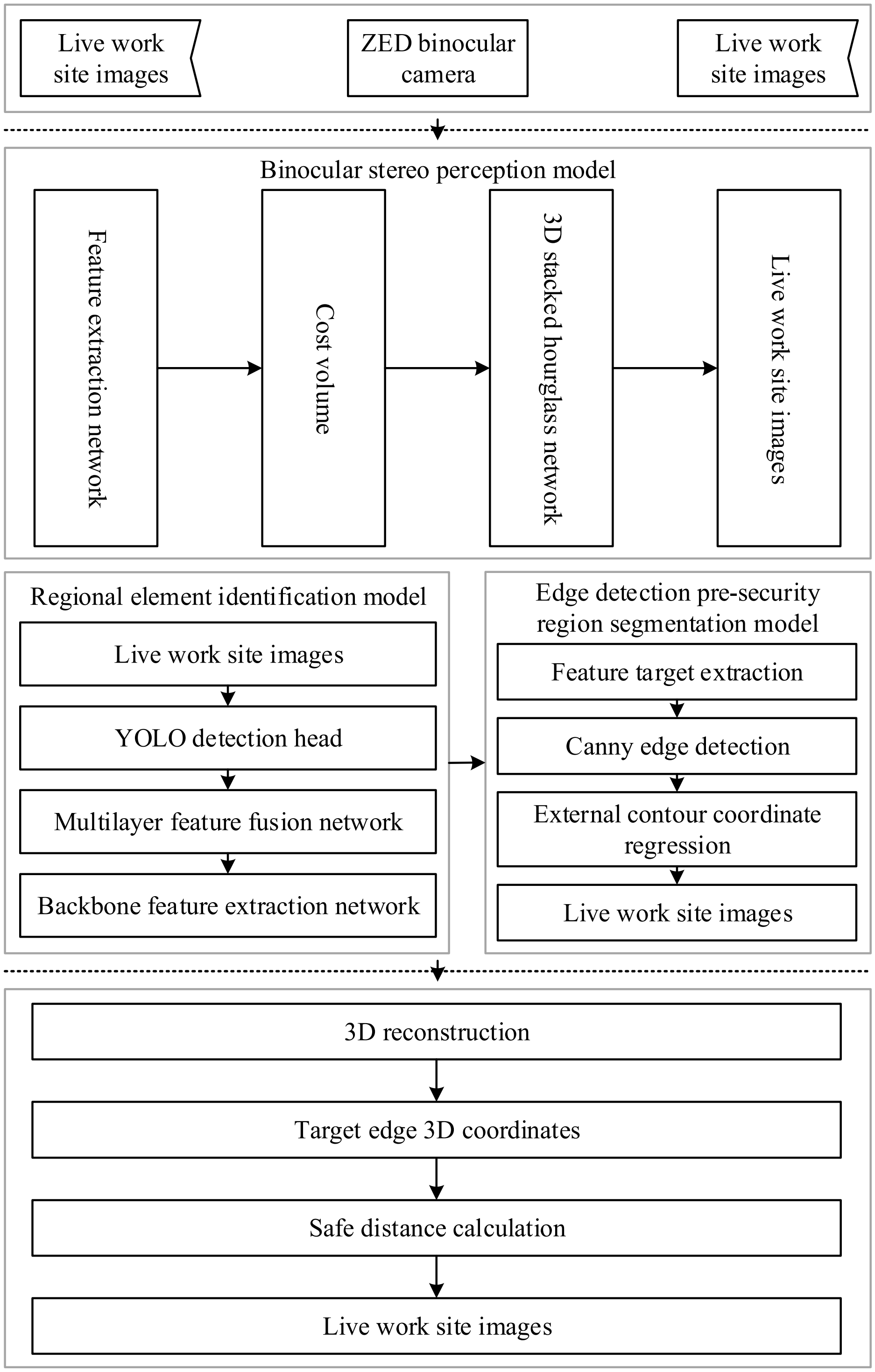


Fig. 2. Safety distance detection of near-electric operation

The method transforms the distance calculation between the staff and the energized equipment into the distance calculation between the staff and the boundary of the safe area, which can effectively deal with the problems such as the loss of the target to be detected or the misdetection of the target occurring during the calculation of the safe distance.

2.2.2. Methods for calculating safe area distances.For the safety distance detection in the process of substation operation, this paper mainly takes the three-dimensional reconstruction coordinates of the staff obtained by binocular stereo sensing as the basis, and then combines the three-dimensional coordinates of the edge contour of the safety area obtained by spatial contextual reasoning, and adopts the Euclidean distance as the evaluation standard. Then the expression of the safety distance of substation operation is:

$$\rho = \sqrt{(x' - x)^2 + (y' - y)^2 + (z' - z)^2} \tag{6}$$

Where (x', y', z') is the coordinates of a point obtained by the method described herein, (x, y, z) is the coordinates of the point obtained by the binocular stereo vision algorithm, and ρ is the calibration accuracy.

In order to ensure the consistency of the acquired three-dimensional coordinates of the edge of the staff and the safety area with the three-dimensional coordinates of the actual outer contour of the target. In this paper, the difference between the parallax value of the edge point acquired by the model and the parallax value of its corresponding local center point is made, and if the difference is less than 1/5 of the parallax value of the local center point, the edge point is spatially contiguous with its local center point and the edge pixel point is considered to be consistent with the 3D coordinates of the target edge. If the condition is not satisfied, then search and calculate point by point in the direction from the edge point to the local center point until the condition is satisfied.

3. Binocular stereo matching and scene safety region segmentation

With the growing global demand for electricity and the deepening of technological advances, substations and power grid systems are becoming more and more complex. As an important part of the power system, the operation and maintenance site of a substation is directly related to the stable operation of the entire power system. However, this link also often faces various potential risks and challenges, from equipment aging, human operation errors to natural disasters, etc., all of which may bring serious threats to the work of substation operation and maintenance. This not only leads to huge economic losses, but also may pose a threat to public safety. Therefore, how to ensure the safety and efficiency of substation O&M field operations has become a pressing issue in the industry.

