

Pedestrian Detection Method in Infrared Images Using Maximum Entropy Threshold and Random Forest

Songze Lei, Xiaoping Li, Feng Xiao and Shifang Zhang

*School of Computer Science and Engineering, Xi'an Technological University,
Xi'an 710032, China
lei_sz@163.com*

Abstract

A novel pedestrian detection method based on maximum entropy threshold and random forest classifier is presented in order to detect effectively pedestrian in infrared images. The entropy of target and background is calculated respectively, and the sum of entropy of target and background is maximized to obtain the maximum information between target and background distributions in the image. According to the rule of maximum entropy, the segmentation threshold value is selected to get interest regions. The morphological and experiential methods are used to further eliminate noise effect. The interest region is transformed to eigenvector by wavelet transform. Random forest classifier is applied to distinguish pedestrians from non-pedestrian candidates. Experimental results in different infrared images show that the method presented has high correct detection rate and low false detection rate with fast calculation speed.

Keywords: *pedestrian object; maximum entropy threshold; infrared image; random forest classifier*

1. Introduction

Pedestrian detection is one of the main issues of computer vision. It provides necessary information for many different applications such as night vision, target acquisition, tracking, surveillance and monitoring [1].

IR images has low signal-to-noise ratio (SNR) [2-3] and hence cannot provide proper visualization of the objects captured. IR images looks like either very dark or bright white-based on the cloud presence in the environment. Non-pedestrian objects and background information interrupt the detection. Thus pedestrian detection in IR images is very difficult. Generally, the temperature of people is higher than the environment temperature and their heat radiation is sufficiently high compared to the background. Therefore, in infrared (IR), images pedestrians belong to the upper range in the gray-level scale and are sufficiently contrasted with respect to the surroundings, thus making IR imagery particularly suited to pedestrian localization. Obviously, other objects that actively radiate heat, such as automobiles, trucks, buses, and motorcycles, have a similar behavior; however, people can be recognized thanks to their shape and aspect ratio.

Driven by the decreasing cost of IR sensors, night vision systems have gained more and more interest in recent years, and thus increasing the need of pedestrian detection at night.

Unfortunately, pedestrian detection is a challenging task due to the non-rigidity of human body and environmental factors. Generally, Pedestrian detection [4] includes two phases: regions of interest (ROI) generation and object classification. Segmentation locates multiple ROIs from infrared images and classification identifies pedestrians from the ROI.

Concerning the various approaches proposed in the literature, for ROI generation, thresholding based methods [5], detection based methods [6], and adaptive thresholding

methods [5][7] are commonly used method for segmenting foreground objects. In [7], the algorithm cannot deal with situations like pedestrians walk shoulder to shoulder or pedestrians staying close to other obstacles like road signs or trees. A dual threshold method is used in [8], also fail in some situations like identical pixels at pedestrian area and adjacent background area. Tian *et al.* [9] extract the target regions from the raw data by using an adaptive thresholding based image segmentation algorithm. Template matching methods include improved template matching [10] and scale invariant template matching [11]. These matching methods fail to detect the shape of the pedestrian and are sensitive to noise.

Broggi *et al.* [12,13] use vertical symmetry derived from gray levels and vertical gradient magnitude to select candidate regions around each relevant symmetry axis. Shashua *et al.* [14] obtain 75 ROIs/frame by filtering out windows based on the lack of distinctive texture properties and noncompliance with perspective constraints on the range and size of the candidates. Cao *et al.* [15] perform an exhaustive scan on the particular rectangle region in which the pedestrians might cause a collision with the vehicle to generate the candidates.

Once the ROI have been obtained, different feature extractions and pattern classifiers can be applied to distinguish pedestrians from non-pedestrian candidates. The shape, appearance, and motion features are the most important cues for pedestrian detection, such as the raw image intensity [16], gradient magnitude [17], and edge [18]. Triggs uses Histogram of Oriented Gradient (HOG) [19,20] feature.

