

A Fast Detection and Recognition Algorithm for Pedestrian at Night Based on Entropy Weight Fast Support Vector Machine

Liang Rui¹, Wei Honglei², Zhu Qingxin^{1*}, Liao Shujiao¹, Deng Hongyao¹

1. School of Information and Software Engineering, University of Electronic Science and Technology of China, Chengdu Sichuan, 610054, China

2. Department of Sport, School of Economics and Management, Southwest Jiaotong University, Chengdu Sichuan, 611756, China.

190386989@qq.com

Abstract

In allusion to such problems as real-time requirement dissatisfaction and significant recognition difference caused by dimension difference existing in the imaging and recognition algorithm for pedestrian in dark scene, a fast head detection and recognition method for pedestrian at night based on fast support vector machine (FC-SVM) algorithm optimization and entropy weight is established in this paper according to relevant principle of statistics. Based on entropy weight, this method aims at improving the extraction process based on histogram gradient features in order to establish three-branch SVM for the deep recognition of pedestrian at night; meanwhile, FC-SVM algorithm is combined to optimize the recognition calculation overhead in order to ensure the real-time property of the recognition algorithm. Furthermore, the falsely detected pedestrians are evaluated on the basis of the head detection mode so as to improve pedestrian imaging matching accuracy. The simulation result shows that this method can not only effectively recognize FIR target of pedestrian at night, but also effectively adapt to such different application environments as urban and suburban areas on the basis of ensuring the real-time requirement for pedestrian recognition, thus presenting good practicability.

Keywords: Pedestrian detection; Head detection; Detection at night; Recognition; Support vector machine

1. Introduction

In such scene, the pedestrian infrared detection includes two processes: the first process is the acquisition of *ROIs* parameter; the second process is the recognition of *ROIs* parameter. In the second process, the parameter recognition accuracy can influence the overall accuracy and the calculation speed of the infrared detection [2], so the recognition algorithm of the second process is the key part of the pedestrian recognition and detection algorithm. *ROIs* detection algorithm for pedestrian at night mainly includes human body recognition mode, wherein this mode mainly includes motion recognition[3] and template matching[4]. Specifically, the moving pedestrian recognition method can be used to effectively recognize the local pedestrian texture change and has strong adaption to relevant scenes, but this method needs relatively accurate local pedestrian action data, so this method is limited for use and also has some disadvantages in the aspect of the real-time index for the recognition algorithm; meanwhile, the pedestrian detection method based on template matching has relatively high calculation efficiency and simple algorithm implementation process, but this method has too strict requirement

¹ Zhu Qingxin is the corresponding author.

for the accuracy of the pedestrian template data, so this method is inapplicable to such complex scenes as non-rigid pedestrian with obvious action changes. For the pedestrian detection method obtained according to the principle of statistics, we need to focus on two key links: the first one is the feature extraction process; the second one is the recognition classifier design. Such feature extraction methods as Haar recognition, LBP (Local Binary Pattern) recognition, Edgelet recognition, SIFT (Scale Invariant Feature Transform) recognition and HOG (Histogram of Oriented Gradient) recognition are usually used for the existing motion feature recognition process, wherein HOG recognition method with wide application scenes can not only effectively detect and describe pedestrian appearance and local feature, but also have other improved forms, for example: HOG scale variable gradient feature extraction method proposed in literature [5] can improve HOG gradient feature recognition speed, but the recognition accuracy is reduced more or less.

2. Three-Branch FC-SVM Pedestrian Detection

2.1. FC-SVM theoretical Analysis

The basic form for marking FC-SVM training samples is $X = \{(x_i, y_i)\}_{i=1}^L$, wherein $y_i = \{+1, -1\}, x_i \in R^d$. The expression form of N groups of support vectors obtained according to s-SVM recognition is $S = \{(s_i, y_i)\}_{i=1}^N$ ($S \subset X$ and $N < L$). In order to reserve P ($p < N$) groups of support vectors in the algorithm, the rest ($N - p$) groups of information data are stored in transformation vector W according to relevant formal transformation. HOG components are obtained from different parts of the pedestrian target and assigned with corresponding weight values. Afterwards, according to the pedestrian infrared imaging features and the imaging edge heterogeneity, the imaging gradient chaos is evaluated on the basis of the entropy value so as to determine the weight values of imaging HOG components. Specifically, k imaging blocks can be described according to the following formulae:

$$W_k = -\sum_{t=1}^{c \times r} P_t \log_2 P_t \quad (1)$$

$$P_m = \frac{h_m}{\sum_{t=1}^{c \times r} h_t}, m = 1, 2, \dots, c \times r \quad (2)$$