3.1. PSMNet-based binocular stereo matching

3.1.1. PSMNet-based stereo sensing.PSMNet combines the 3D cost convolution constructed by comparing the left and right features, feature pyramid pooling (SPP) and hourglass structure makes the stereo matching accuracy improve dramatically, and becomes a common benchmark for stereo matching tasks. First, the left and right images are fully extracted from the binocular image using the CNN module with shared weights and the SPP module, where the convolution module consists of 3 layers of 2D convolution and 4 residual neural network blocks (Resnet), and each residual network block of the Resnet residual basis block structure contains 2 residual basis blocks, and each residual basis block contains 2 layers of 2D convolution.The Resnet network structure The cross-layer aggregation of new and original features largely alleviates the overfitting and degradation problems brought about by the excessive depth of the network layers.The SPP module extracts image features of different scales from pooling kernels with different step sizes and sizes, and the features extracted by the SPP module are able to better represent the information of different depths and sizes of the field in the image [20].

Then, the left and right features extracted by the CNN module and the SPP module are cascaded to construct the matching cost volume, forming an efficient intermediate supervision, and this form of constructing stereo features effectively shows the essence of stereo matching and improves the algorithm accuracy. Firstly, the parallax dimension is added to the left and right features, in which the left and right features are panned to simulate the construction of control feature maps with different parallax levels, and the control feature maps are spliced together to form a 3D costal volume feature with height × width × parallax, and each position of the left and right features is aligned at different parallax levels, thus improving the stereo matching effect.

Finally, the matching cost obtained in the previous step is decoded by 3D convolutional layers and 3 stacked hourglass modules. Each hourglass module contains 2 downsampled 3D convolutions, 2 3D convolutions with a step size of 1, and 2 3D inverse convolutions, and then the output features are upsampled to the original image size using trilinear interpolation. And the SoftMax function is applied in the parallax dimension to generate the probability of each position at all parallax levels, and all parallax levels are aggregated according to the probability weights to obtain a dense parallax map. All the 3D features output from the shallow hourglass are used as inputs to the deep hourglass for coarse to fine parallax prediction.

3.1.2. Parallax regression calculation with loss function.In the parallax regression stage of the PSMNet network, the Soft-argmin function is usually used to get the optimal parallax value of a pixel point within the maximum parallax range, while it can get a relatively smooth parallax map. The matching cost feature body is processed by 3D CNN and up-sampling to get the cost of each pixel in the specified parallax dimension, if the cost is larger it means more mismatch. The probability of each pixel belonging to each parallax is obtained through the SoftMax operation output, and then each parallax and its own corresponding probability value are weighted and summed to obtain the predicted parallax value for each pixel point.

Since this paper improves on the parallax dimension of the matching cost feature body, the value of K is not the maximum parallax D in the traditional sense here, but $D \times q / d$, which is defined mathematically as follows:

$$\widehat{d}_t = \sum_{k=0}^{K} d_t \times \sigma(-c_k) \tag{7}$$

where $\widehat{d}$ is the predicted parallax value, c_k is the cost value, $\sigma(\cdot)$ is the Softmax operation. K is the parallax dimension, and $d_t = k \times D / K$.

In this paper, a smooth $L1$ loss function is used as the basis function to train the network, which is more robust and insensitive to outliers, and the $L1$ loss function is often used in bounding box regression problems. At the same time, the cross-entropy loss function L_{CE} is added, which characterizes the deviation between the predicted results and the real results, and the larger value of the cross-entropy loss function indicates that the training effect is not good, while the smaller value proves that the training is more reasonable. However, there are some pixels in the Sceneflow dataset whose true parallax value exceeds the specified maximum parallax range 192, and these pixels should be excluded from the loss calculation. The mathematical definition is as follows:

$$L_{CE} = \frac{1}{N} \sum_{i=0}^{N} \sum_{k=0}^{K} -\exp\left(-| d_t - d_{gt} | / b\right) \ln\left[\sigma\left(-c_k\right)\right] \tag{8}$$

$$L(d_t, \widehat{d}_t) = \frac{1}{N} \sum_{i=1}^{N} smooth_{L_1}(d_{gt} - \widehat{d}_t) \tag{9}$$

Among them:

$$smooth_{L_1}(x) = \begin{cases} 0.5x^2, | x |< 1 \\ | x | - 0.5, otherwise \end{cases} \tag{10}$$

N represents the number of all pixel points, b is the scatter, d_{gt} represents the true parallax value of each pixel, and $\widehat{d}_t$ is the predicted value.

In summary, the loss function of the proposed network structure is:

$$L = \alpha L1 + L_{CE} \tag{11}$$

where α is the weight coefficient of the $L1$ loss function.

3.1.3. 3D reconstruction of the transformer operation process. The 3D reconstruction in binocular vision methods utilizes the principle of similar triangles to convert the parallax value obtained by the stereo matching algorithm into a depth distance value, which in turn calculates the 3D coordinates of the object [21].

Suppose $O_{c1}-X_{c1}Y_{c1}Z_{c1}$ and $O_{c2}-X_{c2}Y_{c2}Z_{c2}$ are the coordinate systems of the left and right cameras respectively, $O_w-X_wY_wZ_w$ is the world coordinate system, $O_1-X_1Y_1$ and $O_2-X_2Y_2$ are the physical coordinate systems of the left and right images, b represents the length of the optical axis between two shots, and f is the baseline. $P_1(x_1,y_1)$ and $P_2(x_2,y_2)$ are the coordinates of the projection points of the left and right image points P, d is the parallax value of the corresponding point, and W,H is the width and length of the input image, respectively. The coordinates of point P are obtained by combining the parameters of the camera calibration and the similar triangle property. Then:

$$X = Z\frac{W/2 - x_1}{f} \tag{12}$$

$$Y = Z\frac{H/2 - y_1}{f} \tag{13}$$

$$Z = \frac{f \times d}{b} \tag{14}$$

where X,Y,Z denotes the value of a point of the image on the X,Y,Z-axis of the world coordinate system, respectively.

In order to test the robustness of the improved algorithm proposed in this paper, a binocular camera is used in this paper to capture the image pairs of target objects that need to be reconstructed in three dimensions. Subsequently, the camera parameters are calibrated to obtain its internal and external parameters, and the captured target images are processed with polar correction to reduce the stereo matching time. Then the image pairs are inputted into the above stereo matching algorithm for parallax regression to obtain parallax maps, and finally the 3D coordinates of each pixel point are calculated by combining the camera parameters and the above principles to realize the reconstruction of 3D data.