Referring to the classification techniques, the typical classification techniques are template matching [18], Artificial Neural network (ANN) [21], Support Vector Machines (SVM) [16] [20] and Adaboost [15]. Obviously, SVM is the most popular learning method and often produces accurate classifications with most of the features, but it requires heavy computation if there are lots of candidates. The Random Forest (RF) which combines several randomized decision trees, is one of the most popular tree-based classifiers and has produced comparable accuracy in various applications, such as human detection, smoke detection, and object detection.

In this work, the pedestrian detection method in infrared images using maximum entropy threshold and random forest is presented. The proposed method consists of two modules, smart regions of interest (ROI) generation and object classification that is different from others.

The ROI generation module includes image segmentation and candidate selection. The image segmentation adopts maximum entropy threshold method to determine the foreground based on the probability distribution in IR images. The maximum entropy threshold of image is interpreted as indicative of maximum information transfer. The image foreground and background are considered as two different signal sources, so that when the sum of the two class entropies reaches its maximum, the image is said to be optimal threshold. Entropy based segmentation method is the one of the best segmentation methods [22]. The proposed maximum entropy threshold method can segment multiple ROIs that contain exact pedestrian objects. Once the ROIs have been obtained, feature is produced by wavelets transform and random forest classifier is used to recognize the pedestrian objects. Experimental results demonstrate proposed method's effectiveness.

2. Regions of Interest Generation

2.1. Image Segmentation Based on Maximum Entropy Threshold

Entropy is a measure of image information content. Let X denote an image of size $M \times N$ with L gray levels, and $f(x, y)$ is the gray level of a pixel in (x, y) , $f(x, y) \in \{0, 1, \dots, L-1\}$. Let N_i is i th gray-level frequency. Image entropy is defined as follows:

$$H = -\sum_{i=0}^{L-1} p_i \ln(p_i), \quad p_i = \frac{N_i}{N} \quad (1)$$

Let p_1, p_2, \dots, p_n are the gray-level frequencies, and s is the threshold value. The gray of image is divided into target gray and background gray. The target gray distribution is:

$$A: \frac{p_1}{U_s}, \frac{p_2}{U_s}, \dots, \frac{p_s}{U_s} \quad (2)$$

The background gray distribution is:

$$B: \frac{p_{s+1}}{1-U_s}, \frac{p_{s+2}}{1-U_s}, \dots, \frac{p_n}{1-U_s} \quad (3)$$

where,

$$U_s = \sum_{i=1}^s p_i, \quad 1-U_s = \sum_{i=s+1}^n p_i \quad (4)$$

The target entropy is:

$$H(A) = -\sum_{i=1}^s \frac{p_i}{U_s} \ln \frac{p_i}{U_s} = -\frac{1}{U_s} \left[\sum_{i=1}^s p_i \ln p_i - U_s \ln U_s \right] = \ln U_s + \frac{H_s}{U_s} \quad (5)$$

The background entropy is:

$$\begin{aligned} H(B) &= -\sum_{i=1+s}^n \frac{p_i}{1-U_s} \ln \frac{p_i}{1-U_s} = -\frac{1}{1-U_s} \left[\sum_{i=s+1}^n p_i \ln p_i - (1-U_s) \ln(1-U_s) \right] \\ &= \ln(1-U_s) + \frac{H_n - H_s}{1-U_s} \end{aligned} \quad (6)$$

The sum of target entropy and background entropy is defined as follows:

$$\psi(s) = H(A) + H(B) = \ln U_s (1-U_s) + \frac{H_s}{U_s} + \frac{H_n - H_s}{1-U_s} \quad (7)$$

Optimal threshold can be selected by maximizing the entropy of target and background. The maximum information is obtained by maximized $\psi(s)$ between the target and background distributions in the image. The discrete value s which maximizes $\psi(s)$ is the threshold value.

