

An Early Warning Method of Pantograph Horn Drilling Based on Superpixel HOG Algorithm and YOLO v3 smart Detector

Enhong Wang, Shubin Zheng, Qianwen Zhong, Liming Li and Qiaomu Zhang

EasyChair preprints are intended for rapid dissemination of research results and are integrated with the rest of EasyChair.

September 15, 2019

An Early Warning Method of Pantograph Horn Drilling Based on Superpixel HOG Algorithm and YOLO v3 smart Detector

Enhong Wang¹, Shubin Zheng², Qianwen Zhong³, Liming Li⁴, Qiaomu Zhang⁵

School of Urban Rail Transportation, Shanghai University of Engineering Science Songjiang, Shanghai 201620,China; ¹email: aristotle_weh@foxmail.com;²email: shubin.zheng@sues.edu.cn;³email:qianwen.zhong@sues.edu.cn;⁴email:liming0028 @126.com; ⁵email: joeblack929@163.com

ABSTRACT

To slove the contact net drilling problem caused by horn intrusion during pantograph operation, a warning method for the target of pantograph arch is proposed. Firstly, the method performs the target feature area on the received pantograph images by clustering algorithm to realize the target image feature extraction and the pantograph outline feature extraction, and through the clustered contour feature to obtain the location of the largest contact point area of the pantograph and catenary. Then, the target feature area is segmented from the background by superpixel HOG target segmentation algorithm, and the target data set is thus formed by the labeled maximum feature images. Finally, The YOLOv3-smart detector is adopted to build classification model. The results showed that the proposed method could accurately track and extract the contact area of the pantograph and the catenary from the video, and had an effective significance for early warning of the pantograph drilling problem.

KEYWORDS

Pantograph;Feature extraction;Early warning method;Superpixel HOG;YOLOv3-smart;

1 INTRODUCTION

In recent years, a large number of experts and scholars have conducted extensive research on the arcing point detection of the pantograph, the integrity of the pantograph, and the real-time monitoring of the pantograph. (M. Swift et al 2011;). proposed a new computer vision method to control the pantograph, and use image processing to detect the height of the pantograph and test it. The height of the output is controlled and verified by experiments. (Z. Xiao-Heng et al 2010;)in order to monitor the state of the pantograph-contact net system, mainly for tracking the average displacement of the contact line, tracking the contact points and detecting the abnormal points, and positioning the arc, the method can achieve The purpose of monitoring by the pantograph. (S. Rusu-Anghel et al 2011;) proposed a pantograph method. The method is based on a single multiple detector (SSD) method that

collects an in-orbit image creation data set having multiple viewing angles and multiple pantograph types, and the defects of the pantograph horn can be detected by training. (A. Landi et al 2006;) proposed the use of particle swarm tracking to analyze the contact points between the pantograph and the catenary, track the position of the contact line along the frame, analyze the tracking signal, and detect the state of the pantograph, and pass the experiment. Verification of the results, there are good results. Raluca Rob et al. [5] proposed a data acquisition system that can track and record the zigzag motion of the railway pantograph during operation. The hardware system is used to realize communication and real-time processing. In summary, most of the above studies are on the pantograph, and the problem of the horn angle of the pantograph is rarely analyzed. Even if the integrity of the horn is considered, the problem of the line of the pantograph cannot be predicted. Early warning of pantograph horn invasion. In summary, most of the above studies are on the pantograph, and the problem of the horn angle of the pantograph is rarely analyzed. Even if the integrity of the horn is considered, the problem of the line of the pantograph cannot be predicted. Early warning of pantograph horn horn invasion

2 Maximum feature area extraction of contact net image

In order to obtain the characteristics of the hand pantograph image, the experimental video data used in this experiment are all from the video data of the laboratory simulation of the state of the pantograph. The following picture shows the image acquired by the pantograph in the actual train operation and the state of the pantograph in the laboratory. (I. Aydin et al 2012;)The simulated contrast effect diagram of the simulation is carried out, and the key parts of the pantograph monitoring are marked in the figure. From the comparison diagram, the motion state of the pantograph motion state can be found. The pantograph was subjected to monitoring experiments through experimental data from the experimental bench.