3.2. Safe area detection for substation operation scenarios

3.2.1. Spatial Context Reasoning Edge Detection.In the process of spatial description using spatial context matching inference algorithm, polar coordinate transformation is required at first. Taking the ist contour sampling point $p_i(x_i,y_i)$ as the reference coordinate origin for the logarithmic polar coordinate transformation, for any pixel point (x_j,y_j), the polar coordinate transformation formula is as follows:

$$\begin{cases} r_{ij} = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2} \\ \theta_{ij} = \arctan\left(\dfrac{y_i - y_j}{x_i - x_j}\right) \end{cases} \tag{15}$$

where r_{ij} is the radius in log-polar coordinates, θ_{ij} is the angle in log-polar coordinates, (x_i,y_j) corresponds to the coordinates in the Cartesian coordinate system of the ith contour sampling point p_i, and (x_j,y_j) corresponds to the coordinates in the Cartesian coordinate system of the remaining contour sampling point p_j [22].

The shape context of arbitrary pixel point p_i is described as the histogram distribution of the remaining contour sampling points with respect to the angle θ and logarithmic distance $\lg r$ of pixel point p_i. The arbitrary point p_i is used as the reference coordinate origin to establish a logarithmic polar coordinate system, which divides the whole space into $S \times T$ regions by dividing $\lg r$ and θ into S and T equal parts, respectively. Generally $S=5, T=12$ is taken, i.e., partitioned into 60 subregions. After completing the division of the spatial region, the number of distributions of the remaining $N-1$ sampled points except pixel point p_i in these 60 regions is recorded. This distribution number is the shape context of the points, which is calculated as follows:

$$h_i(k) = \#\{p \neq p_i : p \in bin(k)\} \tag{16}$$

where $k = 0,1,2,3,\ldots,60$. The well $\#$ operation denotes the number of points removed p_i and then taken p to be contained by the k th region. The spatial description of the graph is accomplished by computing the shape histogram for each point in the point set.

Since the characterization operator of shape context is mainly based on the histogram distribution, the statistical test of χ^2 distribution in statistics is generally used as the similarity matching cost of shape context, then the matching cost C_{ij} of two points is calculated as follows:

$$C_{ij} = C(p_i, q_i) = \frac{1}{2}\sum_{k-1}^{K} \frac{\left[h_i(k) - h_j(k)\right]^2}{h_i(k) + h_j(k)} \tag{17}$$

where $h_i(k), h_j(k)$ represents the value in the Kth bin in the shape histogram of point p_i on the first graph and point q_j on the second graph, respectively. The larger C_{ij} is, the more costly it is to match the two points, the less similarity between the two points, and the less likely they are to be mutually matched pairs of points.

3.2.2. Segmentation methods for scene safety regions.OTSU is a parameter-free unsupervised image segmentation algorithm for automatic threshold selection. First assume that the threshold value is T. According to the T value, the image pixels are divided into two parts greater than and less than T. The number is denoted as c_1, c_2 then the probability that a pixel is divided into two parts p_1, p_2 is:

$$p_1 = \frac{c_1}{c} \quad p_2 = \frac{c_2}{c} \tag{18}$$

where c is the total number of image cords and $p_1 + p_2 = 1$ Next, the pixel mean of each section is calculated, noting that m_1 is the pixel mean of the section larger than T and m_2 is the pixel mean of the section smaller than T. Then the inter-class variance is:

$$\sigma^2 = p_1 \times p_2 \times (m_1 - m_2)^2 \tag{19}$$

Then T which maximizes σ^2 is the threshold value.

The traditional one-dimensional OTSU image segmentation method has been widely used in image segmentation due to its universal applicability and simple algorithmic operations. However, under normal circumstances, when the target of the segmented image and the background of the large differences, one-dimensional OTSU image segmentation method can get satisfactory results, but if the histogram is multi-peak due to the influence of noise or illumination and other factors, it is difficult to get satisfactory results with one-dimensional OTSU image segmentation method. At the same time, the one-dimensional OTSU segmentation method only considers the characteristics of the pixel itself, without considering the pixel space around it, which makes the algorithm vulnerable to the influence of external factors, resulting in inaccurate segmentation and other problems. Therefore, in this paper, one-dimensional is extended to two-dimensional, so that the pixel point, while considering its own gray-scale features, combines with its surrounding pixel space to increase the noise resistance.

The improved OTSU algorithm takes T as the threshold value, where $x+y=K$ the pixel frequency P_k can be expressed as:

$$P_K = \frac{n_K}{M \times N}, K = 0,1,\cdots,2(L-1) \tag{20}$$

where n_t is the number of pixel points satisfying $x+y=T$ and $M \times N$ is the total number of pixel points. Then the maximum interclass variance criterion is given based on the probability distribution of T, i.e:

$$\sigma_1^2(T) = m_0(\mu_0 - \mu_T)^2 + m_1(\mu_1 - \mu_T)^2 \tag{21}$$

$$m_0 = \sum_{k=0}^{T-1} P_K \tag{22}$$

$$m_1 = 1 - m_0 \tag{23}$$

$$\mu_0 = \sum_{K=0}^{T-1}\left(K\frac{P_K}{m_0}\right) \tag{24}$$

$$\mu_1 = \sum_{K=T+1}^{2(L-1)}\left(K\frac{P_K}{m_1}\right) \tag{25}$$

$$\mu_T = m_0\mu_0 + m_1\mu_1 \tag{26}$$

Where m_0 is the probability of the part that is less than threshold T, m_1 is the probability of the part that is greater than threshold T, μ_0 is the mean gray value of the part that is less than T, μ_1 is the mean gray value of the part that is greater than T, and μ_T is the mean gray value of the whole image. Then the optimal threshold T_b is:

$$\sigma_1^2(T_b) = \max \sigma_1^2(T), 0 \le T \le 2(L-1) \tag{27}$$

4. Validation of the effectiveness of safety distance testing for substation operations

In the current context of rapid progress of information technology, the traditional operation and maintenance methods have been difficult to meet the dual requirements of security and high efficiency of modern power systems. Therefore, how to utilize emerging technologies to improve the security protection and monitoring management of substations has become an urgent problem in the industry. At the same time, substation O&M field operations face various security risks, the existence of which seriously threaten the stable operation of the power system and the personal safety of O&M personnel. Therefore, exploring efficient and intelligent safety protection and monitoring technology can not only improve the efficiency of substation operation and maintenance, but also effectively prevent and reduce the occurrence of safety accidents, and guarantee the reliable operation of the power system and the safety of operation and maintenance personnel.

