Pedestrian Tracking in Infrared Image Sequence Based on Spatio-temporal Slice Trajectory Analysis

Dongmei Fu and Pan Shu(✉)

School of Automation and Electrical Engineering,
University of Science and Technology Beijing, Beijing 100083, China
panshupansy@gmail.com

Abstract. The research of pedestrian tracking algorithm in infrared image sequence is a curial part in video surveillance. But due to the special characteristics of infrared image, such as low contrast, fuzzy edge and unknown noises interference, the study of infrared pedestrian tracking algorithm becomes a great challenge. Spatio-temporal slice method is an effective analysis method considering both space and time. In view of the existing tracking methods based on spatio-temporal slice only considering horizontal slice analysis and usually with large amount of calculation, this paper proposes a novel tracking algorithm based on spatio-temporal slice trajectory analysis. The proposed algorithm uses multi-layer horizontal and vertical slices to obtain the complete trajectories in order to determine the information of target boundary and position, and then tracks the target in each frame of the infrared image sequence. The experimental results show that the algorithm has relatively high tracking accuracy at a fast computing speed. Moreover, it performs effectively in different infrared image sequences with various motion modes by single pedestrian from OTCBVS/05 Terravic Motion IR Database, namely, the algorithm has good robustness to some extent.

Keywords: Infrared image sequence · Spatio-temporal slice · Trajectory analysis · Pedestrian tracking

1 Introduction

Pedestrian tracking is a very important research hotspot in computer vision field and tracking pedestrian targets in infrared (IR) image sequences is especially a crucial part in many military or civil fields, such as video surveillance, position orientation and behavior analysis. However, the gray contrast between the targets and the background in IR image is quite low and the background in IR image is usually inevitably contaminated by unknown noises. Moreover, the distant surveillance determines that the tracked targets are usually small and blurred. Hence, developing an effective infrared pedestrian tracking algorithm is a great challenge.

Many researchers have focused on the tracking of infrared targets and numerous tracking algorithms have been proposed in this field, such as Meanshift [1], Camshift [2], histograms of oriented gradients [3] feature, SVM Classifier [4], local binary patterns [5], particle filter [6] and spatio-temporal slices analysis, etc. Among these

H. Zha et al. (Eds.): CCCV 2015, Part II, CCIS 547, pp. 148–158, 2015.
DOI: 10.1007/978-3-662-48570-5_15

methods, spatio-temporal slices analysis has gained special attention for its unique advantages. Compared to traditional methods, it uses long time scale of information rather than a small amount of short time frames in image sequence for data processing, which enriches the visual information and track information and improves the tracking accuracy of the algorithm. Ngos, etc. [7] uses structure tensor histogram to do a statistics of the visual information in spatio-temporal slices, and do several video segmentations in space domain to realize the detection of foreground and background. Sato, etc. [8] proposes a TSV (Temporal Spatio - Velocity) transform method to calculate the rates of pixels, and to extract target in video image by binary TSV in order to implement the target tracking and interactive behavior recognition. Jingjing Yang, etc. [9] presents a slice-based approach for pedestrian detection in still images. Limited numbers of horizontal spatio-temporal sub-regions are employed to represent pedestrians. Meanwhile, a classifier is constructed to classify multiple sub-regions.

The existing tracking methods based on spatio-temporal slice almost only consider horizontal slice analysis, and usually have complex associated algorithms. Besides, they generally need a large amount of calculation. Considering the above mentioned, in this paper, a novel tracking algorithm based on spatio-temporal slice trajectory (both in horizontal and vertical direction) analysis is proposed by analyzing the spatio-temporal slice trajectory superposed by multiple layers to obtain target boundary and position in order to realize reliable and accurate infrared pedestrian tracking.

The remainder of the paper is arranged as follows. A brief review on spatio-temporal slice is provided in Section 2. Section 3 describes the detailed principles of the proposed tracking algorithm. Experimental results and conclusions are given in Sections 4 and 5.

2 Spatio-temporal Slices

The concept of spatio-temparal slice was initially proposed in 1985 by E. H. Adelson and J. Bergen [10] in the article ‘Spatiotemporal energy models for the perception of motion ’. Assuming a video as a 3-D image sequence in which the three dimensions are respectively the x, y, and time t, if you do segmentation along the axis of t, then the cross section is rightly the so-called spatio-temparal slice.

