A Visual Measurement Algorithm of Approaching Vehicle Speed Based on Deep Learning

Zhu Yurong
School of Electrical and Electronic Engineering,
Shanghai Institute of Technology, Shanghai,201418, China

Zhao Huailin✉
School of Electrical and Electronic Engineering,
Shanghai Institute of Technology, Shanghai, 201418, China

Liu Junjie
School of Electrical and Electronic Engineering,
Shanghai Institute of Technology, Shanghai, 201418, China

Zhang Jinping
School of Electrical and Electronic Engineering,
Shanghai Institute of Technology, Shanghai, 201418, China

Ji Xiaojun
Shanghai Gazecloud Software Technology Co.Ltd.
Shanghai Institute of Technology, Shanghai, 201418, China
E-mail: zhao_huailin@yahoo.com
www.sit.edu.cn

Abstract

With the urbanizational process expediting and the national economy developing rapidly and healthily, the amount of private cars is on the rise, and traffic accidents occur frequently due to speeding and other reasons, and the difficulty of traffic supervision has also increased. This topic will use semantic segmentation and feature extraction and matching. Based on the video data of the traffic surveillance camera, an algorithm is designed to quickly calculate the matching of feature points in adjacent frames with low computing power to achieve the calculation. The same vehicle moves within the two frames of the target, so as to calculate the speed of the vehicle. Firstly, performing semantic segmentation based on deep learning, we choose a fully convolutional network to achieve semantic segmentation of depth maps, and distinguish the picture's principal part. After that, we can realize features extraction and mapping. The HOG algorithm is used on the matching step, the target's relative movement is calculated based on these matched point pairs to measure the moving speed of the vehicle. The experiment and the test prove that the system can realize the efficient speed measurement of moving vehicles.

Keywords: Vehicle speed measurement; semantic segmentation; HOG feature extraction; SVM classification

1. Introduction

There are more and more cars on the road, and the traffic has become heavier and heavier, which has led to the frequent traffic accidents and the difficult traffic supervision. Some data show that on the road sections equipped with "electronic eyes", the number of people and vehicles observing traffic regulations has increased significantly, indicating that the use of some technological methods will enhance people's cautiousness and have a certain binding force on people's irregular behaviors[1]. Strengthen people's attention to traffic rules, thereby reducing the incidence of accidents and reducing conflicts between personnel. The "electronic eye" for the speed monitoring system of mobile vehicles is one of them, which reduces the incidence of traffic accidents caused by speeding, indicating that the installation of these monitoring facilities can avoid the occurrence of some major traffic accidents and effectively reduce the difficulty of traffic supervision[2,3]. In order to improve the level of public security and reduce the incidence of traffic incidents, it is necessary to research and improve vehicle detection and monitoring technologies.

The world is strengthening the development of emerging technologies, science and technology are changing with each passing day, society is becoming more and more informatized, network information technology, communication transmission technology and multimedia technology and other high-tech are constantly advancing with the times, penetrating all aspects of life, and intelligent transportation systems are gradually gaining popularity and become favored by people. As the number of vehicles on the road increases, the monitoring of their speed becomes more and more important[4].

The core of this subject is to realize the speed measurement of moving vehicles. For a few frames of image speed measurement, the image needs to be processed first. If the entire image is processed directly, the amount of calculation is large and complicated, so semantic segmentation is needed to divide the main body of the image. As a key technology, semantic segmentation is used in computer vision and other fields. The realization is that the pixels of the image are processed and classified to obtain several different semantic categories, so that the image is divided into many sub-regions, and the semantics of these sub-regions have different specific meanings. Since the advent of Convolutional Neural Network (CNN), this technology has been continuously improving. Researchers have tried to use it to solve various problems[5,6]. Through unremitting efforts, it has proved its potential in the processing of semantic segmentation problems, whether in automatic driving, medical image segmentation, pedestrian and vehicle detection all have good application prospects.