Actual pantograph system

pantograph dynamic simulation system

Figure 1. Comparison of actual pantograph running image and pantograph motion state simulation test bench

2.1 Pantograph k-means clustering feature extraction

The k-means clustering algorithm classifies the data according to the degree of membership of the data, and classifies the pixels with high similarity ,The similarity of the undivided is low. The k-means clustering algorithm is described in detail:

The collection of pixels of an image $M = \{m_1, m_2, m_3, \dots, m_n\}$ is aggregated into k classes, producing a class result $V = \{V_i\}(i = 1, 2, \dots, k)$ that is centered at the cluster. According to the iteration making the objective function converges, and finally the clustering and image segmentation are completed:

$$J_{\rm m} = \sum_{i=1}^{k} \sum_{j=1}^{n} u_{ij}^{y} ||\mathbf{m}_{\rm j} - v_{i}||^{2}$$
(1)

The two-dimensional membership matrix U are defined in $k \times n$ the element in U expression indicates that the j data point m_j belongs to the membership degree of the cluster, and the table weighted index and, ||.|| representing the Euclidean distance between two pixels, must satisfy the following constraints:

$$\sum_{i=1}^{k} u_{ij} = 1, u_{ij} \in [0,1], 0 \le \sum_{j=1}^{n} u_{ij} \le n$$
(2)

Using the Lagrangian multiplier method, it can be found that when the objective function is minimized, the solution of the membership cluster center is completed by iterating the following formula to obtain the minimum value of the objective function:

$$u_{ij} = \frac{1}{\sum_{r=1}^{k} \left(\frac{\|\mathbf{p}_{j} - \mathbf{v}_{i}\|}{\|\mathbf{p}_{j} - \mathbf{v}_{r}\|}\right)^{\frac{2}{(m-1)}}}$$

$$v_{i} = \frac{\sum_{j=1}^{n} u_{ij}^{m} \mathbf{p}_{j}}{\sum_{j=1}^{n} u_{ij}^{m}}$$
(3)
(3)

The k-means algorithm separates the original image from the pantograph foreground and the background image, highlighting the obvious difference between the fastener and the background, and has a significant effect. (I. Aydin et al 2013;)



Figure 2. Pantograph image clustering feature renderings

2.2 Feature feature localization of maximum feature area of pantograph image

To ensure the integrity of the target contour extraction, using a closed loop search of the shortest path scheme for virtual connection of FIG significantly in nonclosed contour based on the use of the method to obtain the shortest path.(I. Aydin et al 2012;)Suppose there are M_e non-closed endpoints in the contour image, The path length between any two non-closed e_p , e_q , $L(e_p, e_q)$ has a positive correlation with the Euclidean distance between the two points Negative correlation with connection e_p . Normalize two points.

$$L(\mathbf{e}_{p}, \ \mathbf{e}_{q}) = \frac{||\mathbf{e}_{p}, \ \mathbf{e}_{q}||_{2}}{\xi_{i}}$$
(5)

By repeatedly iteratively connecting the end points of the short paths, a plurality of closed annular contour lines can be formed, which can effectively ensure the integrity of the contour target, and FIG. 7 is an effect diagram of the fastener contour extraction.(S. Barmada,2013)



Figure 3. Pantograph contour feature positioning effect diagram

3 Superpixel HOG image background segmentation

In order to obtain the largest pantograph target area from the pantograph image to reduce the interference of other pantograph areas, the pantograph adopts superpixel HOG background segmentation, and uses the clustering contour feature in the pantograph image to locate According to the pantograph feature, the super-pixel Hog complex background segmentation method is used to separate the largest feature target region in the pantograph image from the pantograph image to obtain the maximum feature data set.(S.-H. Chiu)

3.1 Pantograph image entropy rate superpixel feature

Superpixels aggregate pixel points with similar features in an image into pixel blocks of different sizes according to a certain rule, so that these pixel blocks express a visual meaning. The super-pixel segmentation based on the entropy rate achieves the effect of reducing the complexity and increasing the calculation speed by processing the aggregated pixel blocks.(I. Aydin et al 2012;)

The objective function of the entropy rate is constructed, set the discrete random variable, the entropy rate H(M) measures the uncertainty of the random variable, so you can define the probability mass function of H(M) to φ as the support set of M:

$$H(x) = -\sum_{\mathbf{x} \in \varphi} P_X(\mathbf{x}) \log P_X(\mathbf{x})$$
(6)

