DETECTION AND TRACKING OF PEDESTRIAN CROSSWALKS USING RECURRENT NEURAL NETWORK CLASSIFIER AND HISTOGRAMS OF ORIENTED GRADIENTS IN THE INFRARED IMAGE SEQUENCE

SEYED MOHAMMAD SHAHROKHI[1], MOHAMMADREZA AMINI[2]

[1,2]Department of Electrical Engineering, Borujerd Branch, Islamic Azad University, Borujerd, Iran

ABSTRACT

Recently, detection of pedestrian crosswalks has been identified as one of the important issues in the field of image processing and statistical identification. The simultaneous detection and tracking of pedestrians are much important, however, there are challenges such as time-consuming and uncertainty in determining the position of a person. In the past, automated methods have been proposed that often have low accuracy and uncertainty in achieving an optimal response. These methods lack comprehensiveness and suffer over fitting problem. In this paper, an algorithm is proposed for the automatic detecting of the pedestrians' position and tracking based on an efficient approach consisting of recurrent neural networks and the gradient histogram in infrared images, which is more accurate and faster than the similar methods. In the first step, the best features are selected by the gradient histogram algorithm and then detection and tracking will be done in the recurrent neural network algorithm. A K-fold validation technique is used to divide the training and test data with a value of K equals to 10. The proposed algorithm has an acceptable performance with an error of less than 5% in detecting and tracking pedestrians.

Keywords: Pedestrian crosswalks detection, Infrared images, Tracking, Recurrent neural network, Histograms of oriented gradients

1. INTRODUCTION

Visible and infrared lights are one of the radiating and non-visible parts of electromagnetic waves spectrum that contain electromagnetic radiation with a wavelength higher than normal and visible light [1, 2]. This part of spectrum starts from the red edge and ends with visible light and varies from a few hundred nm (430 terahertz) to 1 mm (300 GHz) [3]. Of course, someone might suppose it with the maximum spectral width up to 1000 nm [4, 5]. Most of the thermal radiations

from the objects at room temperature are almost infrared [6, 7], and the largest source of this spectrum emission is the sun.

Infrared wave has employed in various applications, including infrared detectors, telescopes, night vision devices, airplanes, food heating and driving [8]. One of the main application of these waves is to use them in imaging, in which objects that have fewer temperatures are pale, and this property can be used in moving human tracking with the body temperature above ambient temperature [9]. In previous researches, this property has been used to track pedestrians [10]. But in this area, there are also problems that have made human detection difficult. In general, in infrared images, the human format is lighter and brighter than fixed objects in the environment such as trees, streets, and so on. However, detecting people in the visible images is a complicated and difficult process and it is not possible to detect people and separate them simply. For example, objects such as transformers, animals, electric boxes, cars, especially in the summer, complicate human tracking. Another problem is that if supposed that tracking of the people is to be sent elsewhere, the large amount of data cannot be effective. On the other hand, if objects or targets are not large in video frames, then the tracking process will be greatly complicated. Another major problem seen in this area is the inability to track the multi-objective targets or the human in the infrared image sequence, which received less attention in the literature.. For this reason, an efficient and robust pedestrian tracking algorithm is important in thermal and infrared image sequence. Considering the importance of using images in human video surveillance, one or more of the following main aims may be noticed [11]:

❖ Protection and safety
❖ Control, management and surveillance
❖ Training and research

Today, human video surveillance systems are used to protect and monitor important, high-risk, and crowded places. Museums, stores, prisons, airports, metro, and railway stations [12, 13], and even hospitals and schools [14, 15] are other places that are equipped these systems to track people. By existing of this system, it will be possible to detect and track special people more easily [16].

2. PREVIOUS METHODS

Research in this field has followed various goals; however, there are several methods to detect pedestrians that can be categorized according to the type of tracking technique they used:

1. Moving based methods [9, 12]
2. Intensity methods [11, 13]

On one hand, it is possible to categorize according to the classification tools:

1. Neural networks [14]
2. Adaptive AdaBoost [3]
3. Support vector machine [17]

Finally, the techniques used in tracking will also be as follow:

1. Kalman filter
2. Time graph
3. Particle filter
4. Gaussian mixture model [4]
5. Maximal mathematical algorithm [18]

Elguebaly et al. (2013) [4] studied the use of asymmetric Gaussian mixture models to detect targets in infrared images.