This paper is based on the traffic monitoring video and uses low computing power to calculate the speed of moving vehicles. It helps to enhance the supervision of vehicle speed, better determine the responsibility of the accident, and can also serve as a warning to the driver. It can reduce the incidence of accidents. It has good application prospects in traffic monitoring, autonomous driving and other fields.

2. System scheme

2.1. *Overall design*

For the video data obtained by the traffic camera, deframe processing is first performed to obtain a frame-by-frame image sequence. Then perform semantic segmentation based on deep learning on the image sequence to obtain the main body of the image, that is, the vehicle. Then for the image sequence containing the vehicle, the vehicle detection based on HOG and SVM is performed[7]. Finally, some algorithms are used to measure the speed of moving vehicles. The overall design is shown in Figure 1.

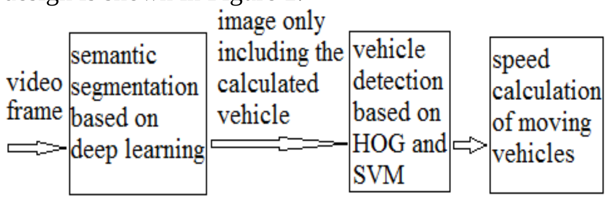

Fig. 1. The overall design

2.2. *Semantic segmentation based on deep learning*

Semantic segmentation, as a key research field of computer vision, has good applications for flat data, three-dimensional or high-dimensional data. In daily life, more and more scenes need to be understood, such as unmanned autonomous driving, reality augmentation, medical image processing, etc. We need to consider relevant semantics from each scene to effectively solve problems, and semantic segmentation is very important. Conducive to the understanding of the scene, so we need to continue to develop semantic segmentation technology.

The network model of semantic segmentation is usually composed of an encoder and a decoder. The encoder, which is a trained network that can be used in classification. The decoder determines the different structures of semantic segmentation. Its work is mainly to process the features. The low-resolution obtained in the encoder is based on semantic projection to obtain high-resolution features, so that it can be realized classification.

Semantic segmentation based on CNN is generally divided into three stages[8,9], namely pixel-level classification, down-sampling, and up-sampling. Figure 2 shows a common semantic segmentation architecture. For the convolutional network, down sampling achieves a feature map that reduces its resolution and increases the number of channels. It is mainly used to extract low-level semantic features in the image, as well as some abstract features. The characteristics of up-sampling are completely different. The size of the feature map will increase and the number of channels will decrease. After up-sampling is followed by down-sampling, the semantic content can be recovered slowly. In the classification stage, multi-classification can be achieved, by classifying all pixels, mainly based on semantic categories.

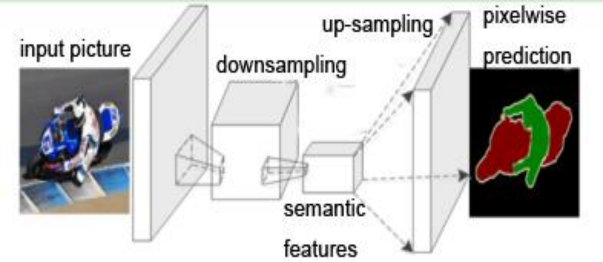

Fig. 2. Semantic segmentation based on CNN

2.3. *HOG feature extraction*

HOG is called Histogram of oriented gradient. It can use gradient information to describe part of the limited features of the image, and then use a series of algorithms to obtain the gradient feature information of the entire image, which is mainly used in computer vision field.

When performing HOG feature extraction on an image, the local information of the image can be depicted more completely through gradient information or edge information. To achieve HOG feature extraction need four steps: image preprocessing, calculation of the gradient vector of the image, extracting the histogram of the gradient direction, block normalization. The details are as follows.

2.3.1. *Image preprocessing*

The first step is to process the input image to achieve grayscale. Because the effect of color information in the HOG algorithm is not obvious, the impact on detection accuracy is also small, and the amount of information contained in the grayscale image will be less, and the calculation pressure will also be reduced. So usually the color image is grayed out first.