In formulae (1~2), r is the number of histogram projection intervals, c is the number of image blocks, and h_t is the histogram information strength along direction t in image block. Firstly, it is necessary to obtain the information entropy and normalize the local parameters for HOG gradient features in order to reduce the local gray level interference of the image. In HOG gradient feature expression, although trilinear interpolation has good effect, yet it cannot completely eliminate such condition as $P_n = 0$. In this paper, the limitation mode thereof is improved as follows:

$$P_n \log_2 P_n = 0 \quad (3)$$

$$M_q = W^T M_p \quad (4)$$

$$M_p = \begin{bmatrix} K(x_1, s_1) K(x_2, s_1) \dots K(x_L, s_1) \\ K(x_1, s_2) K(x_2, s_2) \dots K(x_L, s_2) \\ \vdots \\ K(x_1, s_p) K(x_2, s_p) \dots K(x_L, s_p) \end{bmatrix} \quad (5)$$

$$M_q = \begin{bmatrix} K(x_1, s_{p+1}) K(x_2, s_{p+1}) \dots K(x_L, s_{p+1}) \\ K(x_1, s_{p+2}) K(x_2, s_{p+2}) \dots K(x_L, s_{p+2}) \\ \vdots \\ K(x_1, s_N) K(x_2, s_N) \dots K(x_L, s_N) \end{bmatrix} \quad (6)$$

In Formulae (4~6), $K(\square)$ is the kernel function and meets Mercer limitation. In order to give consideration to the reservation of the transformation vector decision data and p groups of vectors for FC-SVM algorithm, the following decision form is obtained according to s-SVM recognition process:

$$f(x) = \sum_{i=1}^N \alpha_i y_i K(x, s_i) + b$$

$$= \left[\sum_{j=1}^p \alpha_j y_j K(x, s_j) + \sum_{j=p+1}^N \alpha_j y_j K(x, s_j) \right] + b \quad (7)$$

$$= (A_p^T + A_q^T W^T) M_x + b$$

$$A_p^T = (\alpha_1 y_1, \alpha_2 y_2, \dots, \alpha_p y_p) \quad (8)$$

$$A_q^T = (\alpha_{p+1} y_{p+1}, \alpha_{p+2} y_{p+2}, \dots, \alpha_N y_N) \quad (9)$$

$$M_x = [K(x, s_1), K(x, s_2), \dots, K(x, s_p)]^T \quad (10)$$

In Formula (7), the recognition calculation overhead of any sample vector x is mainly focused on sample dimensionality d and decision reservation of support vector number p . In this formula, the three parameters, namely A_p , A_q and W can be obtained in advance. During the recognition process, FC-SVM algorithm is superior to s-SVM algorithm in the aspect of calculation speed, and the calculation complexity thereof is reduced from $O(Nd)$ to $O(pd)$. Under general conditions, parameter p is significantly less than parameter N . Compared with s-SVM algorithm, FC-SVM algorithm has lower calculation complexity in the decision and storage processes. Meanwhile, since W storage space is significantly less than rest $(N-p)$ groups of vector space, FC-SVM algorithm has higher support vector dimensionality d .

2.2. Classifier Recognition Framework

Due to the pedestrian modality problems, the modalities of the pedestrians with the same distance from the photographing point are different each other, thus causing significant difference in the horizontal lengths of the minimum external rectangle of the pedestrian imaging. For example, for two persons with the same height, although the imaging widths thereof are significantly different, yet the heights H thereof are also different. Therefore, the original pedestrian image is divided into three non-intersected pedestrian modality data subsets as the training samples according to heights H , as shown in Figure 1. Meanwhile, according to the sample features, the pedestrian modality with the height more than 70 pixels is

regarded as the long-shot pixel, the pedestrian modality with the height less than 32 pixels is regarded as the close-shot pixel and the pedestrian with the height between 32 pixels and 70 pixels is regarded as the mid-shot pixel. Afterwards, the sample height values are taken as the training subsets to train the classifier so as to obtain the sample distribution rule and IE_HOG feature classifier under FC-SVM framework.