2.2. Interesting Candidate Regions Generation and Selection

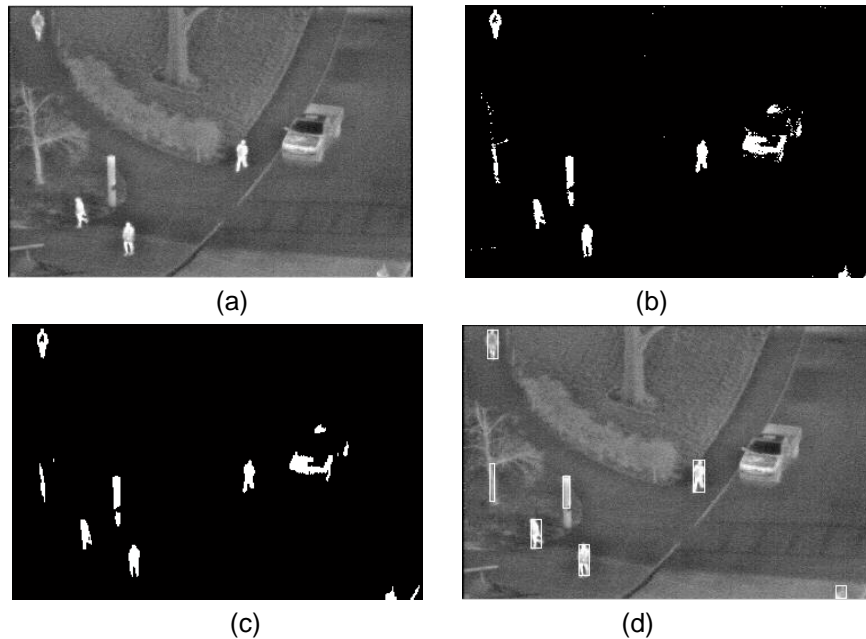
The steps of ROI generation which includes image segmentation based on maximum entropy threshold and candidate selection are as follows:

Step1: Image preprocessing. Median filter is used to filter noise, which is benefit to the subsequent steps.

Step2: Thresholding. The threshold for segment is obtained by maximum entropy threshold method to get binary image of pedestrian object candidates.

Step3: Postprocessing. Morphological method and experience-based judgment are used for eliminating interfering regions or noise. There are many noise points after image preprocessing and thresholding. The open-close morphology operation is used to eliminate noise and small object with vertical rectangle-shaped structure element. The

characters such as shape, size and position are utilized to remove the interfering regions. Considering the width-to-height ratio of pedestrian, which is about 0.4, some interfering regions can be eliminated when width-to-height ratio is larger than the threshold. To prevent misjudgement, the width-to-height ratio could be set 0.8 to 0.9. Figure 1 shows the process of Regions of Interest Generation.



(a) Original Image (b) Maximum Entropy Thresholding after Filtering (c) Processing by Morphology Operation (d) Candidate Regions

Figure 1. The Process of Our Method

3. Feature Extraction

The train samples are selected from candidate regions segmented by maximum entropy threshold. Sample images are normalized to 30×16 size, and done histogram equalization to eliminate the influence of illumination. Part enlarged samples are shown in Figure 2.

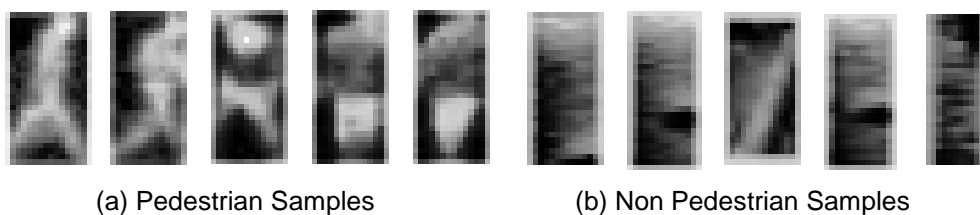
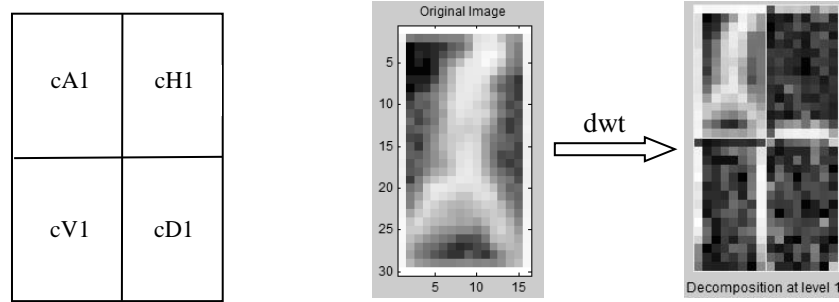