4.1. Binocular Stereo Perception Matching Effectiveness

4.1.1. PSMNet model validation

1) Comparison of output weights of stacked hourglass modules

There exists a stacked hourglass module in the PSMNet model, whose Output1, Output2 and Output3 outputs are used for model training, which can facilitate the training process of neural networks, and each output has its own weights, which have a large impact on the final model training results. In order to select the optimal set of weights, this paper conducts experiments on the Sceneflow dataset for various combinations of weights between 0 and 1, with endpoint error (EPE) as the evaluation index. For the baseline, the weights of all three outputs are selected as 1 in this experiment, then the comparison results of the output weights of the stacked hourglass module under different combinations of weights are shown in Table 1.

In the process of stacking the neural network, each hourglass module will contain the feature information obtained from the convolution of the previous module, and this feature information will be maximized in the last stacking module, the closer to the end of the network the more important the loss calculation is for the training of the network, and each sub-block plays an auxiliary role in the training process of the network, so that the error can be effectively reversed. From the data in the table, it can be seen that when the weights of Loss1 and Loss2 are increasing in the process, the endpoint error will continue to decrease, and the information contained in Loss1 and Loss2 can effectively increase the accuracy of the resulting parallax map. However, increasing the weights one by one does not guarantee that the accuracy will continue to improve, and appropriate weights can make the error effectively slewed in the network, and the outputs of multiple coding sub-blocks ensure balanced training of the network. The results show that Loss1 weights set to 0.6, Loss2 weights set to 0.8, and Loss3 weights set to 1.0 minimize the error on the Sceneflow dataset to only 0.95%.

Table 1. Comparison of output weights of stack sand-drain module

No.	Weight			EPE
	Loss1	Loss2	Loss3	
1	0.1	0.1	1.0	1.56%
2	0.2	0.4	1.0	1.28%
3	0.4	0.6	1.0	1.04%
4	0.6	0.8	1.0	0.95%
5	0.8	1.0	1.0	1.07%

2) Comparison results on different datasets

In order to verify the effectiveness of the proposed algorithm, this experiment evaluates the performance of PSMNet-based binocular stereo matching network on the Sceneflow dataset against five algorithms with advanced and characterized performance in the ranking list, including CRL, DispNetC, GC-Net, PSMNet, and AANet.In addition, the percentage of mis-matched pixels is used as an evaluation index, the errors under different pixels were analyzed on the KITTI dataset. Table 2 shows the evaluation comparison results of binocular stereo perception matching network on different datasets, where px denotes pixel, and Noc and All represent non-occluded region pixel and all region pixel, respectively.

As can be seen from the table, the EPE result accuracy of this paper's model on the Sceneflow dataset is superior to the five selected models. It is reduced by 0.71px compared to CRL and 0.04px compared to AANet, and the error accuracy is reduced by 51.08% and 5.56% compared to the two models, respectively. It can be seen that the SPP module in the PSMNet network can effectively improve the feature extraction ability of the model, increase the useful information contained in the feature map, and improve the network accuracy. In addition, the model in this paper has the best error accuracy in the non-obscured region with greater than 4px and greater than 6px error and in all regions, which is better than several other models. These two errors are 1.45 and 0.74 in the non-obscured region and 1.73 and 0.93 in all regions, respectively.Adding 3D convolutional layers to the stacked hourglass module can effectively reduce the model's errors in the non-obscured region and in all regions with good results.

Table 2. The comparison results of different data sets

Model	Sceneflow	KITTI							
		>2px%		>4px%		>6px%		Mean (%)	
	EPE (px)	Noc	All	Noc	All	Noc	All	Noc	All
CRL	1.39	2.64	3.18	1.69	2.04	1.02	1.23	0.51	0.62

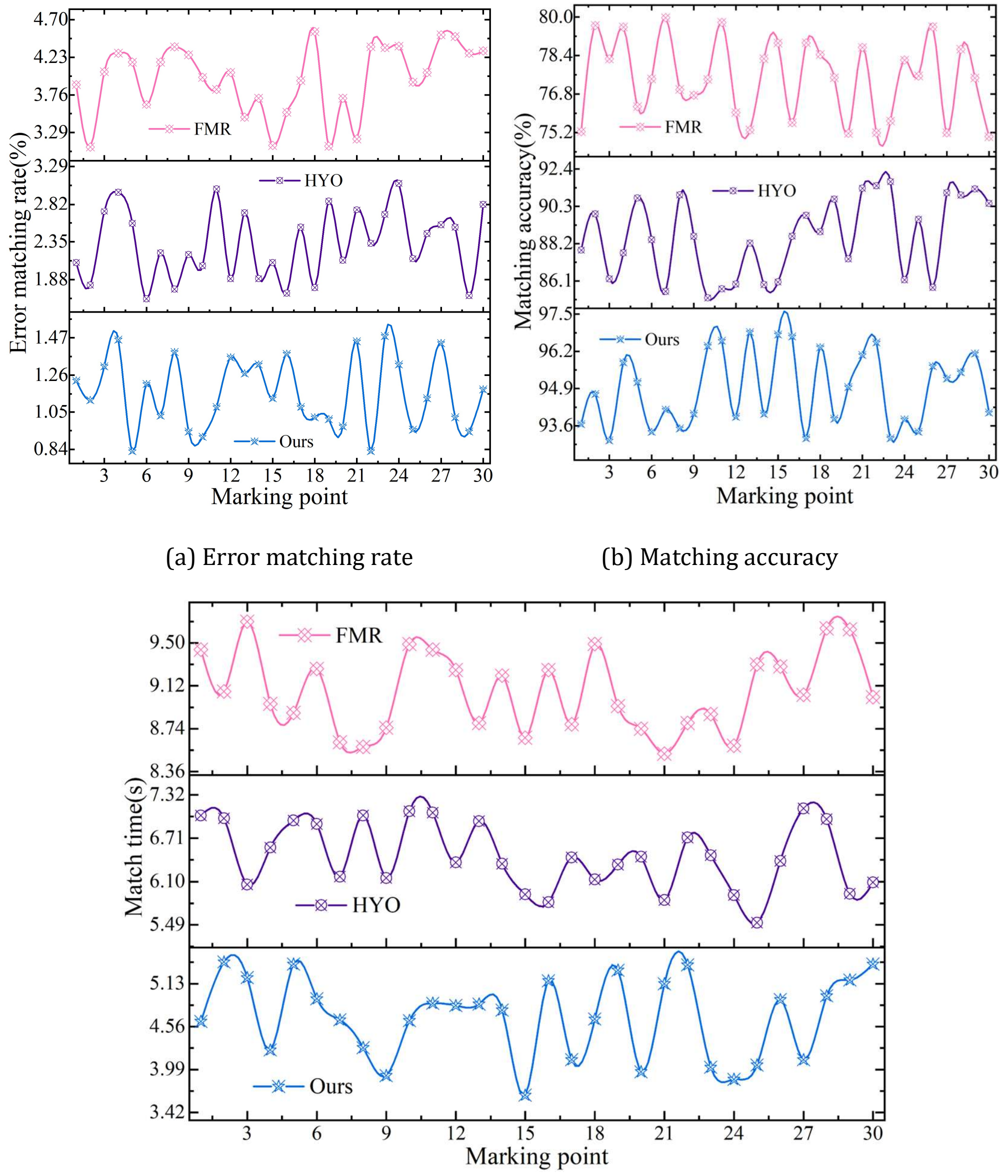
DispNetC	1.65	7.28	8.13	4.13	4.66	2.08	2.37	0.92	1.07
GC-Net	2.84	2.79	3.45	1.78	2.35	1.15	1.46	0.63	0.74
AANet	0.72	2.35	2.99	1.54	2.01	0.97	1.35	0.40	0.51
PSMNet	0.68	2.47	3.21	1.45	1.72	0.74	0.93	0.48	0.59