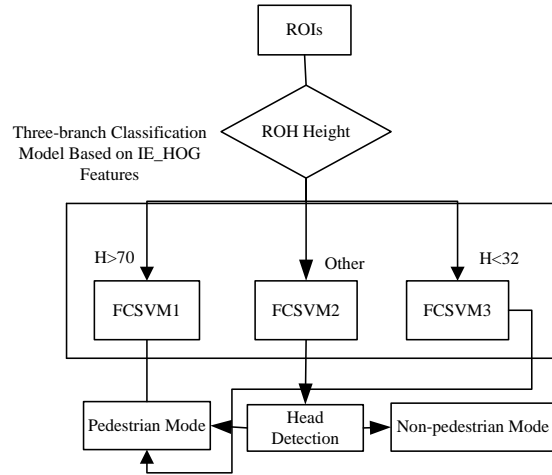


Figure 1. Principle of Pedestrian Detection in Infrared

3. Head Detection

The head image calibration mode for pedestrian imaging is adopted to improve the recognition reliability of the system and reduce the false positive rate. Generally speaking, due to the small dependence on the pedestrian posture, the head image in pedestrian imaging is favorable for obtaining stable pedestrian imaging recognition and detection effect; meanwhile, the head image in pedestrian imaging is usually located at the top imaging area, thus ensuring that the adoption of head modality for imaging calibration has many advantages and can improve system stability. Haar recognition method is designed on the basis of such advantages. Specifically, the gray level difference of the head image and the surrounding background is detected according to the specific rectangle modality template based on Haar features, as shown in Figure 2. Such information difference can be expressed as follows according to Formula (11):

$$F = S_b - S_w \quad (11)$$

In Formula (11), S_w and S_b respectively represent the gray average values (GAV) of white and black rectangles as shown in Figure 2 [9]. In the practical application of the scheme proposed thereby, the pedestrian images detected by infrared sensors are the images with gray level treatment, so S_w and S_b can be simply obtained as follows[10]:

$$GAV = \frac{1}{nm} \sum_{i=1}^m \sum_{j=1}^n g_{ij} \quad (12)$$

In Formula (12), m and n respectively refer to the number of the pixels in horizontal distribution and vertical distribution, and g_{ij} refers to the corresponding gray level value of pixel (i, j) .

Meanwhile, according to the judgment about whether the person is a pedestrian, the ideal gray level difference threshold value can be determined between the two

intervals, and such threshold value obtained thereby can be used to verify the algorithm stability so as to reduce the false positive rate.