In 2014, Teutsch et al. [5] performed pedestrian detection in infrared images by the Hot Spot classification. Authors in [6] tracked pedestrians through infrared image fusion for the various visible spectrums. In 2015, Soundrapandiyan et al. [7] tracked targets by adaptive thresholding and background subtracting of sample infrared images. Another research in 2015 by Rajkumar et al. [8] was suggested the use of local thresholding for detecting of the pedestrians. Berg et al. [9] proposed object tracking with different temperatures in 2015, which was assumed as a basis factor. Recently, extensive researches have been done in this field [15-10] and most researchers have been looking for new methods with the proper accuracy and performance [17-23].

3. PROPOSED METHOD

The proposed method is a combination of feature extraction by gradient histogram and the use of the recurrent neural network for the classification, as shown in Fig. 1. The tracking process is based on the Kalman filter method which is described in the subsequent sections.
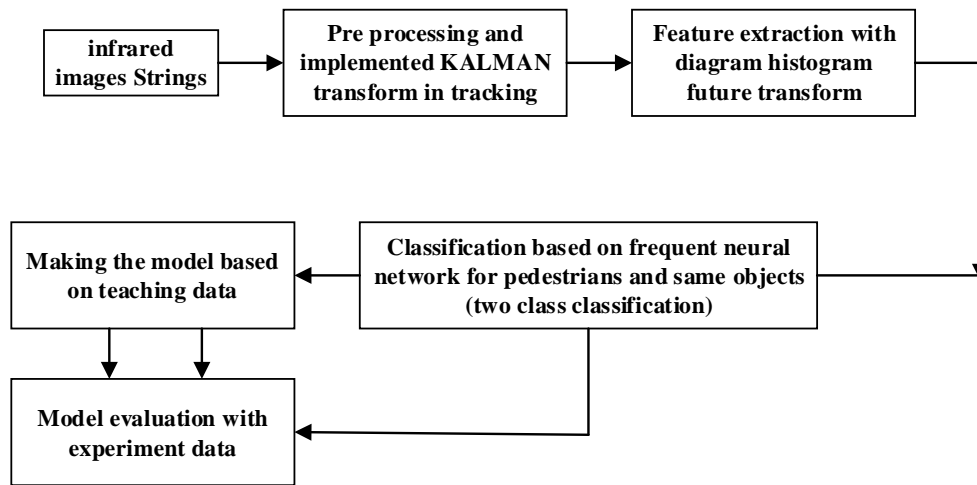
```
┌──────────────┐     ┌──────────────────┐     ┌──────────────────┐
│   infrared   │     │ Pre processing and│    │Feature extraction with│
│images Strings│ ──▶ │implemented KALMAN│ ──▶ │ diagram histogram │
│              │     │transform in tracking│  │  future transform │
└──────────────┘     └──────────────────┘     └──────────────────┘
                                                          │
                                                          ▼
┌──────────────────┐    ┌──────────────────────────────┐
│Making the model  │    │Classification based on frequent neural│
│based on teaching │ ◀──│network for pedestrians and same objects│
│data              │    │  (two class classification)   │
└──────────────────┘    └──────────────────────────────┘
        │                              │
        ▼                              │
┌──────────────────┐                  │
│Model evaluation  │ ◀────────────────┘
│with experiment   │
│data              │
└──────────────────┘
```

Fig. 1: Detection model and pedestrian tracking

3.1 Kalman Filter

Kalman filter predicts a new state of the system, compares the prediction value with the measured value, and weighs the difference of predicted and measured value, and creates a new estimation. The estimation at the time is displayed t_i in equation (1):

$$\hat{E}(t_i) = G(t_i).[x(t_i) - M(t_i).\hat{P}(t_i)] \qquad (1)$$

In this equation, $G(t_i)$ is the Kalman gain, $M(t_i)$ is the measurement matrix, $x(t_i)$ is the system input, and $\hat{P}(t_i)$ is the predictive value obtained in accordance with equation (2):

$$\hat{P}(t_i) = A(t_i).\hat{E}(t_{i-1}) \qquad (2)$$

In this equation, $A(t_i)$ is also called the system matrix, and the gain matrix is obtained from the covariance matrix of the error. The Kalman filter algorithm is recursive; it means that the estimation includes all the system information in the past, without saving all measured values. The intensity of a pixel is estimated as a state of the system to be assumed as a background pixel. In this case, the threshold is considered based on the equations that if the moving object in the image follows the equations, then it will be a part of the background; otherwise it is a part of the foreground.