4.1.2. Binocular stereo perception matching.In order to further verify the image matching effect of the PSMNet binocular stereo perception matching method proposed in this paper in the process of substation operation, a video surveillance image of a substation is taken as an example, in which a certain frame of the image is called, and the feature points of the staff and the safety area in the image are calibrated. Corresponding to its image and feature marker points, the noise in the image is adjusted with the network loss value generated by the normalization processing model after normalizing the above image pixel channels. After the training processing, two traditional feature point matching methods are prepared for experiments with the designed matching method to compare the performance of the three matching methods, namely, this paper's method, Fused Multi-Feature Representation Binocular Stereo Matching (FMR), and Hyperpixel Optimized Binocular Stereo Matching (HYO). Based on the experimental preparation, 300 neighboring pixel blocks around the marked points in the image are selected, and ten groups of pixels are used as an experimental group, when the pixel blocks appear to be occluded, it is defined as an error matching process, and the error matching rate, matching accuracy, and running time produced by each marked pixel point are calculated and counted, and the final results are obtained as shown in Fig. 3. Among them, Fig. 3(a)~(c) shows the comparison results of the error matching rate, matching accuracy and running time, respectively.

From the results of the error matching rate, it can be seen that controlling the three matching methods to process images with the same pixel size, the overall error matching rate value of the binocular stereo matching method with fused multi-feature representation ranges from 3.11% to 4.49%, and there are more pixel points that are incorrectly matched. The binocular stereo matching method with superpixel optimization has an error matching value between 1.68% and 3.01%, with more actual mismatched pixel points. In contrast, the binocular stereo perception matching method designed based on PSMNet network in this paper produces a mismatch pixel rate between 0.83 and 1.48%, and compared with the traditional matching method, the designed matching method has fewer mismatched pixel points and the strongest actual substation operation process image matching effect.

Based on the accuracy results, it can be seen that the average matching accuracies of fusion multi-feature representation, super-pixel optimization, and the methods in this paper are 77.43%, 88.59%, and 94.89%, respectively. The pixel accuracy obtained from the actual matching of the first two algorithms is small, and the method of this paper obtains the largest pixel accuracy value and the best image pixel accuracy. After setting the image convolution parameters, according to the results of the running time of the three matching methods, it can be seen that the running time of the binocular stereo matching method that incorporates multi-feature representation as well as super-pixel optimization ranges from 5.52s to 9.03, and the actual time required for matching is relatively long, whereas the running time of this paper's model for substation operation image matching ranges from 3.65s to 5.42s, which is comparable to that of the two traditional Compared with the two traditional matching methods, the running time required by the designed and obtained matching method is the shortest.

In summary, the binocular stereo sensing model constructed by PSMNet network can realize the accurate matching of substation operation images, lay the foundation for the 3D reconstruction of substation operation images, and also provide support for the accurate detection of safety distances in the process of substation operation.



(a) Error matching rate (b) Matching accuracy

(c) Match time

Fig. 3. The matching effect of binocular sense

4.2. Verification of safe area testing for substation operations

4.2.1. Safe area edge detection.In order to verify the effectiveness of the spatial context inference algorithm proposed in this paper in edge detection, it is evaluated in comparison with the traditional Canny edge algorithm, Sobel algorithm and the improved Canny algorithm. Four different scene images (scenes S1~S4) of the substation operation site containing pretzel noise are selected as experimental objects based on the metrics A, B and C, where A represents the total number of edge points, B represents the number of points whose pixels have their connected components in their 4-neighborhood, and C represents the number of points whose pixels have their connected components in their 8-neighborhood. Here, an increase in the value of A indicates that the higher the number of edge points, the better the extraction, while smaller values of C/A and C/B indicate better linear connectivity. Table 3 shows the statistical results of edge detection for different algorithms.