Figure 2. Part Samples of Train Sets

Feature vector of candidate region is extracted using wavelet transform. We select a proper wavelet function *coif1* to decompose image at one level. The low frequency component, horizontal, vertical and diagonal high frequency components are obtained. The four components are combined into a column vector, which is the desired feature of ROI.

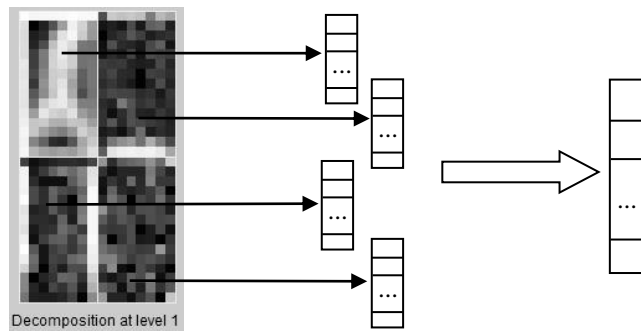
Figure 3 is a sample of pedestrian image wavelet decomposition. The original image is decomposed into four components: the approximation (cA1), and the details in three orientations: horizontal (cH1), vertical (cV1), and diagonal (cD1).



(a) Wavelet Decomposition (b) Wavelet Decomposition at One Level

Figure 3. Wavelet Decomposition of Pedestrian Image Sample

Figure 4 shows feature extracting method. Firstly the ROI image is decomposed into four components by wavelet decomposition. Then each $m \times n$ component image is mapped to $(m \times n) \times 1$ single vector, as shown in Figure 4 (b). Lastly 4 $(m \times n) \times 1$ vectors are connected to $(4 \times m \times n) \times 1$ feature vector, as shown in Figure 4 (c).



(a) Wavelet Decomposition Components (b) Combined Vector of Each Wavelet Decomposition Component (c) Feature Vector of Image

Figure 4. Flow Chart of Feature Extraction Applying Wavelet Decomposition

4. Pedestrian Classification Based on Random Forest

Random Forest (RF) is an ensemble classification method [23]. RF is a combination of multiple decision trees, where each tree is constructed by evaluating a random subset of feature dimensions.

During training, a collection of samples propagates through the tree. At each node, the training set is split in two new sets and each set is transferred to one of the two child nodes. This process is repeated until a stopping criterion is reached, such as the maximum node depth, or the minimum number of samples in a node. The split function maximizes the homogeneity after the split, thereby implicitly maximizing the homogeneity in the leaf nodes. A random subset of feature dimensions is tested at each split, after which the best single or subset of feature dimension is used for splitting. Because of this randomization, RF requires strong single feature dimensions and a large number of training samples per dimension. Consequently, high-dimensional data is challenging. Also, bagging is used, where the initial training set is randomly sampled with replacement.

During testing, each test sample traverses all decision trees in the forest. In each tree, the unknown sample reaches a single leaf node. The tree returns the class distribution of this leaf, normalized to the prior distribution of the training set of the tree. The final classification score is constructed by combining the distributions of the individual trees.

The probability value of classifier is output at each tree. In learning step, random forests randomly select feature dimensions to split tree nodes. Information gain is applied

in node splitting. A leaf node has a probability distribution, and random forests perform classification with bagging algorithm.