3.2 Feature extraction

The main idea in the histograms of oriented gradients method is that the distribution of local gradients or edge orientations can describe the image well, even if we do not have information about the exact position of the gradient or the corresponding edges. This feature shows the orientation of the image in a local neighbourhood. Each of these is called a cell. Depending on whether the gradient is unsigned or signed, the distance between 0-180 degrees or 0-360 degrees are divided into *n* equal parts, which *n* represents the number of gradient orientations or the histogram intervals, and each of these distances is a histogram channel. In each cell, the histogram of the oriented gradients is calculated for the pixels within the cell. Then, to resistance intensity changing, the histograms create a block and normalize. Normalized histograms are called oriented gradients. Blocks can be chosen as overlapped and non-overlapped, and usually, choose them overlapped. However, the length of the feature vector increases, but in this case, the accuracy of the performance will improve. For colour images, the gradient for each colour channel is calculated individually and the largest value for each pixel is calculated as the gradient vector in each cell.

3.3 Recurrent neural network

The best dynamic model that can be adapted to the type of the new data is the neural network model. The reason for this choice is that it can be achieved to efficient models by adjusting weights. It is noteworthy that the exact adjustment parameters cannot be assumed to be effective in this model, except the number of neurons in each layer and the number of layers corresponding to the problem. To this end, we use the repetition process as one of the methods for capturing the best neural network classifier model that provides the highest level of accuracy. The proposed model has solved in the following:

❖ Solve the overfitting problem:

The model is only created by using training patterns, but if another data, even a bit far from the training set, is applied to the model, the model is not able to correctly answer the new data and classified them with a lot of error.

❖ Solve the underfitting problem:

The algorithm obtains a general model of the training set, but cannot decide on the data with which it has made the decision model. Usually, the structure of the neural networks avoids underfeeding.

❖ Possibility to save the best model:

Due to that, we need to test after finding the best model, we save the model so that it can decide with new data.

❖ Data split into three parts, training, validation and testing data:

The model is based on training data, and validation data is used in the K-fold cross-validation step and finally, the selected model is tested with test data.

4. EXPERIMENTAL RESULTS

The results of the simulation include tracking and the classification step, which is mentioned separately. In the tracking step, the Kalman filter is used, and in the second step, the detection is done using histograms of oriented gradients and recurrent neural network.

❖ Tracking step

In the tracking step performed by the Kalman algorithm, the pedestrian location is found, and for each mapped template onto the body of the person, the desired features are extracted. In Figure 2, an example of the primary tracking is displayed. Also, in Figure 3, the result of tracking is displayed beside the color frames.

**Fig. 2: Results of the first step of the tracking algorithm**
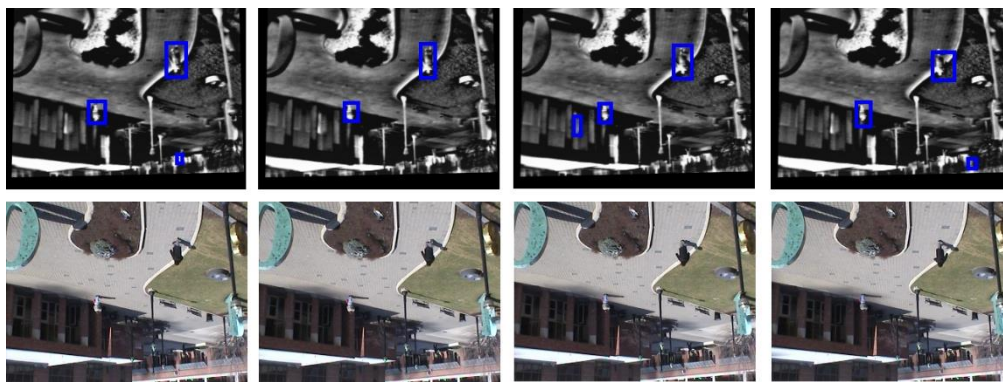
Fig. 3: Tracking in infrared image dataset

❖ Classification step

All videos are randomly converted into 20 clips in AVI format with a resolution of 320x340 pixels at 15 frames per second. The entire videos that include pedestrians and thermal objects are in 10 clips, and all the clips similar to pedestrians were placed in 10 other clips. The whole of

these clips took about 30 minutes. Approximately 26438 frames were selected; videos filmed in an environment (pedestrian or objects similar to the human body format) were computed together. For example, Clip 1 and Clip 10 were filmed in an environment and Clip 1 included pedestrians, and the Clip 2 frames contained objects similar to the human body format. Table 1 shows the results.