From the table, it can be seen that the algorithm of this paper is lower than the traditional Canny algorithm, Sobel algorithm and Improved Canny algorithm in terms of C/A and C/B values. In terms of

C/A values than Canny algorithm, Sobel algorithm and Improved Canny algorithm, it is lower by 11.43% to 72.73%, 20.83% to 40.91% and 8.82% to 23.53%, respectively. In terms of C/B values, it is lower than Canny algorithm, Sobel algorithm and Improved Canny algorithm by 12.90%~59.55%, 46.55%~69.18%, 6.67%~38.16% respectively. In addition, from subjective observation, it can be seen that the Sobel algorithm can roughly outline the image but is basically covered by the pretzel noise, and the Canny algorithm is slightly improved compared to the Sobel algorithm, but it cannot remove the interference of the pretzel noise, and there are discontinuities at the local edges, and the edge connection is not complete enough. The improved Canny algorithm is effective in filtering pretzel noise, and the overall edge extraction effect is also better, but there are interruptions at the local edges, resulting in the overall integrity of the image contour is poor. Compared with the above algorithms, the edge detection algorithm based on spatial contextual inference in this paper is generally better than the previous three algorithms, and there is little admixture of pretzel noise in the edge detection results, the image contour is more complete, the edge details are richer, the edge connection is more continuous, and the appearance of pseudo edges is smaller. Therefore, in this paper, edge detection based on spatial contextual inference for different scenes at the substation operation site can better retain the detailed features of the image edges and shows excellent results in maintaining the edge integrity.

Table 3. The edge of different algorithms is tested for statistical results

Scene	Index	Canny	Sobel	Improve Canny	Ours
S1	C/A	0.025	0.022	0.017	0.013
	C/B	0.178	0.173	0.083	0.072
S2	C/A	0.017	0.018	0.022	0.011
	C/B	0.093	0.254	0.076	0.105
S3	C/A	0.035	0.041	0.034	0.031
	C/B	0.108	0.305	0.015	0.014
S4	C/A	0.011	0.024	0.022	0.019
	C/B	0.072	0.116	0.075	0.062

4.2.2. Threshold Segmentation Execution Efficiency.In this paper, after using spatial contextual reasoning to detect the edges of the scene at the substation operation site, the improved OTSU threshold segmentation method is utilized to segment the scene at the substation operation site, and then better obtain the spatial coordinates of its safety region. In order to verify the operational efficiency of the improved OTSU threshold segmentation method, the improved MRF (Algorithm 1) and the image segmentation method incorporating multi-featured marking information (Algorithm 2) are selected as a comparison, and the images of a substation substation operation scene collected in the previous paper are used as an example, and 200 images are randomly selected from them, to record the efficiency of the three algorithms for image segmentation, and the efficiency of image segmentation of different algorithms is shown in Fig. 4.

On the whole, the improved OTSU threshold segmentation method proposed in this paper consumes between 3.44ms and 9.76ms when performing image segmentation of substation operation scenes, and the average value of its image segmentation time consumed is 6.34ms, which is 59.18% and 72.65% lower than the average value of time consumed by the image segmentation of the improved MRF and the image segmentation of fused multi-feature marking information, respectively. This indicates that the method in this paper consumes less time when performing image segmentation of substation operation scenes and has higher operational segmentation efficiency. In this paper, the spatial context inference method is used for edge detection of substation operation scene images, and by setting reasonable upper and lower thresholds and connecting the edges according to the height of the pixel grayscale value, the efficient segmentation of substation operation scene images is realized, which effectively improves the segmentation accuracy and efficiency of substation operation scene images.

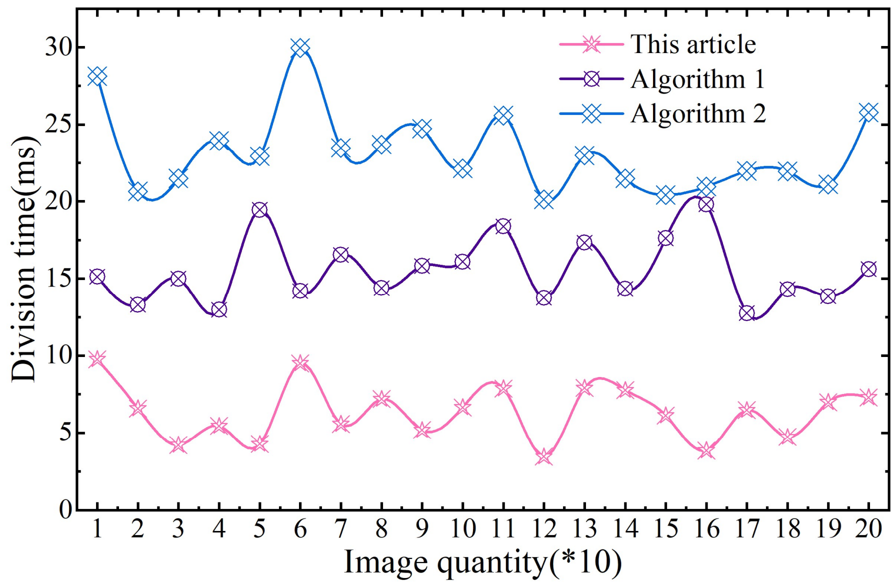


Fig. 4. Image segmentation efficiency of different algorithms

4.3. Transformer Safety Distance Detection Model Performance

4.3.1. Distance detection model performance.In this paper, the binocular stereo perception matching model is used to obtain the three-dimensional coordinates of the staff in the substation operation scene, and then combined with the spatial contextual reasoning of the edge detection of the safe region to obtain the three-dimensional coordinates of the safe region, and combined with the two Euclidean distance to construct the safe distance detection of the substation operation. In order to verify the feasibility and accuracy of the method proposed in this paper, the distance measurement experiment is carried out in this paper with the operator as the object. In 300mm ~ 30m every 300mm to take points for binocular ranging, and high-precision rangefinder measurement data as the real distance, the relative error to measure the method of ranging performance, the experimental results are shown in Figure 5.

As can be seen from the figure, within 0~2m, the error is larger because the target is located in the blind area of the field of view where the binocular field of view meets, and its relative error is 5.00% at most. Between 2~12m, the ranging error of this method increases with the distance. Within 12~21m, the error curve of this method is relatively smooth, and it can maintain high accuracy and high anti-interference ability, and the average relative error is kept near about 1.61%, and this method can effectively control small and medium-sized substation substation operation site. At the distance of 21~30m, the difficulty of stereo matching increases rapidly due to the fact that the scene elements account for too few pixels in the image, at which time the error of this method increases significantly, and the maximum relative error is about 2.78%. Overall, combining the binocular stereo perception matching model with the spatial context inference algorithm to detect the safety distance of the substation operation has high feasibility, and its detection relative error is small, which can satisfy the detection of the safety area in the process of substation operation.