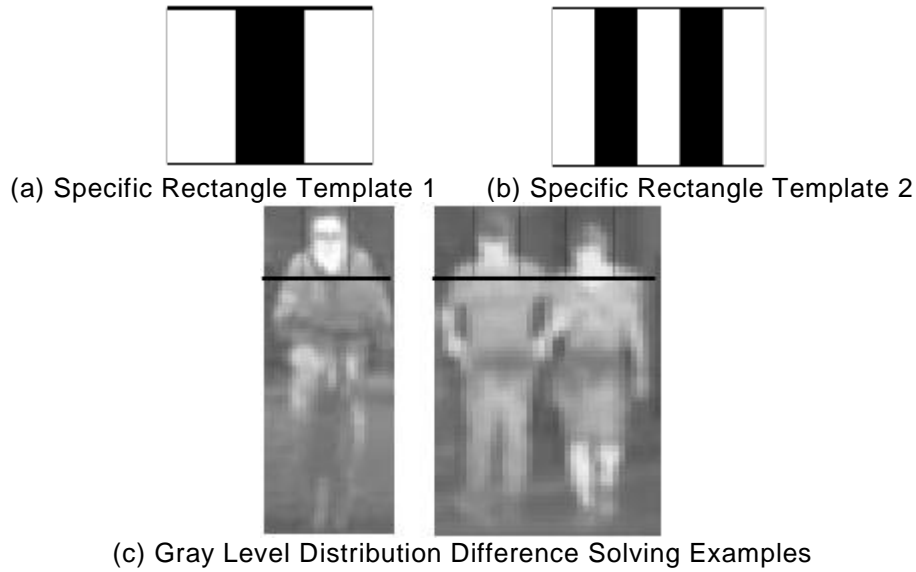


Figure 2. Head Calibration Process for Pedestrian Imaging

4. Experiment Analysis

4.1. Hardware Platform

In order to objectively verify the effectiveness of the proposed method, it is necessary to ensure that the case data can comprehensively reflect the field scene. Mainly because of single scene setting and no consideration of universal interference factors, OCTBVS case data resource (current common benchmark test database) [11] is not adopted for this experiment. Instead, the pedestrian infrared detection equipment based on auto-aided driving monitoring system is established in this experiment and such equipment is composed of onboard thermal infrared imager. Figure 3 shows the schematic diagram of the thermal imager and the carrying vehicle, wherein the horizontal ground is regarded as the reference plane and the thermal imager is set in the vehicle at the height of 26cm. Hardware configuration of the thermal imager: space imaging resolution 384×288 pixels; imaging acquisition rate 25 frame/s; thermal imager focus 3m; angle of view $28^\circ \times 21^\circ$.



Figure 3. Data Acquisition Equipment

During the process of acquiring pedestrian imaging, the vehicle is driven at a constant speed 60 km/h; the imaging acquisition range covers suburban and urban areas and other scenes; the imaging acquisition time is 19:30 ~ 21:30; the imaging

data are saved in avi lossless format; the sequence image resolution is 352×288 pixels.

4.2. Process Setting

In order to comprehensively evaluate the effectiveness of the proposed pedestrian detection algorithm, a certain video test sequence (as shown in Figure 4) acquired by the established platform is selected for the experiment. The above video sequence covers 5 suburban areas and other urban areas, namely S03, S04, S05, S06, S07, U01, U03, U04, U05 and U06. Additionally, the pedestrians in the video are all annotated as the algorithm evaluation standard. The pedestrians are annotated in the video every three frames, and the total length of the video is 250 frames. DR index is evaluated for the detection rate as follows:

$$DR = \frac{n_D}{n_A} \quad (13)$$

In Formula (13), n_D refers to the number of accurate detections and n_A is the total number of pedestrian in the video sequence.

The false positive rate of the proposed algorithm is evaluated according to FPPW (False Positive Per Window), and relevant setting is as the same as the above setting. The condition for accurate pedestrian detection is defined as follows: the specific value of external rectangle B_d (minimum) of the recognized pedestrian and the external standard rectangle B_g is above 60%. Specifically, the following formula is defined:

$$A = \frac{\text{area}(B_d \cap B_g)}{\text{area}(B_d \cup B_g)} \cdot 100\% \quad (14)$$



(a) Pedestrian Target ICOs



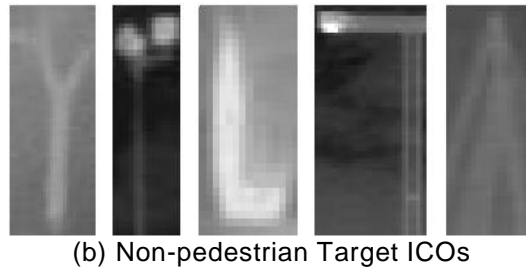


Figure 4. Case Sample

4.3. Classification and Recognition Comparison

The performance of the pedestrian recognition algorithm based on IE_HOG three-branch features is regarded as a pedestrian recognition and classification problem. 5 video clips are selected from the above video test sequence as the experiment cases, and HOG pedestrian detection algorithm is adopted to experimentally analyze and compare the pedestrian recognition. Meanwhile, since the pedestrian recognition is related to feature description and the composition structure of the pedestrian detection classifier [12~13], thus different classifier frameworks are selected for the comprehensive evaluation of pedestrian detection, wherein the three-branch framework is a common structure of pedestrian classifier and such classifier is firstly based on single classifier for pedestrian detection judgment and then based on branch framework for s-SVM algorithm training. The parameter setting of three-branch classifier is as follows:

(1) Close-shot, mid-shot and long-shot pedestrian pixels are respectively set as follows: 32×96 pixels, 24×64 pixels and 12×32 pixels; (2) Image blocks are divided for the adjusted ROIs, and the image block pixels are respectively set as 8×8 pixels, 4×8 pixels and 3×4 pixels; (3) Each image block includes 2×2 photos; (4) In order to minimize the calculation overheads of IE_HOG recognition and HOG recognition, the overlapping parameter of two adjacent image blocks is set as 0; (5) The gradient index is obtained according to the first-order gradient center operator; (6) The image gradient interval is set as 9; (7) The normalized HOG components are obtained according to $L2$ norm.