The second step is to perform Gamma normalization processing on the grayscale image. After the implementation of the first step, there are still lighting, shadows, and noise that affect HOG feature extraction. In order to reduce the impact of these factors, Gamma normalization (normalization) of the image obtained in the first step is used to solve the problem. The Gamma compression formula is shown in the following formula.

$$\mathrm{I}(\mathrm{x},\mathrm{y}) = \mathrm{I}(\mathrm{x},\mathrm{y})^{\mathrm{gamma}} \tag{1}$$

Where, $I(x, y)$ represents the value of (x, y). And gamma index refers to gamma correction.

2.3.2. The gradient vector of the image

The difference in gradient value can reflect the change of discontinuous features, such as gray value. Based on the pixel level, if the grayscale difference is larger for the image, then the gradient difference will be larger accordingly. For a certain object to be tested, the gradient change of the edge feature is generally the largest, and we usually use this feature to determine the edge position. Generally speaking, the first-order derivative of the function is used to express the image gradient, and the second-order derivative is used to express its change.

The usual method is to first use convolution in the original image to calculate the gradient components of the plan image, and then divide it horizontally and vertically. For the pixel (x, y) in a certain image, the pixel can be calculated by the following formula.

$$\mathrm{G_x}(\mathrm{x},\mathrm{y}) = \mathrm{H}(\mathrm{x}+1,\mathrm{y}) - \mathrm{H}(\mathrm{x}-1,\mathrm{y}) \tag{2}$$

$$\mathrm{G_y}(\mathrm{x},\mathrm{y}) = \mathrm{H}(\mathrm{x},\mathrm{y}+1) - \mathrm{H}(\mathrm{x},\mathrm{y}-1) \tag{3}$$

$$\mathrm{G}(\mathrm{x},\mathrm{y}) = \sqrt{\mathrm{G_x}(\mathrm{x},\mathrm{y})^2 + \mathrm{G_y}(\mathrm{x},\mathrm{y})^2} \tag{4}$$

$$\alpha(\mathrm{x},\mathrm{y}) = \tan^{-1}\left(\frac{\mathrm{G_y}(\mathrm{x},\mathrm{y})}{\mathrm{G_x}(\mathrm{x},\mathrm{y})}\right) \tag{5}$$

Among them, for a certain pixel in the input image, $\mathrm{G_x}(\mathrm{x},\mathrm{y})$ is its horizontal component, $\mathrm{G_y}(\mathrm{x},\mathrm{y})$ is vertical component, $\mathrm{H}(\mathrm{x},\mathrm{y})$ is pixel value, $\mathrm{G}(\mathrm{x},\mathrm{y})$ is gradient magnitude, and $\alpha(\mathrm{x},\mathrm{y})$ is gradient direction.

2.3.3. Extracting the histogram of the gradient direction

In HOG feature extraction, each picture is divided into a finite number of cells of the same size and independent of each other. Different shapes can be divided into rectangles and stars. The calculation methods used are different, and the feature dimensions are also different. Generally, the picture is first divided into rectangular unit blocks of the same size, and then the unit blocks are equally divided into 9 directions, and the direction of the unit block can be obtained by weighting according to the gradient magnitude of each direction.

2.3.4. Block (interval) normalization within the image block

In the third step, the image has been divided into many Cell blocks, and then the characteristics of each Cell block are analyzed. Then, according to the obtained feature attributes of the Cell blocks, the Cell blocks need to be combined, the Cell blocks are concatenated to obtain a Block block, and the features of each Cell block are concatenated to obtain the entire HOG feature vector of the corresponding Block. Block blocks are allowed to overlap, but Cell blocks must be independent of each other. From this it can also be known that in the entire feature descriptor, there may be situations where the same feature gets different description results. In order to obtain the HOG feature vector of the complete image, each Block block needs to be connected in series, and this descriptor can be used for the processing of classification problems.

Through the above process, it can be known that Block blocks will overlap. When this phenomenon occurs, the pixels within it will be calculated multiple times, and the obtained gradient values will be more different, so that the accuracy of classification will be effectively improved.