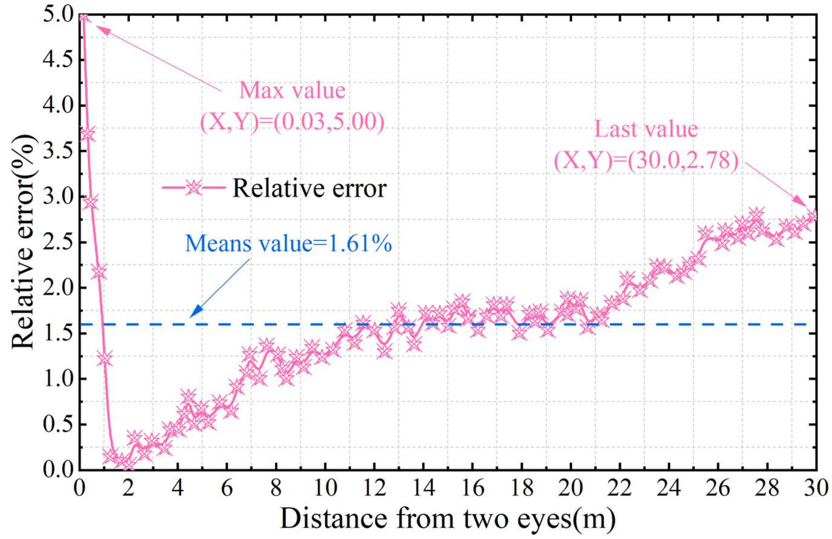


Fig. 5. Safety distance detection performance test

4.3.2. Detection performance of different methods.In this paper, the method is tested in a substation, and the method is used to control the operation process of single person and multiple persons at the operation site from the horizontal viewpoint and inclined viewpoint. At the same time, in order to verify the feasibility and accuracy of the method proposed in this paper, this paper adopts high-precision rangefinder measurement data as the source of the real distance between the operator and the energized equipment, and measures the accuracy of this method by absolute and relative errors, and this paper shows 12 groups of experimental results here. Figure 6 shows the quantitative demonstration results of the performance of this paper's method.

In the test environment of this paper, the detection speed of the method proposed in this paper is 14.52 frames/s in the real scenario, which can meet the demand of industrial-grade real-time. The average relative error of the method proposed in this paper is 0.85%, and the minimum and maximum values of the absolute error are 0.02m and 0.13m, respectively, and the minimum and maximum values of its relative error are 0.21% and 2.39%, respectively, which has the performance of high-precision detection of the safety distance of the operator substation.

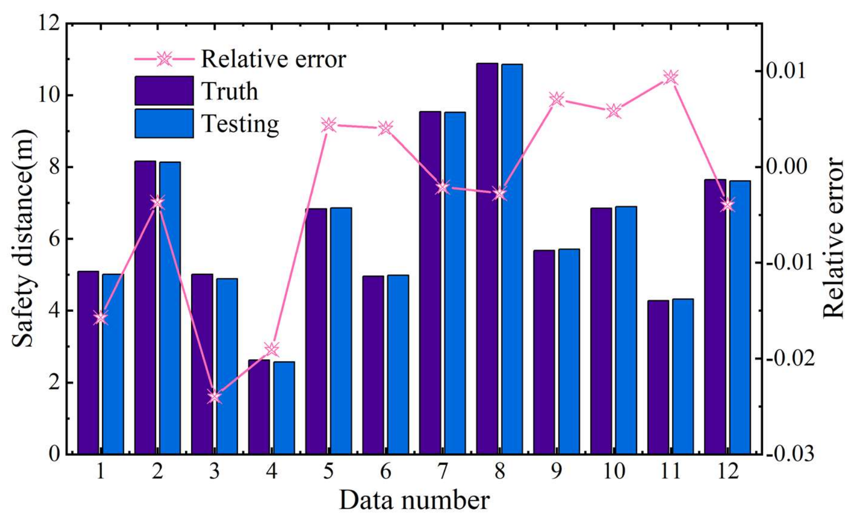


Fig. 6. Performance comparison of different object detection models

Meanwhile, this paper compares the proposed method with the three methods of electric field monitoring (FM), stereo matching (SM), and laser point cloud diagnosis (LPC), and this paper analyzes the performance of the four methods from three aspects, namely, ranging error, processing speed, and usage scenarios, respectively, and the results of the performance comparison of different safety distance detection methods are shown in Table 4.

As can be seen from the table, the average relative error of the adopted safety distance detection method based on electric field detection is 16.38%, and although it has real-time detection speed, its detection accuracy is relatively low due to the susceptibility of on-site electric field to interference, and it can not be adapted to complex scenes. The average relative error of the distance detection method based on stereo matching technology is 7.93%, with low detection accuracy and detection speed, while the method has poor robustness to long-distance targets. Highly accurate safety distance detection is achieved by laser point cloud data with an average relative error of 2.46%, but the method has high cost and insufficient real-time performance due to the need to collect point cloud data, as well as low generalization due to the limitation of less semantic information in the point cloud data. The method proposed in this paper has higher detection accuracy and faster processing speed, and shows stronger robustness and generalization ability when facing different substation substation operation scenarios.

Table 4. Performance comparison results of different detection methods

Method	Mean absolute error	Mean relative error	Processing speed	Whether to support multiple scenarios
FM	1.165m	16.38%	Real time	Unsupported
SM	0.012m	7.93%	Nonreal-time	Unsupported
LPC	0.083m	2.46%	Nonreal-time	Unsupported
Ours	0.044m	0.85%	Real time	Support

The error sources of the method proposed in this paper have several aspects, including the error of training data acquisition, the error generated by the calibration process of internal and external parameters of the camera, and the error generated by the selection of the coordinates of the center of the three-dimensional surfaces of the operator and the energized equipment, and other factors. Overall, the proposed method of substation safety distance detection for substation personnel based on binocular stereo matching and safe area edge detection is a general and convenient method for personnel substation safety distance detection. By collecting real substation operation data to train the model, the method in this paper can accurately measure the safety distance between on-site operators and energized equipment in real time and report to the police.