The equation of random forests classification shown below:

$$p(c | v) = \frac{1}{T} \sum_{t=1}^T p_t(c | v) \quad (8)$$

where T is the number of trees in random forests ($t \in \{1, 2, \dots, T\}$), v is the input feature vector, c is the class of activity. $p(c/v)$ is the predicted value of posteriori by t -th decision tree.

Forest structure. The structure of the forest can be changed by altering the maximum depth of each tree and the maximum number of trees in the forest. Increasing the depth of each tree results in stronger trees. However, the maximum depth of the tree should be set in relation to the problem complexity, as shallow trees cannot model the required complex boundary, whereas allowing too many splits can result in over-training. By increasing the number of trees, more feature dimension combinations are tested and the generalization of the forest increases. Adding more trees always increases classification accuracy, until it converges to the maximum achievable accuracy. However, computation time also increases linearly with the number of trees.

Randomization. The degree of randomization is the main characteristic of the RF and needs to be carefully chosen. The degree of randomization is defined as the number of feature dimensions tested at each split. By growing degree of randomization, the strength of each split increases, but the correlation between the trees also increases. When degree of randomization is reduced, individual trees become weaker, but the correlation between the trees is lower. RF has the best performance when individual trees are strong and the correlation amongst trees is low, thus the key is to balance both, using the degree of randomization.

Split function. The amount of randomization in the RF is also influenced by the split function. The split function optimizes the homogeneity of the train set, commonly performed by maximizing the Gini index or the Information Gain per split. The most common split functions distribute samples using a single feature dimension. More information can be injected into the tree by evaluating more dimensions in each node.

During the course of our experiment, we vary the maximum depth from 4 to 20 and the maximum allowed trees from 10 to 100 using wavelet features. The accuracy converges at a depth of 9 and a maximum of 50 trees. It should be noted that the result stays constant when increasing the depth or maximum number of trees beyond the values depicted in the graph. Literature states that increasing the number of trees will increase the generalization accuracy until a maximum is reached, which is confirmed by our experiments.

There are 306 pedestrian samples and 431 non-pedestrian samples to be used for the train experiment. The feature vector is sent into RF classifier to train. When the image is detected, the object segmentation is first made by maximum entropy threshold to get candidate regions. Then the sub-image of candidate region is normalized to 30×16 size with histogram equalization, and transformed into feature vector. The candidate region is recognized by the trained random forest classifier.

5. Experiments Results

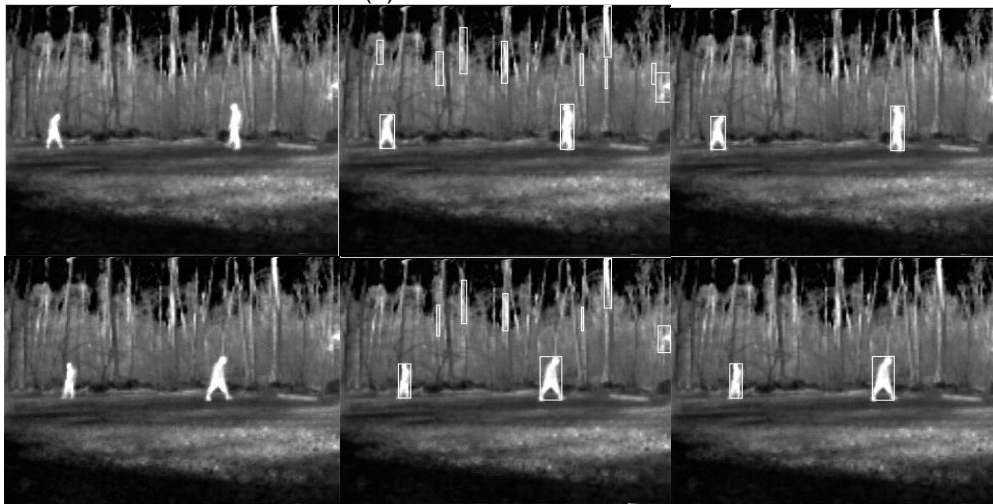
To evaluate our proposed method, we have used part images of OTCBVS (Object Tracking and Classification Beyond the Visible Spectrum) Benchmark Dataset [24] and Terravic Motion IR Database [25]. The image sizes of two database are 360×240 and 320×240 separately. All the calculation is implemented by using Matlab 7 on an Intel i5 2.67 G processor with 4 GB memory.