The superpixel entropy rate is to establish a topology through the superpixel of the image. The essence is to find a topology of the graph, define it as A, and connect all the superpixel blocks to maximize the target feature function:

$$\max_{A} \{ H(A) + \lambda B(A) \}$$
(7)

Where: H(A) is the entropy rate term used to form the relative regular superpixel, λ is the parameter, refers to the balance factor, and B(A) is the balance term, the purpose is to constrain the size of the super pixel. The H(A) construction is:

$$H(A) = -\sum_{i} \mu_{i} \sum_{j} P_{i,j}(A) \log(P_{i,j}(A))$$
(8)

$$P_{i, j}(A) = \begin{cases} \frac{W_{i, j}}{W_{i}} & \text{if } i \neq j \quad and \, e_{i, j} \in A \\ 0 & \text{if } i \neq j \quad and \, e_{i, j} \in A \\ 1 - \frac{\sum j_{e_{i}, j \in Aw_{i, j}}}{W_{i}} & \text{if } i = j \end{cases}$$

$$(9)$$

The B(A) construction formula is:

$$B(A) = -\sum_{i} P_{Z_{A}}(i) \log(P_{Z_{A}}(i)) - N_{A}$$
(10)

Suppose the original image is G, N_A is the number of connected subgraphs in G, Z_A is the subgraph distribution, and $P_{Z_A}(i)$ is the distribution law:

$$P_{Z_A}(i) = \frac{|S_i|}{|v|}, \quad i = \{1, \dots, N_A\}$$
(11)

In order to obtain the optimal objective function, the greedy algorithm is used to process, and the obtained segmentation result σ is the balance factor parameter. The larger the balance factor σ is, the more irregular the segmentation superpixel shape is, but the shape is more uniform, and the σ is smaller than the super pixel shape. The more regular, but the size is not uniform. Assume that the initial value is ξ , *K* is the number of super pixels, and β is the dynamic factor, then the balance factor $\sigma = K\beta\xi$, where:

$$\beta = \frac{\max e_{i, j} H(e_{i, j}) - H(\phi)}{\max e_{i, j} B(e_{i, j}) - B(\phi)}$$
(12)



Figure 4. Entropy rate super pixel feature rendering

3.2 Superpixel HOG image background segmentation

The HOG is for the local area of the image, and describes the local area shape and texture features of the image by the local gradient or the target edge direction density distribution.(F. Takens et al 2012;) First, the image is divided into small connected regions, also called cell units, and then the pixel gradient or edge direction histogram in each cell unit is calculated. Finally, all cell unit gradient direction histograms are serially described to describe the HOG features of the image, ie, The extraction of HOG features is as follows:

a. Image normalization

Convert images into 2D images, normalize 2D images using Gamma correction to reduce imbalances in the image and reduce image noise interference:

$$I(\mathbf{x}, \mathbf{y}) = I(\mathbf{x}, \mathbf{y})^{Gamma}$$
(13)

b. Calculate the image gradient

Calculate the horizontal and vertical gradients of the image separately:

$$G_{y}(x, y) = I(x, y+1) - I(x, y-1)$$
(14)

$$G_{x}(x, y) = I(x+1, y) - I(x-1, y)$$
(15)

$$G(x, y) = \sqrt{G_x(x, y)^2 + G_y(x, y)^2}$$
(16)

$$\theta = \arctan(\frac{G_y(x, y)}{G_x(x, y)})$$
(17)

c. Establish cell units

Create a histogram for each cell unit. Divide the image into 64*32 cell units. 16*16 steps are 8. The input image will get:

$$(\frac{64-16}{8}+1)*(\frac{32-16}{8}+1) \tag{18}$$

The gradient direction of the image ranges from 0 to 360. The histogram in each cell unit is calculated, and all the histograms are connected in series to form a four-segment unit gradient histogram to form a one-dimensional array of HOG.



Figure 5. Obtain the maximum target HOG feature area

3.3 Maximum Entropy Rate and Minimum ROI Feature Location of HOG

By comparing the superpixel entropy rate characteristics of the pantograph frame in complex background and comparing the entropy rate maximum feature of the pantograph target region, the most significant region of the entropy rate of the pantograph target image can be extracted. The image containing the target feature is subjected to entropy rate superpixel segmentation to obtain the image feature, and combined with its HOG feature, the minimum range including the two features in the target region can be obtained, and the minimum ROI range of the largest feature is located from the image(P. Sermanet et al 2013;). To remove complex backgrounds.