Notice for parameter setting: in the latter classifier, ROIs parameter is adjusted as 24×64 , and the size of the photo included therein is 4×8 pixels, and other image parameters are as the same as the above corresponding parameters. Figure 5 shows the comparison of DR indexes of four algorithms in five video clips. The pedestrian recognition comparison algorithms adopted thereby can all meet specific limitations, so the corresponding false positive rate is relatively low. During the process of drawing ROC comparison curve, only the pixel intervals with small false positive rate are concerned.

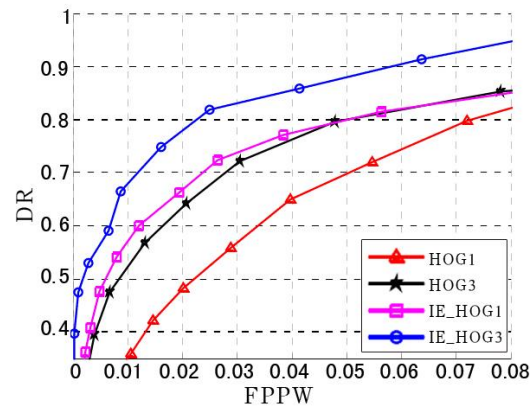


Figure 5. Performance Comparison

According to the experiment data in Figure 5, the classification and recognition structure based on IE_HOG can obtain good pedestrian recognition performance index, so it is necessary to deeply elucidate IE_HOG feature advantages. Obviously, IE_HOG feature extraction is suitable for describing pedestrian infrared detection mode. Meanwhile, the classification and recognition algorithm based on three-branch structure is laterally verified to have better detection and recognition capability.

4.4. Pedestrian Recognition Performance Comparison

According to the above experiment analysis, the three-branch pedestrian classification recognizer has good performance. The algorithms with different structures are adopted in different branch frameworks to train the classifiers so as to evaluate various training methods and accordingly verify the real-time property and accuracy of pedestrian recognition. Table 1 shows the recognition performance parameter comparison for ten video clips, and Tables 1 & 2 show the mean values of the experiment data of different training algorithms in two different scenes, wherein FPS refers to frame processing speed and can be calculated as follows:

$$FPS = \frac{n_f}{t_D} \quad (15)$$

Table 1. Performance Comparison of Classification Models

Model	Scene	Recognition Quantity	FPPW	DR (%)	FPS
Entropy Weight FC-SVM	Urban Area	493/565	8.81	83.72	40.40
	Suburban Area	1436/1489	7.23	89.72	
FC-SVM	Urban Area	461/565	8.24	85.63	25.31
	Suburban Area	1371/1489	7.01	93.49	
s-SVM	Urban Area	479/565	7.62	84.78	18.72
	Suburban Area	1365/1489	6.22	91.67	
HIKSVM	Urban Area	487/565	6.33	86.19	17.14
	Suburban Area	1420/1489	4.93	95.37	

According to Table 1, under the precondition of ensuring the pedestrian detection accuracy (mean false positive rate and mean detection rate are respectively reduced

by 1.10% and 1.51%), the calculation speed of FC-SVM algorithm based on entropy weight is about 1.16 times of that of s-SVM algorithm. Under such algorithm framework, the pedestrian detection speed of FC-SVM algorithm based on entropy weight is about 45frames/s and can meet the real-time requirement for pedestrian detection. Meanwhile, compared with HIKSVM pedestrian detection and recognition algorithm, FC-SVM algorithm based on entropy weight has higher calculation efficiency, with the increment as 1.36 times. Table 1 shows the mean pedestrian detection rate of the system. Obviously, the performance of FC-SVM algorithm based on entropy weight is better than that of common FC-SVM algorithm. Specifically, after the introduction of entropy weight, the urban pedestrian recognition quantity is increased from 461 to 493, urban pedestrian recognition index FPPW is improved from 8.24 to 8.81 and urban pedestrian recognition index FPS is increased from 25.31 to 40.40. In conclusion, relevant performances are obviously improved, mainly because the entropy weight can increase the contrast ratio for image recognition and further improve the detection and recognition accuracy.

5. Conclusion

In order to conquer such problems as large intra-class deviation caused by imaging scale difference and low calculation efficiency in the recognition of pedestrian at night, a head verification and recognition algorithm for pedestrian at night based on entropy weight and FC-SVM is proposed in this paper according to relevant principle of statistics. Firstly, the entropy weight method is adopted to optimize HOG features and describe the local gradient information of pedestrian at night. Then, in order to conquer the disadvantage of large pedestrian intra-class deviation, the recognition and detection algorithm based on three-branch SVM is also introduced in this paper. Then, for the purpose of reducing calculation and storage overheads and ensuring the real-time property, fast classification and detection are realized on the basis of FC-SVM. Finally, in order to improve the matching accuracy, the pedestrian target is deeply evaluated according to the head verification method. Relevant experiment shows that the algorithm proposed for night scene has such features as ideal recognition accuracy, reliable detection and practicability.