Spatio-temporal slice images record the history of long time scales of video information. There are usually three kinds of spatio-temparal slice depending on the segmentation direction: vertical, horizontal and diagonal slice. Among these slices, vertical and horizontal slices are more commonly used. One dimension of the cross section is time t, another dimension is x or y. If the cross section parallels to the x axis, it is called a vertical slice; if the cross section parallels to the y axis, then it is called a horizontal slice. The illustration of horizontal and vertical spatial temporal slices is shown in Fig.1. The image sequence includes 160 frames (T=160) and the size of each frame is 240 × 360 (M=240, N=360) pixels.

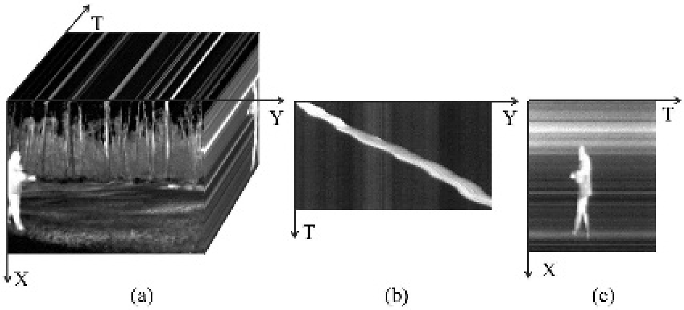

Fig. 1. Illustration of slicing: (a) image sequence (b) a horizontal slice (c) a vertical slice

3 Proposed Tracking Algorithm

In this paper, a novel tracking algorithm based on spatio-temporal slice trajectory analysis is proposed. First of all, multiple horizontal slices at different heights are obtained, then we respectively use background subtraction method to extract target trajectory in each slice. Next, do a superposition of multi-layer trajectories in order to get the complete horizontal trajectory. Meanwhile, the complete vertical target trajectory can be gained by using the similar process (instead, vertical slices are obtained from different widths).And then, vertical coordinate and height of moving target can be obtained via analyzing horizontal trajectory and abscissa along with width from vertical trajectory. Finally, tracking pedestrian with a rectangular box which is also called bounding box can be realized by using location information from the previous step. The process of this algorithm is depicted in Fig. 2.

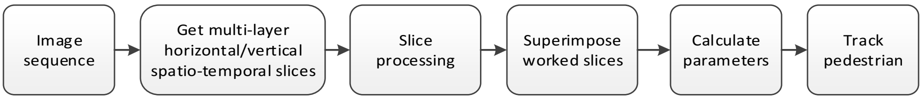

Fig. 2. Flow chart of the proposed tracking algorithm

3.1 Multi-layer Spatio-temporal Slices Acquisition

First of all, assume a surveillance video as an XYT 3-D image sequence, where x and y are the image dimensions and t is the temporal dimension. Select one row of all pixels continuously from each frame in the 3-D image sequence at the same height-level, and put them together by order of dimension t which forms a horizontal slice. It can be seen as an image with dimension y and the temporal dimension t. In the same way, select one column of all pixels from each frame in the 3-D image sequence at the same width-level we can gain a vertical slice.

Take a video sequence with T frames for example, if the size of each frame is $M \times N$, then the size of one horizontal slice is $M \times T$, and the size of one vertical slice is $N \times T$. Assume that I is a video sequence and Ii represents each frame (i ranges from 1 to T), if we respectively select one row at the height of m and one column at the width of n, then the horizontal slice I_H and vertical slice I_V can be represented as the following formulas.

$$\begin{cases} I_H\left(i,:\right) = I_i(\mathrm{m},:) \\ I_V\left(:,i\right) = I_i(:,\mathrm{n}) \end{cases} \tag{1}$$

In order to gain complete trajectories, different rows and columns should be selected at an abundant and appropriate amount in the proper region. The amount of slices will have influence on the performance of this algorithm. It will be discussed in Section 4.

3.2 Slice Processing

In order to extract trajectory from slice image, background subtraction method is selected. So background modeling is a key issue to this method. The complete process is shown in Fig.3.

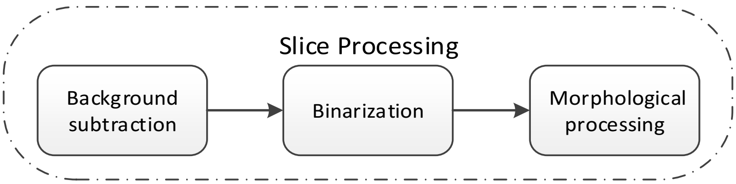

Fig. 3. Process of slice processing

There are many kinds of background modeling methods, such as Gaussian Mixture Model (GMM) [11], Codebook [12] Algorithm and so on. If one slice image is seen as a sequence of lines in the time domain, then according to the reference definition of foreground and background in video sequence, one line is equivalent to one frame in a sequence and the trajectory is equivalent to the target section. In this section, single Gaussian Model was used for multi-layer slices background modeling. Each layer of the spatio-temporal slice corresponds to a row background model. In the same layer, the y value of all pixels are fixed, but in different layers y value are different. Therefore, on the whole, layers of time slice is still a two-dimensional Gaussian background model, which can be represented as the following formula.

$$I_B\left(x,y\right) \sim N\left(\mu,\sigma^2\right) \tag{2}$$

Assume B is a pixel from background in one slice, and x is its abscissa y is its ordinate, μ is average value, σ is variance.