The test dataset is selected from part of OTCBVS Benchmark Dataset and Terravic Motion IR Database, which does not overlap with the train dataset. Test dataset includes single or multiple pedestrian in infrared image. Test dataset is divided into two sub-datasets, which are test set 1 and test set 2. Test set 1 is selected from OTCBVS Benchmark Dataset with 101 images and 386 pedestrians. Test set 2 is selected from Terravic Motion IR Database with 316 images and 632 pedestrians.

Figure 5 shows detection results of two test sets. The images are source image, candidate image and final detection image from left to right in every row respectively. The candidate regions and pedestrians are labelled with rectangle in image. As shown in Figure 5, although candidate regions include some non-pedestrian by maximum entropy threshold, they are well judged by trained random forest classifier. It is shown that the proposed infrared image segment and classification methods are effective.



(a) Result on Test Set 1



(b) Result on Test Set 2

Figure 5. Pedestrian Detection Result

The results of two test sets and average detection time are given in Table 1. As shown in Table 1, detection rate is equal to or greater than 95.9%. As the image brightness of pedestrian object is greater than its background on test set 2, and also pedestrian gray is more uniform than other regions, the detection rate on test set 2 is better than it on test set 1.

Table 1. Pedestrian Detection Result

	Image size	Image number	Pedestrian number	False detection rate/%	Detection rate/%	Average detection time/s
Test set 1	360×240	101	386	1.8	95.9	0.18
Test set 2	320×240	316	632	2.5	100	0.21

The test image set has a special situation, as shown in Figure 6. Two pedestrians walk in the opposite direction, and overlap when they meet in the middle of image. After image is segmented, the candidate regions may include two overlapping pedestrians. It is hard to separate overlapping pedestrian region into two pedestrians. In fact, we need only locate the position of pedestrian, and do not care about the number of the pedestrian at the position. While the overlapping pedestrian region is trained by RF classifier, it can get the correct result.

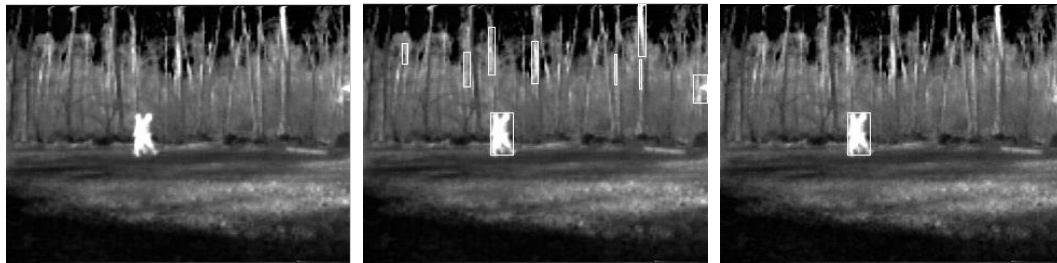


Figure 6. Overlapping pedestrian detection result of test set 2

There is some error in the test set 1 that some pedestrian candidates do not detect. This is mainly because the quality of some images in test set 1 is poor and pedestrian object is very small with uneven gray. Also, there is some false detection in test set 2, but it is low.

The average detection time is 0.1~0.3 second in Matlab, and detection time can be shortened in other real-time system. The proposed method fully satisfies real time need according to the experiment results.