Figure 6. Superpixel Hog feature image segmentation

3.4 Maximum feature ROI data set

The super-pixel complex background segmentation can obtain the minimum ROI of the largest feature of the target region. Removing the complex background in the original video image in the ROI can reduce the interference of other non-target objects on the monitoring formation and can preserve the maximum characteristics of the monitoring target.(R. Girshick 2015;)The segmented consecutive frames are saved as video data of the smallest ROI feature with the largest feature after removing the complex background, and Canny boundary extraction is performed on the ROI, and a maximum entropy ROICanny video file of the boundary effect is generated.



Figure 7. Hydropower eco-efficiency index (HEI) with descendant trend

The ROICanny boundary processing was used as the test training, and the test results are shown in Fig. 7. The original video data is extracted from the maximum feature ROI video and the Canny-processed ROICanny video data.

4 YOLOv3-smart recognition

4.1 YOLO v3-tiny structure

Like YOLO v3, YOLO v3-tiny still uses a multi-scale detection mechanism. The slight difference is that tiny uses only two scales: 13*13 and 26*26.

$$\begin{aligned} \lambda_{\text{coord}} \sum_{i=0}^{S^{2}} \sum_{j=0}^{B} I_{ij^{obj}} [(x_{i} - x_{i})^{2} + (y_{i} - y)] + \\ \lambda_{\text{coord}} \sum_{i=0}^{S^{2}} \sum_{j=0}^{B} I_{ij^{obj}} [(\sqrt{\omega}_{i} - \sqrt{\omega_{i}})^{2} + (\sqrt{h_{i}} - \sqrt{h_{i}})^{2}] + \\ \sum_{i=0}^{S^{2}} \sum_{j=0}^{B} I_{ij^{obj}} [(C_{i} - C_{i})^{2} + \\ \lambda_{noobj} \sum_{i=0}^{S^{2}} \sum_{j=0}^{B} I_{ij}^{noobj} (C_{i} + C_{i})^{2} + \\ \sum_{i=0}^{S^{2}} I_{i}^{obj} \sum_{c \in classes} (p_{i}(c) - p_{i}(c))^{2} \end{aligned}$$
(19)

At the same time, the network structure of YOLO v3-tiny is similar to other network structures, mainly composed of convolution layer and pooling layer. The whole network structure is simple and tidy. The main feature of the tiny structure is that the overall network parameters are small, the amount of calculation is small, the computing power of the computer is low, and the inference calculation of the input information can be completed quickly. The shortcomings are also obvious. In order to improve the speed, the detection accuracy is sacrificed. Due to the simple layer structure, the feature extraction is not sufficient. In view of the situation of YOLO v3-tiny itself, based on the actual needs of this design, the detection accuracy is improved without significantly reducing the detection speed. Its loss function is shown in equation.

4.2 YOLOv3-tiny-smart structure

Based on YOLO v3-tiny, the mechanism has been optimized to make it more suitable for pantograph video recognition.First, expand the network structure. For the deep learning network, the features extracted by the shallow network include the rigid structures and texture features of the object. At the same time, convolution operations using large feature maps can also extract richer and more detailed features. In the original YOLO v3-tiny network, the second layer downsamples the feature map to reduce the amount of computation on the network. In response to this situation, a 1*1 convolutional layer and a 3*3 convolutional layer two-layer network are added between the first layer and the second layer of the network to enhance the extraction of features of the shallow layer. There are two main functions of the 1*1 convolutional layer. One is the smooth transition of feature extraction, which can make the smoother transition of the features extracted by the first layer into the added 3*3 convolution layer, so that Features are not lost during this transfer. The second is to reduce the parameter amount of the network. If only a layer of 3*3 network is added to extract the shallow features, because the feature map just input is 416*416 size, the feature map has too many parameters of the network, which will cause this. The calculation of one layer is too much, which reduces the overall detection speed. Therefore, the 1*1 convolutional layer is added in front of the 3*3

convolutional layer to reduce the overall network parameters in the feature extraction operation. Then, add more convolutional layers for feature extraction before the 26*26 scale detection. In the original YOLO v3-tiny network, the 26*26 scale detection is only routed by the 3*3 convolutional layer of the 19th layer and the 3*3 convolutional layer of the 8th layer, and then sent to the detection. Although the shallow feature and the deep feature can be merged by route merging, since the deep feature is extracted by only a single layer of 3*3 convolution layer, the extracted abstract features are less. Here we refer to the layer structure setting of YOLO v3 multi-scale detection. On the basis of the original 3*3 network of the deep features.