Figure (4): the cost function for several of random iterations of the algorithm in the test and validated data for the best display, worst model error (high classification error) of the best model (low classification error) from high error to low error.
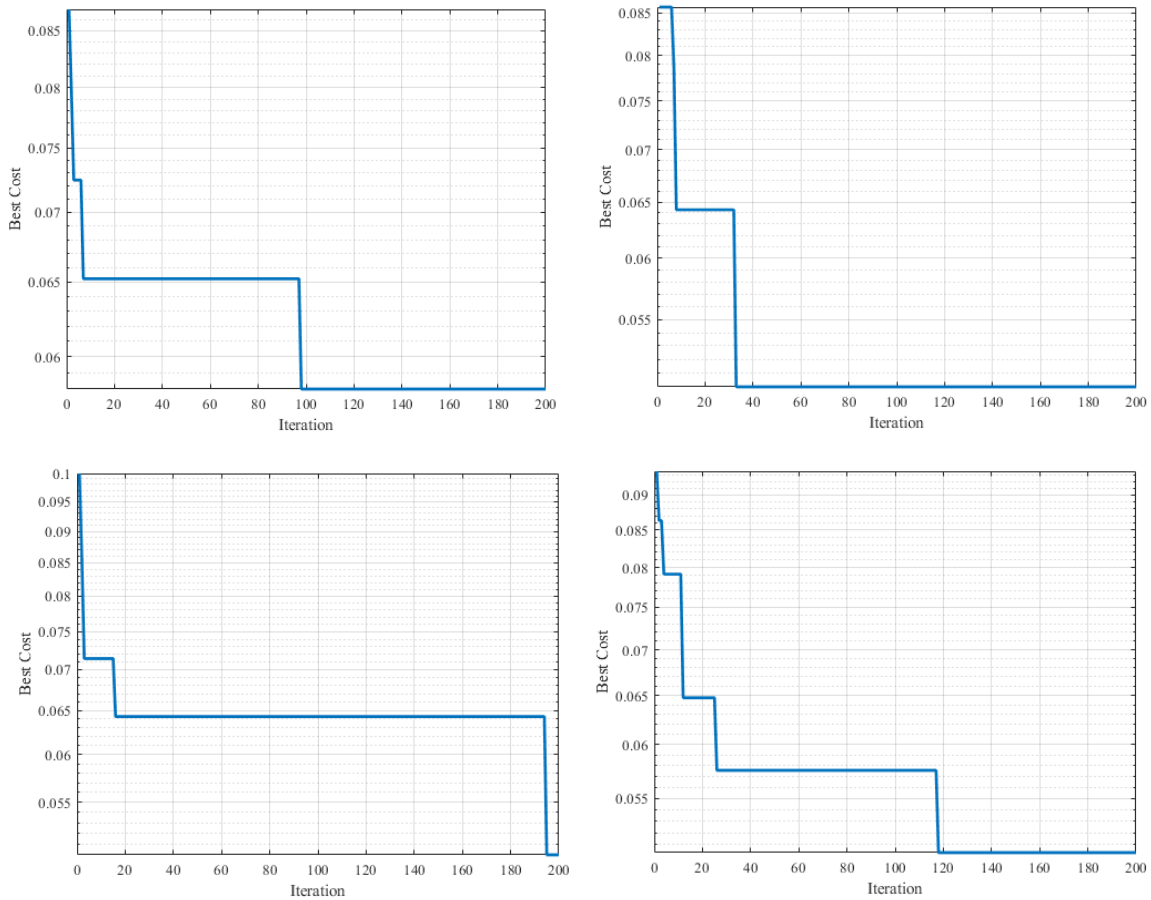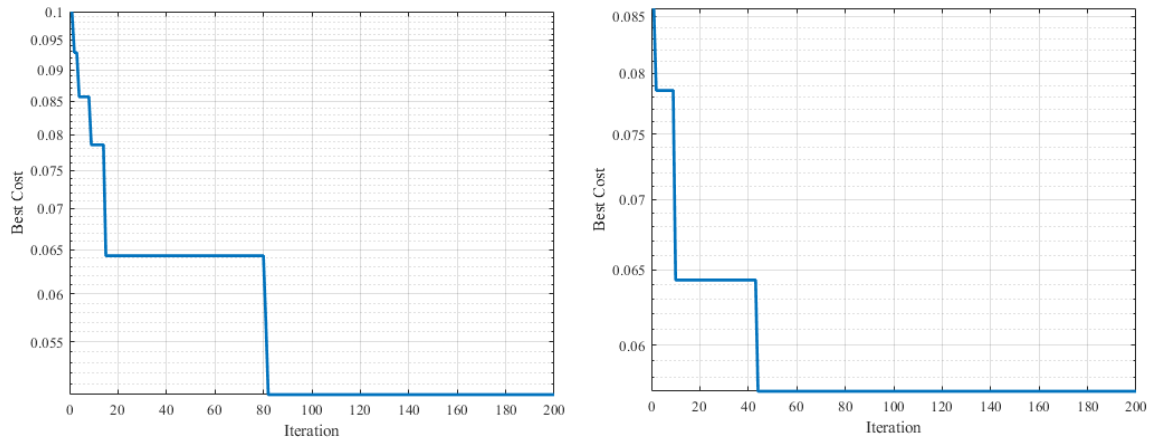
Figure (4): Cost function for selecting the best model from the masses of models selected for the purpose of individual tracking of thermal images; the right column is for validating training data and the left column is for training data to be tested.