Compared with the classical method HOG+SVM [20] which was used by Dalal and Triggs for pedestrian detection, as shown in Table 2, our method gets better result. The detection rate of HOG+SVM is 93.8% in the Test set 1 while detection rate of our method is 95.9%. Also false detection rate of HOG+SVM are higher than the rate of our method. For the characteristic of Infrared image, the Infrared image of the dataset has not rich gradient information, and gradient information of image is extracted by HOG. Therefore the HOG feature is not suitable for classification and recognition in the Infrared image of the dataset. Our wavelet feature includes four kinds of information, which have high discriminability for recognition. Random forest classifier and SVM have similar performance. Nevertheless data preprocessing will be necessary for the SVM, while it will not be necessary random forest classifier.

Table 2. Results of Different Methods

	False detection rate/%	Detection rate/%
HOG+SVM[20]	2.8	93.8
Our method	1.8	95.9

6. Conclusions

The applications of pedestrian detection in infrared image are becoming increasingly wider. Infrared image is formed by infrared thermal imaging technology, and it is hardly influenced by illumination and weather *et al.* For body temperature is higher than environmental temperature, threshold segment method is efficient way to obtain interesting candidate regions. Entropy is the good segment method as it represents distribution of target and background. The method based on maximum entropy threshold obtains good ROIs with less time. The feature is extracted by wavelets from ROIs and is suitable for classification in the Infrared image of the dataset.

About the classifier, the random forest and SVM have no obvious difference in generalization capability for two-classification problem. Also two classifier have no obvious difference in robustness for noise of the data classes. However data preprocessing will be necessary for the SVM, while it will not be necessary for random forest classifier. The experiment results show that the method is efficient to detect pedestrians in the Infrared image.

Acknowledgements

This work is supported by the National Natural Science Foundation of China (Grant No. 61572392) and President Fund Project of Xi'an Technological University (Grant No. XGYXJJ1006), and National Joint Engineering Laboratory of New Network and Detection Foundation (Grant No. GSYSJ2016008).

References

- [1] S. Rajkumar and P.C. Mouli, "Perceptual visualization enhancement of infrared images using fuzzy sets", *Transactions on Computational Science XXV*, Springer, (2015), pp.3-19.
- [2] S. Rajkumar and P.C. Mouli, "Target detection in infrared images using block-based approach", *Informatics and Communication Technologies for Societal Development*, Springer, (2015), pp.9-16.
- [3] R. Liu, X. Li, L. Han and J. Meng, "Track infrared point targets based on projection coefficient templates and non-linear correlation combined with kalman prediction", *Infrared Physics & Technology*, vol. 57, (2013), March, pp.68-75.
- [4] T. Gandhi and M. Trivedi, "Pedestrian protection systems: Issues, survey, and challenges", *IEEE Trans. Intell. Transp. Syst.*, vol. 8, no. 3, (2007), pp. 413-430.
- [5] J. Ge, Y. Luo, and G. Tei, "Real-Time Pedestrian Detection and Tracking at Nighttime for Driver-Assistance Systems", *IEEE Transactions on Intelligent Transportation Systems*, vol. 10, no. 2, (2009), pp.283-298.
- [6] M. Soga, S. Hiratsuka, H. Fukamachi, and Y. Ninomiya, "Pedestrian Detection for a Near Infrared imaging System", *IEEE Conference on Intelligent Transportation Systems*, China, Beijing, (2008) October 1167-1172.
- [7] Q. M. Tian, Y. P. Luo, and D. C. Hu, "Pedestrian detection in nighttime driving", in *International Conference on Image and Graphics*, Hong Kong, China, (2004) December 116-119.
- [8] J. Dong, J. Ge, and Y. Luo, "Nighttime Pedestrian Detection with Near Infrared using Cascaded Classifiers", in *IEEE International Conference on Image Processing*, San Antonio, Texas, USA (2007) September VI - 185-VI - 188.
- [9] Q. Tian, H. Sun, Y. Luo, and D. Hu, "Nighttime pedestrian detection with a normal camera using SVM classifier," in *Second International Symposium on Neural Networks*, Chongqing, China, (2005) May 189-194.
- [10] R. Liu, Y.Lu, C. Gong and Y. Liu, "Infrared point target detection with improved template matching", *Infrared Physics & Technology*, vol. 55, no. 4, (2012), pp. 380-387.