2.4. *SVM classification*

SVM is called Support Vector Machine (Support Vector Machine). SVM has a good application when dealing with classification and extraction problems. It is a good algorithm that can be used for machine learning. It was first studied in 1995. Support vector machine is mainly used to deal with the problem of feature classification, a supervised learning algorithm based on boundary classification. The purpose of this algorithm is to find a boundary in a plane or a boundary in a high-dimensional space to realize the classification of features. When dealing with simple binary classification problems, this dividing line is easy to find and meet the requirements, but when dealing with multi-feature classification or complex binary classification problems, it is relatively more complicated. For dividing lines or interfaces that need to be expressed by mathematical expressions, it is important to master the relevant theories of support vector machines.

SVM needs the support of mathematical tools to effectively solve super-high-dimensional classification problems. Practice has proved that SVM works best when the number of samples is lower than the spatial dimension. In addition, SVM only needs to store support vectors, so it can effectively save memory space. Usually, in the construction of the classifier, the method of using HOG and SVM together is used. Because at this time, the gradient histogram can be used as a very robust descriptor, which can accurately reflect the characteristics of a certain class, which will make the classification effect better.

For linearly separable data, the classification process of SVM is shown in Figure 3. In addition to the processing of linearly separable problems, it can also handle data linearly inseparable. The usual method is to use some non-linear transformations to map this difficult problem in low-dimensional space to high-dimensional space, so that It becomes a linearly separable data to be solved, as shown in Figure 4. In order to reduce the complexity, the kernel function will be quoted in high-dimensional operations.

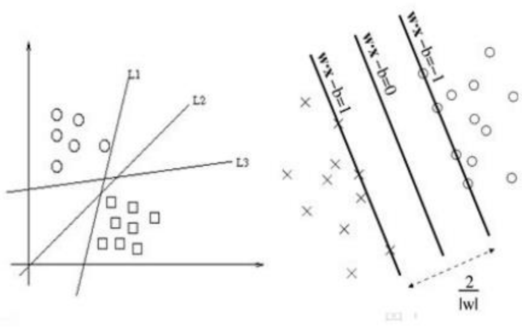

Fig. 3. Linear case

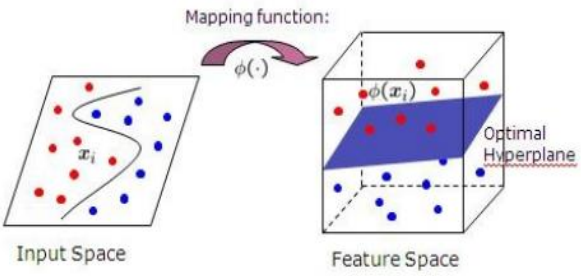

Fig. 4. Non-linear case

3. Programming

According to the analysis and implementation steps of the above-mentioned related technologies. First of all, for a traffic video data, we use an algorithm to extract the images of adjacent frames in the video data, and then perform semantic segmentation based on the full convolutional neural network to divide the main body and the background of the image. Next, input the vehicle (the main body of image) into the SVM classifier for training to further ensure the accuracy of the segmentation. After that, extract and match the HOG features of the main body in the image. When the gap in the number of video frames is within 10 frames, target tracking is performed to obtain a certain distance between the different frames. When moving distance of a vehicle is calculated, we could get its moving speed. Finally, the video where the speed of the moving vehicle in the original video has been measured is stored, and the export the video.

From the above methods, the program flow chart of this article can be made, based on the algorithm realization process, as shown in Figure 5.