In some responses, the cost function is zero, indicating that the performance of the proposed neural network as a cluster is along with extracting the optimal properties and finding the best movement position of the individuals in the thermal images.

Table 1: Results of detecting pedestrians from others by calculation of accuracy and error rate

Video Strings	Number Of Frames	Number Of Frames		The number of frames that had infrared spectrum objects		Number of frames with pedestrians		Average Accuracy (AAC)	Detection Rate (DR)	False Alarm Rate (FAR)
		1	2	N_{TP}	N_{FN}	N_{TN}	N_{FP}			
Clips 1,10	1256	432	824	4	0	11	0	100%	1.00	0
Clips 2,11	3596	792	2804	5	1	18	2	86.46%	0.8333	10%
Clips 3,12	1791	984	807	3	1	9	0	92.37%	0.75	0
Clips 4,13	3356	1511	1845	5	0	16	1	95.45%	1.00	5.88%
Clips 5,14	3345	1376	1969	7	1	23	3	85.23%	0.875	11.53%
Clips 6,15	2096	1195	901	3	0	9	1	92.31%	1.00	10%
Clips 7,16	1697	440	1257	3	0	7	1	90.90%	0.875	12.50%
Clips 8,17	4569	1417	3152	8	2	31	2	90.69%	0.80	6.06%
Clips 9,18	3709	928	2781	5	1	19	4	82.75%	0.8333	17.39%
Clips 10,20	1023	457	566	2	0	10	1	92.30%	1.00	9.09%
Total	**26438**	**9532**	**16906**	**45**	**6**	**153**	**13**	**91.24%**	**0.8823**	**7.83%**

5. CONCLUSION

Pedestrian detection has been recently paid immense attention and most researchers have tried to detect and track at the same time. This process is being considered in various researches. In this paper, we employed Kalman filter algorithm to estimate and track pedestrians. In the next step, after detecting the pedestrians and objects similar to the human body format, by using histograms of oriented gradients and recurrent neural network with the ability to avoid from overfitting, we identified the pedestrian with less than 9% error. In the future, for more data, the implementation process will be done. The application of this model can greatly be effective in public places for the detection of pedestrians.