- [11] J. Yoo, S.S. Hwang, S.D. Kim, M.S. Ki and J. Cha, "Scale-invariant template matching using histogram of dominant gradients", *Pattern Recognition*, vol. 47, no. 9, (2014), pp. 3006-3018.
- [12] A. Broggi, M. Bertozzi, A. Fascioli, and M. Sechi, "Shape-based pedestrian detection," in *Proceedings of IEEE Intelligent Vehicles Symposium*, Dearborn, USA (2000) October 200-215.
- [13] M. Bertozzi, A. Broggi, A. Fascioli, A. Tibaldi, R. Chapuis, and F. Chausse, "Pedestrian localization and tracking system with Kalman filtering," in *Proceedings of IEEE Intelligent Vehicles Symposium*, Parma, Italy, (2004), June 584-589.
- [14] A. Shashua, Y. Gdalyahu, and G. Hayun, "Pedestrian detection for driving assistance systems: Single-frame classification and system level performance," in *Proceedings of IEEE Intelligent Vehicles Symposium*, Parma, Italy, (2004) June 1-6.
- [15] X. Cao, H. Qiao, and J. Keane, "A low-cost pedestrian-detection system with a single optical camera," *IEEE Transactions Intelligent Transportation Systems*, vol. 9, no. 1, (2008), pp. 58-67.
- [16] F. Xu, X. Liu, and K. Fujimura, "Pedestrian detection and tracking with night vision", *IEEE Transactions Intelligent Transportation Systems*, vol. 6, no. 1, (2005), pp. 63-71.
- [17] L. Zhao and C. Thorpe, "Stereo-and neural network-based pedestrian detection", *IEEE Transactions Intelligent Transportation Systems*, vol. 1, no. 3, (2000), pp. 148-154.
- [18] D. M. Gavrila and S. Munder, "Multi-cue pedestrian detection and tracking from a moving vehicle", *International Journal of Computer Vision*, vol. 73, no. 1, (2007), pp. 41-59.
- [19] Yu-Chun Lin, Yi-Ming Chan, Luo-Chieh Chuang, Li-Chen Fu, Shih-Shinh Huang, Pei-Yung Hsiao and Min-Fang Luo, "Near-Infrared Based Pedestrian Detection by Combining Multiple Features", *IEEE Conference on Intelligent Transportation Systems*, Washington, DC, USA, (2011) October 1549-1554.
- [20] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection", *International Conference on Computer Vision and Pattern Recognition*, San Diego, CA, USA, (2005) June 886-893.
- [21] L. Zhao and C. Thorpe, "Stereo-and neural network-based pedestrian detection", *IEEE Transactions Intelligent Transportation Systems*, vol. 1, no. 3, (2000), pp. 148-154.
- [22] Mehmet Sezgin and Bülent Sankur, "Survey over Image Thresholding Techniques and Quantitative Performance Evaluation", *Journal of Electronic Imaging*, vol. 13, no. 1, (2004), pp. 146-165.
- [23] C.D. Geelen, R.G.J. Wijnhoven, G. Dubbelman and P.H.N.D. With, "Gender Classification in Low-Resolution Surveillance Video: In-depth Comparison of Random Forests and SVMs", *Proceedings SPIE 9407, Video Surveillance and Transportation Imaging Applications*, San Francisco, United States, (2015) March.
- [24] Davis J and Keck M, "A two-stage approach to person detection in thermal imagery", *IEEE Workshop on Applications of Computer Vision*, vol. 1, Washington, USA, (2005) 364-369.
- [25] Roland Mieziako, *Terravic Research Infrared Database*. <http://www.terravic.com>.

Author



Songze Lei, Associate Professor. He received PhD in information and communication engineering from Northwestern Polytechnical University. Now he works at Xi'an Technological University. His main research interests include image processing and pattern recognition.