Take the initial m rows in a single layer slice, μ_0 and σ_0 are initial average value and variance, and it can be obtained as the formula below.

$$\begin{cases} \mu_0 = \dfrac{1}{m}\sum_{t=1}^{m} x_t \\ \sigma_0^2 = \dfrac{1}{m}\sum_{t=1}^{m}(x_t - \mu_0)^2 \end{cases} \tag{3}$$

Then Judge by the formula (5) of pixel matching, if meet, then the pixels should update according to formula (6), otherwise, remain the same.

$$|x_t - u_t| < 2.5\sigma_t \tag{4}$$

$$\begin{cases} \mu_t = (1-\alpha)\mu_{t-1} + \alpha X_t \\ \sigma_t^2 = (1-\alpha)\sigma_{t-1}^2 + \alpha(X_t - \mu_t)^T (X_t - \mu_t) \end{cases} \tag{5}$$

Among them, α is the update rate of the background. From this process of Gaussian modeling, we can gain the background B_H and B_V. Then each line in one slice image make subtraction with this model B_H and B_V, the new slice is almost the trajectory T_H and T_V.

$$\begin{cases} T_H(\mathrm{i},\mathrm{j}) = \left|I_H(\mathrm{i},\mathrm{j}) - B_H(\mathrm{i},\mathrm{j})\right| \\ T_V(\mathrm{i},\mathrm{j}) = \left|I_V(\mathrm{i},\mathrm{j}) - B_V(\mathrm{i},\mathrm{j})\right| \end{cases} \tag{6}$$

Then the trajectory should be converted to binary image at an appropriate threshold.

$$T_{H/V}(\mathrm{i},\mathrm{j}) = \begin{cases} 1 & T_{H/V}(\mathrm{i},\mathrm{j}) \geq Threshold \\ 0 & otherwise \end{cases} \tag{7}$$

At last, some morphological processing were done on the trajectories, such as median filtering, open operation and close operation in order to obtain clear multi-layer trajectories.

3.3 Trajectory Extraction

The appearance of trajectory has its certain uncertainty. However, when the target motion in the video, the different parts of the same target at any of the same moment have similar levels of coordinates or abscissa as well as a consistent motion mode or movement trend. Thus, in the trajectory superposition phase, the adjacent layer trajectory images with similar levels of coordinate or abscissa in the path do a superposition then the complete trajectory can be obtained. This process can eliminate the target area of the shade or loss caused by the trajectory of fracture to some extent.

$$\begin{cases} TH = \sum_{h=1}^{rows} \sum_{i,j=1}^{M,T} T_H(i,j) \\ TV = \sum_{v=1}^{columns} \sum_{i,j=1}^{N,T} T_V(i,j) \end{cases} \tag{8}$$

In this formula, rows is the number of horizontal slices and columns is the number of vertical slices.TH is the complete horizontal trajectory and TV is the complete vertical trajectory.

3.4 Parameter Calculation and Pedestrian Tracking

The complete trajectory can reflect the scene changes and target motion states, such as the abscissa, ordinate, instantaneous velocity, direction, width and height information of the target as above mentioned. First state, the trajectory image is a binary image. It means that if one pixel is trace of trajectory, then its value is 1, otherwise it is background and its value is 0.

For horizontal trajectory, it is analyzed it by row. In each row, count the pixels with the value of 1, and record the coordinate of first appearance position as Min_y and last as Max_y. Then w, the width of the bounding box is determined: that Max_y minus Min_y. As for vertical trajectory, it is analyzed by column. In each column, count the pixels with the value of 1, and record the abscissa of first appearance position as Min_x and last as Max_x. Then h, the height of the bounding box is determined: that Max_x minus Min_x. Meanwhile, the Min_x and Min_y can be used as the position information of the starting point along with this bounding box.

$$\begin{cases} w = Max_y - Min_y \\ h = Max_x - Min_x \end{cases} \tag{9}$$

With the information about the starting point and bounding box, rectangles can be drawn to track pedestrian in the corresponding frames of one video sequence. Take one frame at time t as example, w is the width of bounding box and h is the height of bounding box. Fig. 4 illustrates this tracking process.