REFERENCES

1. Zou, H., Sun, H., & Ji, K. (2012, December). Real-time infrared pedestrian detection via sparse representation. In Computer Vision in Remote Sensing (CVRS), 2012 International Conference on (pp. 195-198). IEEE.
2. Wang, J. T., Chen, D. B., Chen, H. Y., & Yang, J. Y. (2012). On pedestrian detection and tracking in infrared videos. Pattern Recognition Letters, 33(6), 775-785.
3. Teutsch, M., & Müller, T. (2013, May). Hot spot detection and classification in LWIR videos for person recognition. In SPIE Defense, Security, and Sensing (pp. 87440F-87440F). International Society for Optics and Photonics.
4. Elguebaly, T., & Bouguila, N. (2013). Finite asymmetric generalized Gaussian mixture models learning for infrared object detection. Computer Vision and Image Understanding, 117(12), 1659-1671.
5. Teutsch, M., Muller, T., Huber, M., & Beyerer, J. (2014). Low resolution person detection with a moving thermal infrared camera by hot spot classification. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops (pp. 209-216).
6. M. R. Amini, M. Moghadasi, and I. Fatehi, "A BFSK Neural Network Demodulator with Fast Training Hints," in *2010 Second International Conference on Communication Software and Networks*, 2010, pp. 578-582.
7. Soundrapandiyan, R., & Mouli, P. C. (2015). Adaptive Pedestrian Detection in Infrared Images Using Background Subtraction and Local Thresholding. Procedia Computer Science, 58, 706-713.
8. Rajkumar, S., & Mouli, P. C. (2015, February). Pedestrian detection in infrared images using local thresholding. In Electronics and Communication Systems (ICECS), 2015 2nd International Conference on (pp. 259-263). IEEE.

9. M. R. Amini and E. Balarastaghi, "Improving ann bfsk demodulator performance with training data sequence sent by transmitter," presented at the Machine Learning and Computing (ICMLC), 2010 Second International Conference on, Bangalore, India, 2010.

10. M. Amini, M. Mahdavi, and M. J. Omidi, "Energy Efficiency Optimization of Secondary Network Considering Primary User Return with Alternating-Phase-Type Traffic," *IEEE Transactions on Communications,* vol. PP, pp. 1-1, 2017Pawłowski, P., Piniarski, K., & Dąbrowski, A. (2015, September). Pedestrian detection in low resolution night vision images. In Signal Processing: Algorithms, Architectures, Arrangements, and Applications (SPA), 2015 (pp. 185-190). IEEE.

11. Yang, C., Liu, H., Liao, S., & Wang, S. (2015). Pedestrian Detection in Thermal Infrared Image Using Extreme Learning Machine. In Proceedings of ELM-2014 Volume 2 (pp. 31-40). Springer International Publishing.

12. Zhao, X., He, Z., Zhang, S., & Liang, D. (2015). Robust pedestrian detection in thermal infrared imagery using a shape distribution histogram feature and modified sparse representation classification. Pattern Recognition, 48(6), 1947-1960.

13. M. Amini and A. Mirzavandi, "Phase-Type Model Spectrum Sensing for Cognitive Radios," *IETE Journal of Research,* vol. 61, pp. 1-7, 2015 2015.

14. Ma, Y., Wu, X., Yu, G., Xu, Y., & Wang, Y. (2016). Pedestrian Detection and Tracking from Low-Resolution Unmanned Aerial Vehicle Thermal Imagery. Sensors, 16(4), 446.

15. Dai, C., Zheng, Y., & Li, X. (2007). Pedestrian detection and tracking in infrared imagery using shape and appearance. Computer Vision and Image Understanding, 106(2), 288-299.

16. Biswas, S. K., & Milanfar, P. (2017). Linear support tensor machine with LSK channels: Pedestrian detection in thermal infrared images. IEEE transactions on image processing, 26(9), 4229-4242.

17. Cai, Y., Liu, Z., Wang, H., & Sun, X. (2017). Saliency-based pedestrian detection in far infrared images. IEEE Access, 5, 5013-5019.

18. Ma, M. (2019). Infrared pedestrian detection algorithm based on multimedia image recombination and matrix restoration. Multimedia Tools and Applications, 1-16.

19. Kwak, J. Y., Ko, B. C., & Nam, J. Y. (2017). Pedestrian tracking using online boosted random ferns learning in far-infrared imagery for safe driving at night. IEEE Transactions on Intelligent Transportation Systems, 18(1), 69-81.

20. Bai, X., Wang, Y., Liu, H., & Guo, S. (2018). Symmetry information based fuzzy clustering for infrared pedestrian segmentation. IEEE Transactions on Fuzzy Systems, 26(4), 1946-1959.

21. Shen, G., Zhu, L., Jihan, L. O. U., Shen, S., Liu, Z., & Tang, L. (2019). Infrared multi-pedestrian tracking in vertical view via Siamese Convolution Network (December 2018). IEEE Access.

22. Lahmyed, R., El Ansari, M., & Ellahyani, A. (2018). A new thermal infrared and visible spectrum images-based pedestrian detection system. Multimedia Tools and Applications, 1-2.