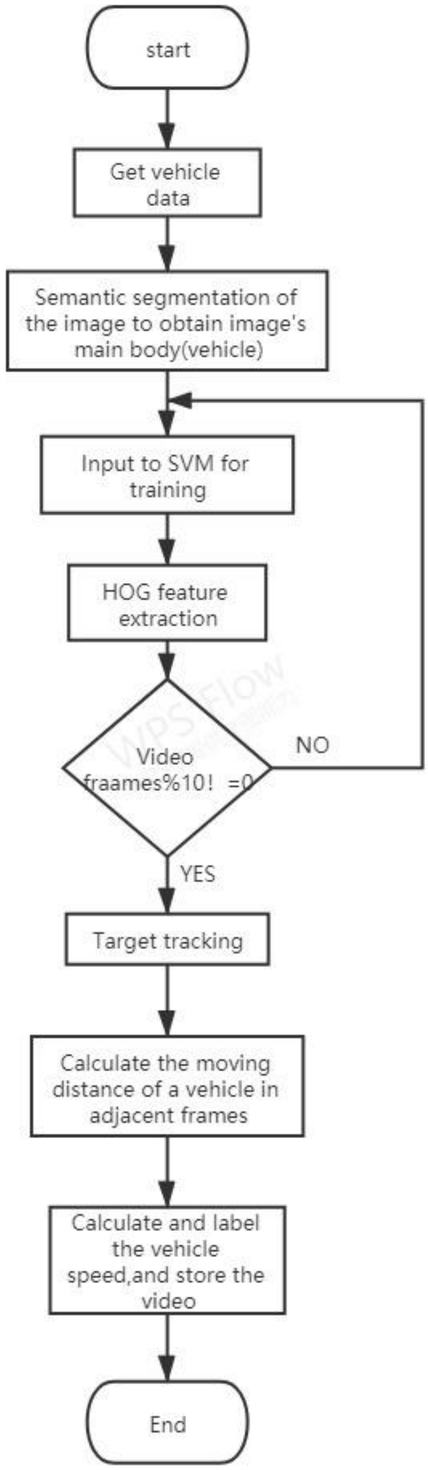

Fig. 5. Program flow chart

4. Experiment

In the previous sections, we mainly introduced and analyzed the related technologies and algorithms used in the speed measurement of moving vehicles, including CNN, HOG, and SVM classifiers and so on. In this section, it is mainly based on the following implementation process to obtain and analyze the corresponding results.

4.1. *Experimental Results*

First of all, for a traffic video data, we first use an algorithm to extract the images of adjacent frames in the video data, and then perform semantic segmentation based on the full convolutional neural network to achieve the segmentation of the vehicle body and the background. Then for the image body, also That is the vehicle. Then input into the SVM classifier for training, and then extract and match the HOG features of the main image. When the difference in the number of video frames is within 10 frames, perform target tracking to obtain a certain vehicle moving distance between adjacent frames of images, and then calculate Find out its moving speed, and finally store the video where the speed of the moving vehicle in the original video has been measured, and export the video.

After the original video data is run and debugged by the program, the video data that has been marked with the moving speed of the vehicle will be output. The screenshot of the video part is shown in Figure 6. Through the video data obtained from the data set, debugging and running in the program, the result shown in Figure 6 has been obtained, which can meet the requirements for the speed measurement of moving vehicles. Next, add some programs and add a yellow area to the output to better see how the program implements the vehicle speed measurement. Figure 7 shows three original video frames compared with the corresponding experiment results.