5. Conclusion

Electric power production should be completed through the work of the substation, so in the electric power production for the substation operation and maintenance site operations carried out by certain safety protection and monitoring technology is indispensable to implement countermeasures, which can give substation operation and maintenance site to bring a more reliable safety guarantee. The article proposes a safe distance detection frame for substation operation that combines binocular stereo matching perception and spatial context reasoning, and verifies the feasibility of the method in practical application through quantitative analysis. The binocular stereo sensing matching method based on PSMNet network design produces a mismatch pixel rate between 0.83 and 1.48%, and the overall operation time is between 3.65s and 5.42s. The minimum C/A value for safe area detection for substation operation through spatial contextual reasoning is only 0.011, and the average relative error for safe distance detection for substation operation is only 0.85%. The combination of deep learning technology and spatial contextual reasoning algorithm can realize the safe detection of substation operation and fully guarantee that the staff in the substation carries out the substation-related work in the safe area of near-electric operation.

References

[1] Khalikov, S., Boqijonov, U., Muminov, H., Ikromov, N., & Yuldashev, E. (2023, January). Operational management of substation safety power supply. In AIP Conference Proceedings (Vol. 2552, No. 1). AIP Publishing.

[2] Li, B., Zhang, W., Zhao, H., Wang, C., Cao, H., Jin, J., ... & Wang, T. (2018, June). Research on Prevention and Management of Substation Operation. In 2018 6th International Conference on Machinery, Materials and Computing Technology (ICMMCT 2018) (pp. 175-177). Atlantis Press.

[3] Wen, B., Moustafa, M. A., & Junwu, D. (2019). Seismic Fragilities of High-Voltage Substation Disconnect Switches. Earthquake Spectra, 35(4), 1559-1582.

[4] Gurevich, V. (2017). Protection of Substation Critical Equipment Against Intentional Electromagnetic Threats. John Wiley & Sons.

[5] FÎȚĂ, N. D., OBRETENOVA, M. I., LAZAR, T., MURESAN-GRECU, F., SAFTA, G. E., & SCHIOPU, A. M. (2023). RISK ASSESSMENTS IN TERMS OF OCCUPATIONAL HEALTH AND SAFETY FOR 220/110/20 kV SARDANESTI POWER SUBSTATION. Fiability & Durability/Fiabilitate si Durabilitate, 32(2).

[6] Vilaplana, J. A. L., Kabbara, N., Coste, T., Morais, H., Zerriffi, H., & Gibescu, M. (2024). Virtualized Protection, Automation, and Control in Electrical Substations: An Open-Source Dynamic Cost-Benefit Assessment Model. IEEE Access, 12, 107488-107504.

[7] Oputa, O., Ezeonye, C. S., Obi, P. I., & Onwuka, I. K. (2023). Mitigating Faults Effects on Equipment and Personnel on Substations. NIPES-Journal of Science and Technology Research, 5(1), 101-111.

[8] Wu, Y., Fang, X., Men, F., Qin, X., & Zhou, L. (2020, May). Research on simulation training system for substation equipment maintenance. In Journal of Physics: Conference Series (Vol. 1550, No. 5, p. 052006). IOP Publishing.

[9] Rui, H., Zhaoli, G., Mingkai, X., Xiaoping, L., & Chunguang, Y. (2018, August). Research on Fault Emergency Strategy for Substation DC System. In 2018 International Conference on Advanced Mechatronic Systems (ICAMechS) (pp. 27-30). IEEE.

[10] Behera, C., Banik, A., Nandi, J., Reddy, G. H., Chakrapani, P., & Goswami, A. K. (2020). A probabilistic approach for assessment of financial loss due to equipment outage caused by voltage sag using cost matrix. International Transactions on Electrical Energy Systems, 30(3), e12202.

[11] Thomas, D., & Fung, J. (2022). Measuring downstream supply chain losses due to power disturbances. Energy economics, 114.

[12] Ivanova, M., Dimitrova, R., & Filipov, A. (2020, September). Analysis of power outages and human errors in the operation of equipment in power grids. In 2020 12th Electrical Engineering Faculty Conference (BulEF) (pp. 1-5). IEEE.

[13] Hao, J., Wang, J., Liu, Z., Yu, X., Zhang, H., & Lan, C. (2024, August). Measures to Shorten the Outage Time of Substation Comprehensive Maintenance. In 2024 IEEE 2nd International Conference on Sensors, Electronics and Computer Engineering (ICSECE) (pp. 387-390). IEEE.

[14] Idoniboyeobu, D. C., Wokoma, B. C., & Ibanibo, V. C. (2018). Preventive maintenance for substation with aging equipment using Weibull distribution. American Journal of Engineering Research, 7, 95-101.

[15] Liang, H., Blagojević, N., Xie, Q., & Stojadinović, B. (2023). Seismic resilience assessment and improvement framework for electrical substations. Earthquake Engineering & Structural Dynamics, 52(4), 1040-1058.

[16] Chattopadhyay, A., Ukil, A., Jap, D., & Bhasin, S. (2017). Toward threat of implementation attacks on substation security: Case study on fault detection and isolation. IEEE Transactions on Industrial

Informatics, 14(6), 2442-2451.

[17] Chen, B., Xu, Z., Kirmani, S., Ichter, B., Sadigh, D., Guibas, L., & Xia, F. (2024). Spatialvlm: Endowing vision-language models with spatial reasoning capabilities. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 14455-14465).

[18] Resnick, I., Hegarty, M., Uttal, D., & Lowrie, T. (2023). Spatial Reasoning: The Role of Context in Learning. In Proceedings of the Annual Meeting of the Cognitive Science Society (Vol. 45, No. 45).

[19] Zhangping You,Dajian Yi,Zheng Fang & Wenhui Zhang. (2024). Image Enhancement ANPSO Processing Technology Based on Improved Particle Swarm Optimization Algorithm. IAENG International Journal of Computer Science(11).

[20] Guo Jianbo,Wang Guoqiang,Guan Wei,Chen Zeren & Liu Zhengbin. (2023). A feasible region detection method for vehicles in unstructured environments based on PSMNet and improved RANSAC. Multimedia Tools and Applications(28),43967-43989.

[21] Jiazhen Dou,Jiongshen Fang,Wenjun Jiang,Jianglei Di & Yuwen Qin. (2025). High-precision 3D reconstruction of terahertz computed tomography under extremely sparse view. Optics and Lasers in Engineering108783-108783.

[22] Lee Jae Woong & Helal Abdelsalam. (2022). Modeling and Reasoning of Contexts in Smart Spaces Based on Stochastic Analysis of Sensor Data. Applied Sciences(5),2452-2452.