Figure 8. YOLOv3-tiny-smart system flow chart

The modified network model is called YOLO v3-tiny-smart. Although its detection speed is lower than that of YOLO v3-tiny, the network detection accuracy is improved by enhancing the network's ability to extract features. The modified model was compared horizontally with YOLO v3-tiny, YOLO v3, SSD and Fast R-CNN. The results of the comparison are shown in Table 1:

Network model	Map
YOLO v3-tiny-strong	98%
YOLO v3-tiny	54.5%
YOLO v3	86.5%
SSD	81.7%
Fast R-CNN	90.3%

 Table 1. Network Model Map Comparison Table

From the horizontal comparison of the five networks, the Fast R-CNN has better detection accuracy than the one stage algorithm such as YOLO v3 and SSD due to its inherent two stage network structure, especially its candidate window recommendation mechanism. Enhance the ability to detect targets. The YOLO v3 network model adopts the new network structure of darknet-53, deepens the feature extraction ability of the network layer as a whole, and adopts a multi-scale detection mechanism to improve the detection accuracy of small targets. At the same time, YOLO v3's downsampling operation uses a convolutional layer with a step size of 2, which prevents the loss of feature details during the downsampling process. These optimizations have greatly improved the overall detection accuracy, and Fast R - The CNN network is close. (E. Karakose et al 2015;)By comparison, the detection accuracy of YOLO v3-tiny-smart is 15% higher than that of YOLO v3-tiny, which has been greatly improved.The detection speed of the five algorithms was tested with the NVIDIA GTX980 GPU. The test comparison results are shown in Table 2:

Network model	FPS
YOLO v3-tiny-strong	47
YOLO v3-tiny	40
YOLO v3	26
SSD	19
Fast R-CNN	11

Table 2. Network Model FPS Comparison Table

It can be seen from the above comparison that although the detection speed of the modified YOLO v3-tiny-smart model has decreased, the detection accuracy has been greatly improved, and the condition value of real-time detection is satisfied. In other network models, the Fast R-CNN network structure is more complex and the detection speed is the slowest. YOLO v3 has undergone two generations of updates on the basis of YOLO v1. The network structure is more concise than before, and the detection speed is better than It is similar to the SSD model that is also an end-to-end network.

Through the comparison of MAP and FPS above, it can be proved that the overall performance of the modified network model is better than YOLO v3-tiny. The modified network model is more powerful for target feature extraction and more suitable for the application scenario of this design.

5 Experimental results and analysis

Using the YOLOv3-tiny-smart system, the maximum feature ROI data set is trained, and the loss function curve is obtained by iterating 4215 times, as shown in FIG.It can be seen from Fig. 10 that the loss function of YOLOv3-tiny-strong is better than YOLOv3-tiny, so the effect of YOLOv3-tiny-strong training is better than YOLOv3-tiny.



training loss function

Through the above training, weights can be obtained, real-time monitoring using weights, and a stable positional relationship between the pantograph and the

contact net can be recognized from the video image, and the coordinates of the contact point in the image can be output.



Figure 10. Pantograph sheep angle drill bow warning effect chart

The dynamic contact position is accurately monitored while the train is running, and the target dynamic monitoring effect is shown in Figures 11 and 12. In the figure, the green mark is the midpoint position of the pantograph, the red color at both ends represents the safe area, and the framed part is the contact position. Figure 12 is a video detail rendering with the output curve labeled with dynamic contact points, the green line is the midline of the pantograph movement, and the red areas on both sides are safe areas.



Figure 11. Hydropower eco-efficiency index (HEI) with descendant trend

Figure 13 shows the dynamic parameters of the pantograph. The green line is the midline of the pantograph and the red part is the safe area where the pantograph is in contact with the catenary. In the figure, the pantograph indicates power in the safe zone. The bow runs safely, and the part of the figure that exceeds the red line is beyond the safe position.