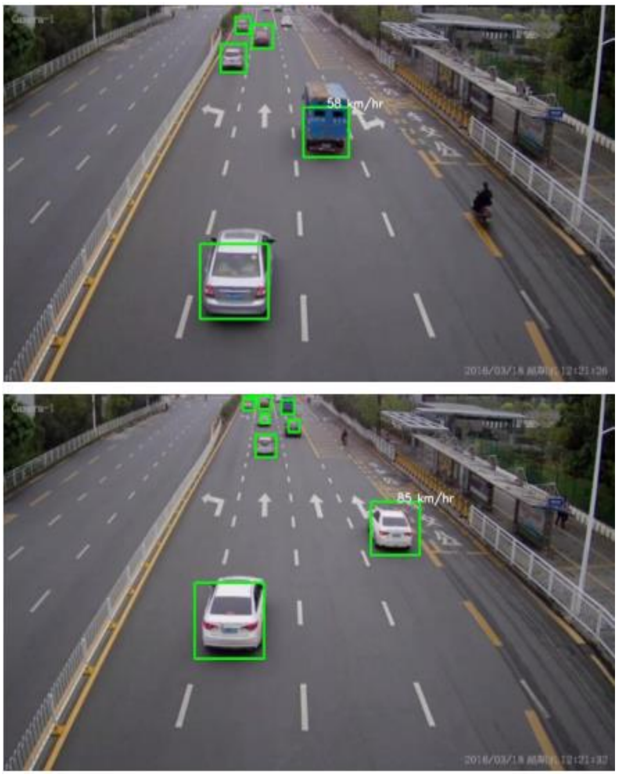

Fig. 6. Output video screenshot

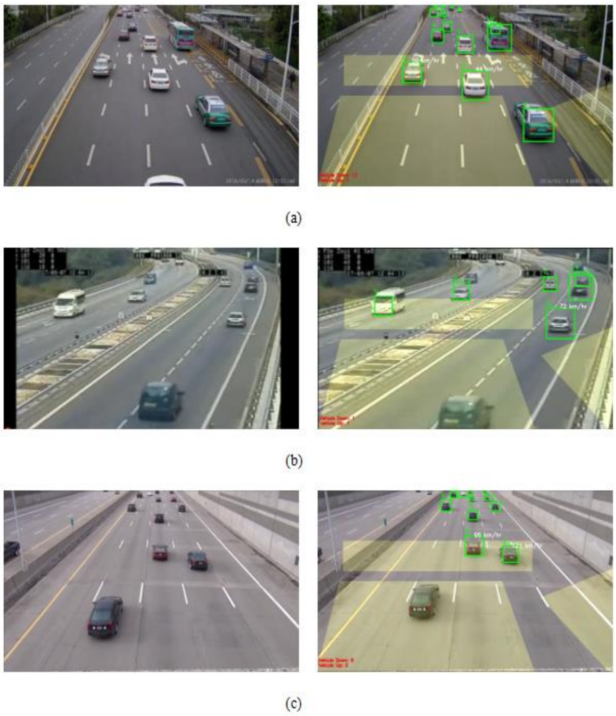

Fig. 7. Comparison of three groups of experiments

4.2. *Experiment Analysis*

Given the corresponding number of samples and the number of pictures to be tested, the number of samples is about 700, and the corresponding number of pictures to be tested is about 1800. The running results are shown in Table 1.

Table 1. Statistics of running results

No. of samples	No. Of pictures to be deteced	No. Of recognition	Accuracy rate
100	300	231	77.0%
100	400	324	81.0%
200	400	324	82.0%
300	700	657	93.9%

It can be seen from the Table1 that as the number of samples increases, the recognition rate is significantly improved, and the accuracy rate is also improved. Therefore, for the relevant data, the target of the vehicle can basically be identified. And the number to be tested is not as large as possible. It is necessary to select an appropriate value so that it can be better applied to engineering practice.

According to the results of the above-mentioned different traffic monitoring videos, it can be known that the program can detect more than 90% of the vehicles in the image, with a high detection rate, and the results will also show the moving speed of the vehicles. Next, the detection accuracy under light changes is further verified. In the three cases of medium, strong, and weak light intensity, 600 test samples are intercepted for accuracy measurement. The results are shown in Table 2.

Table 2. Detection degree under different illumination

Light intensity	Detection accuracy rate(%)	False detection rate(%)	Missing detection rate(%)
Normal light	97.22	2.32	0.46
Strong light	94.31	3.46	2.23
Weak light	91.87	1.77	6.36

It can be seen from Table 2 that the detection accuracy of HOG features can be maintained above 90% under the influence of different illumination, and its anti-interference ability is strong.

The speed measurement of the highway video with known speed is used to judge the accuracy of the moving vehicle speed calculation algorithm used in this article. Figure 8 shows the speed measurement result of a blue truck on a certain highway.

It is known that the actual speed of the blue truck in the picture is 80km/h. In Fig. 8(c), in the first 6 speed calculations, the error is kept within ±3km/h. However, it can be seen that in the 7th and 8th calculations, the data fluctuates greatly. Then it gradually calmed down. On the whole, from the line graph data, we can see that in multiple vehicle speed calculation experiments, the accuracy of the vehicle speed measured by the experiment has reached more than 90%, so the speed measurement method adopted in this paper can better meet the test requirements.