References

- [1] J He, Y Geng, F Liu, C Xu, CC-KF: Enhanced TOA Performance in Multipath and NLOS Indoor Extreme Environment, *IEEE Sensor Journal*, vol. 14, no. 10, pp. 3766-3774, Nov. (2015).
- [2] S Zhou, L Mi, H Chen, Y Geng, Building detection in Digital surface model, 2013 IEEE International Conference on Imaging Systems and Techniques (IST), Oct. (2012).
- [3] J He, Y Geng, K Pahlavan, Toward Accurate Human Tracking: Modeling Time-of-Arrival for Wireless Wearable Sensors in Multipath Environment, *IEEE Sensor Journal*, vol. 14, no. 11, pp. 3996-4006, Nov. (2014).
- [4] N Lu, C Lu, Z Yang, Y Geng, Modeling Framework for Mining Lifecycle Management, *Journal of Networks*, vol. 9, no. 3, pp. 719-725, Jan. (2014).
- [5] Y Geng, K Pahlavan, On the accuracy of rf and image processing based hybrid localization for wireless capsule endoscopy, *IEEE Wireless Communications and Networking Conference (WCNC)*, Mar. (2015).
- [6] G Liu, Y Geng, Pahlavan, Effects of calibration RFID tags on performance of inertial navigation in indoor environment, 2015 International Conference on Computing, Networking and Communications (ICNC), Feb. (2015).
- [7] J He, Y Geng, YWan, S Li, K Pahlavan, A cyber physical test-bed for virtualization of RF access environment for body sensor network, *IEEE Sensor Journal*, vol. 13, no. 10, pp. 3826-3836, Oct. (2013).
- [8] J Bing, X Ke, L Jiankun. Analysis and management of the operation and management of power generation units during the low load period[J]. *Jiangsu electrical engineering*, (2015).
- [9] H Ying. Research on transmission network planning based on improved hybrid differential evolution algorithm [J]. *Jiangsu electrical engineering*, (2015).

- [10] S Mingchi, Y Dong. Research and optimization of Nanjing electric power communication transmission network [J]. Jiangsu electric engineering, (2015).
- [11] Y Lin, J Yang, Z Lv, W Wei, H Song. A Self-Assessment Stereo Capture Model Applicable to the Internet of Things. Sensors. (2015).
- [12] W Ou, Z Lv, Z Xie. Spatially Regularized Latent topic Model for Simultaneous object discovery and segmentation. The 2015 IEEE International Conference on Systems, Man, and Cybernetics (SMC2015).
- [13] W Ke. "Using Simulation to Explore Distributed Key-Value Stores for Exascale System Services." 2nd Greater Chicago Area System Research Workshop (GCASR). (2013).
- [14] Y Wang, Y Su, G Agrawal. A Novel Approach for Approximate Aggregations Over Arrays. In Proceedings of the 27th international conference on scientific and statistical database management, ACM, (2015).
- [15] Z Lv, A Halawani, S Feng, S ur Rehman, H Li. Touch-less Interactive Augmented Reality Game on Vision Based Wearable Device. Personal and Ubiquitous Computing. (2015).
- [16] J Yang, S He, Y Lin, Z Lv. Multimedia cloud transmission and storage system based on internet of things. Multimedia Tools and Applications. (2016).

Authors



Rui Liang. Rui Liang was born in Sichuan,China,in 1985. He received B.S. degree in computer science and technology from Sichuan University, Chengdu, in 2007, and M.S. degree in computer science from University of Electronic Science and Technology of China(UESTC), Chengdu, in 2010. Currently, he is pursuing his PhD degree at UESTC. His interests include Computer Vision, Machine Learning, Video Semantic analysis.



Qingxin Zhu. Qingxin Zhu received the B.S. and M.S. degrees in applied mathematics from Sichuan Normal University and Beijing Institute of Technology, in Sichuan and Beijing, respectively, both in China, in 1982 and 1984, respectively, and the Ph.D. degree in operational research and cybernetics from Ottawa University, Canada, in 1993.