6 CONCLUSION

This paper proposes a warning method for the pantograph arching target of the pantograph and arch net contact target to prevent the pantograph of the pantograph from being drilled during high-speed operation. Firstly, the target image clustering feature extraction and pantograph contour feature extraction are used to locate the target region in the image by the obtained cluster contour feature, and then locate the target region; then, the superpixel HOG background segmentation algorithm is applied to the pantograph background. Segmentation establishes the largest feature dataset; finally, the pantograph dataset is trained with YOLOv3-tiny-smart. The results show that the target tracking method can accurately track the pantograph and the contact net from the video data, obtain the target of the bow mesh contact point, and warn the problem of the pantograph angle of the pantograph.

ACKNOWLEDGEMENT

Shanghai Local University Capacity Building Project (Grant No. 18030501300) Shanghai University of Engineering and Science Research Startup Project Fund (Grant No. 0240-E3-0507-19-05081)

Dynamic Diagnosis Method and Theoretical Study of Vibration and Visual Fusion Characterize Track Diseases supported by NSFC (Grant No. 51975347) **REFERENCES**

- A. Landi, L. Menconi, L. Sani (2006). "Hough Transform and Thermo-Visionfor Monitoring Pantograph – Catenary System", J.Journal of Rail and Rapid Transit, 435-447.
- E. Karakose, M. T. Gencoglu, M. Karakose, I. Aydin, and E. Akin (2016). "A New Experimental Approach Using Image Processing Based Tracking for an Efficient Fault Diagnosis in Pantograph-Catenary Systems,"*IEEE Transactions on Industrial Informatics*, 1-1.
- F. Takens (2012). "Detecting Strange Attractors in Turbulence" J.Proc.Dynamical Systems and Turbulence, 366-381.
- I. Aydin, M. Karakose, and E. Akin(2012). "A New Contactless Fault Diagnosis Approach for Pantograph-Catenary System," *IEEE International Conference On Mechatronika*, 1-6.
- I. Aydin, M. Karakose, E. Akin (2013). "A Robust Anomaly Detection in Pantograph-Catenary System Based on Mean-Shift Tracking and Foreground Detection, "*IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, 4444-4449.

- I. Aydin, M. Karakose, E. Akin (2012). "A new contactless fault diagnosis approach for pantograph-catenary system", *IEEE 15th International Symposium on MECHATRONIKA*, Czech Republic, 1-6.
- M. Swift, G. Aurisicchio, P. Pace(2011). "New practices for railway condition monitoring and predictive analysis," J.Proc. the IET Conf. on Railway Condition Monitoring and Non-Destructive Testing, RCM, 1 - 6.
- P. Sermanet, D. Eigen, X. Zhang, M. Mathieu, R. Fergus, and Y. Lecun (2013)
 "OverFeat: Integrated Recognition, Localization and Detection using Convolutional Networks," Eprint Arxiv.
- R. Girshick (2015) "Fast R-CNN," *IEEE International Conference on Computer Vision*, 1440-1448.
- S. B. Changalasetty, A.S.Badawi, W.Ghribi (2014) "Identification and feature extraction of moving vehicles in LabVIEW", *J.International Conference on Communication and Signal Processing*, 19-20.
- S. Barmada, M. Raugi, M. Tucci1, F. Romano (2013). "Arc detection in pantograph-catenary systems by the use of support vector machines-based classification," *J.IET Electrical Systems in Transportation*,1-8.
- S.-H. Chiu, C.-Y. Wen, J.-H. Lee, K.-H. Lin, and H.-M. Chen (2012) "A fast randomized generalized hough transform for arbitrary shape detection", *J.International Journal of Innovative Computing, Information and Control*,1103-1116.
- S. Rusu-Anghel, C. Miklos, M. Topor, D. Demian, S. Mezinescu (2011) "Pantograph catenary system control using elements of chaos theory," *J.Pantograph Catenary Interaction Framework for Intelligent Control (PACIFIC)*,1-4.
- Z. Xiao-Heng, G. Xiao-rong, W. Ze-yong, W. Li, Y.Kai (2010) "Study on the edge detection and extraction algorithm in the pantographslipper's abrasion", *J.International Conference on Computational and Information Sciences*,474 477.