(a) Blue truck speed measurement start frame

(b) Blue truck speed test end frame

(c) Blue truck speed test line chart

Fig. 8. Blue truck speed measurement

5. Conclusion

By using the data in different data sets for testing, and then using the above method for result analysis and evaluation, it can be known that the program can basically detect all vehicles with a high detection rate, and it can also successfully display the moving speed of most vehicles. However, the correctness of the speed needs to be further verified, and the program will be further optimized and improved afterwards, so as to achieve a more accurate speed measurement of moving vehicles.

In sum, the proposed algorithm can effectively achieve the speed calculation of the moving vehicles. The image segmentation, HOG feature extraction, and the speed measurement are of high accuracy.

References

1. Tong, Zheng, Philippe Xu, and Thierry Denœux. "Evidential fully convolutional network for semantic segmentation." Applied Intelligence (2021): 1-24.
2. D. Y. Chan, D. C. Morris, T. J. Polascik, M. L. Palmeri and K. R. Nightingale, "Deep Convolutional Neural Networks for Displacement Estimation in ARFI Imaging," in IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control, vol. 68, no. 7, pp. 2472-2481.
3. Lahiani, Houssem, and Mahmoud Neji. "Hand gesture recognition method based on HOG-LBP features for mobile devices." Procedia Computer Science 126 (2018): 254-263.
4. Gan, Jiangzhang, Jiaye Li, and Yangcai Xie. "Robust SVM for Cost-Sensitive Learning." Neural Processing Letters (2021): 1-22.
5. Ye, Lu, Ting Duan, and Jiayi Zhu. "Neural network-based semantic segmentation model for robot perception of driverless vision." IET Cyber-systems and Robotics 2, no. 4 (2020): 190-196.
6. Wang, Qi, Xiang Gao, Fan Wang, Zhihang Ji, and Xiaopeng Hu. "Feature Point Matching Method Based on Consistent Edge Structures for Infrared and Visible Images." Applied Sciences 10, no. 7 (2020): 2302.
7. Pozzer, Sandra, Ehsan Rezazadeh Azar, Francisco Dalla Rosa, and Zacarias Martin Chamberlain Pravia. "Semantic segmentation of defects in infrared thermographic images of highly damaged concrete structures." Journal of Performance of Constructed Facilities 35, no. 1 (2021): 04020131.
8. Krishnakumar, K., and S. Indira Gandhi. "Video stitching based on multi-view spatiotemporal feature points and grid-based matching." The Visual Computer 36, no. 9 (2020): 1837-1846.
9. Wang, Qi, Xiang Gao, Fan Wang, Zhihang Ji, and Xiaopeng Hu. "Feature Point Matching Method Based on Consistent Edge Structures for Infrared and Visible Images." Applied Sciences 10, no. 7 (2020): 2302.

Authors Introduction

Ms. Yurong Zhu

She received the B.S. degree from Shanghai Institute of Technology, Shanghai, China. And now she is a master course student. Her main research interests are deep learning and Control Science and Engineering.

Dr. Huailin Zhao

He received his PhD from Oita University, Japan in 2008. He is a professor in the School of Electrical & Electronic Engineering, Shanghai Institute of Technology, China. His main research interests are robotics, multi-agent system and artificial intelligence. He is the members of both IEEE and Sigma Xi.

Mr. Junjie Liu

He received the B.S. degree from Shanghai Institute of Technology, Shanghai, China. His main research interests are deep learning and Control Science and Engineering.

Mr. Jinping Zhang

He received the B.S. degree from Shanghai Institute of Technology, Shanghai, China. And now he is a master course student. His main research interests are deep learning and Control Science and Engineering.

Mr. Xiaojun Ji

He received the master degree from Southeast University, Nanjing, China. He is a senior software engineer, system architect and product manager. And he is the founder of Shanghai Gazecloud Software Technology Co. Ltd..