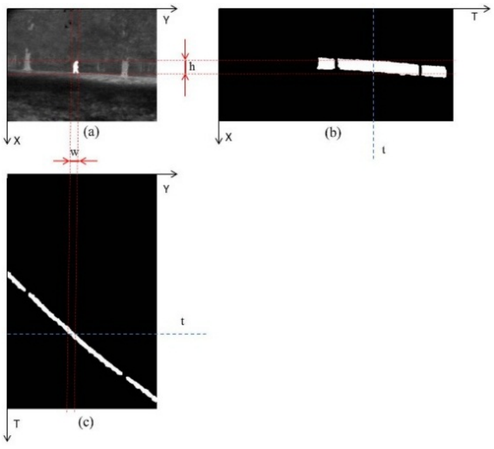

Fig. 4. (a): the t frame; (b): the horizontal trajectory; (c): the vertical trajectory

4 Experimental Results

In order to verify this proposed tracking algorithm based on the spatio-temporal slice trajectory analysis, this section we do simulation on multiple image sequences, including irw07, irw08, irw09, irw10 and irw11(these sequences are all from the OTCBVS/05 Terravic Motion IR Database [13]).

4.1 Complexity Performance

Traditional detection or tracking approaches usually consider all pixels in the image sequence, but in this approach target was detected by the low dimensional spatio-temporal slice images. Therefore, computational cost is relatively reduced in our approach. Their computational costs are shown respectively in formula (10) and formula (11).

$$O\left(M \times N \times T\right) \tag{10}$$

$$O\left[\left(k_1 \times N + M \times k_2\right) \times T\right] \tag{11}$$

where M, N, T represent the image height, width and total image frames, and k1,k2 represent the number of horizontal and vertical slice, meanwhile, $k_1 \ll M, k_2 \ll N$.

To validate time efficiency of our approach, experiments were carried on multiple surveillance videos of OTCBVS/05 Terravic Motion IR Database. The algorithm has been implemented by Matlab R2009a on a PC with Intel Core Quad CPU at 2.50 GHz running the Windows 7 Operating System. The size of each infrared image in this database is 320×240 pixels. Table 1 is the brief description about these image sequences. The average operating time of per frame is shown in Table 2.

Table 1. Sequence Information

Sequence Name	Description		
	Motion direction	Distance	Occlusion
irw07	away from the camera	close	no
irw08	across the scene	far	no
irw09	across the scene	far	no
irw10	across the scene	far	yes
irw11	across the scene	close	yes

4.2 Comparison of Tracking Performance

Just like the horizontal slices, the vertical slices also include motion information of the target. Especially the height and abscissa of the target can be obtained from the vertical trajectory. So it can improve tracking accuracy by analyzing both horizontal and vertical trajectories. The experimental results by only using horizontal slice and using both horizontal and vertical slices are shown in Fig. 5.

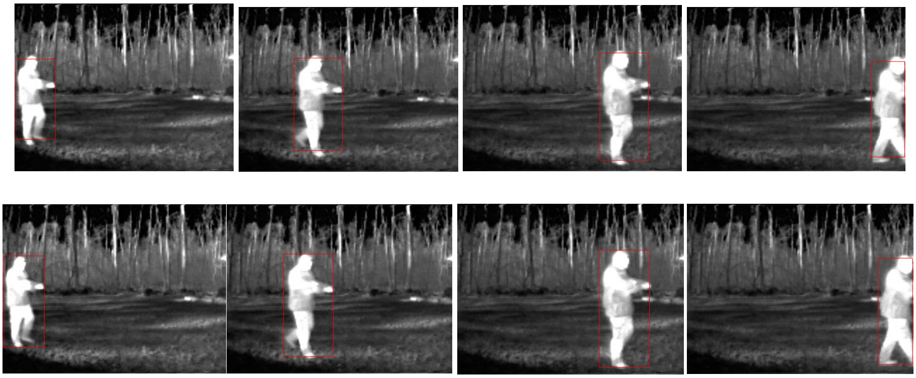

Fig. 5. Tracking results of irw08 (1) Row1: only use horizontal slices (2) Row2: the proposed method

From the comparison of these tracking results, we can see it can indeed track pedestrian with higher accuracy by using horizontal and vertical slices together.

4.3 Tracking Results

Sequences with different scenes and motion modes will have different trajectory patterns. The following images are different horizontal and vertical trajectories.

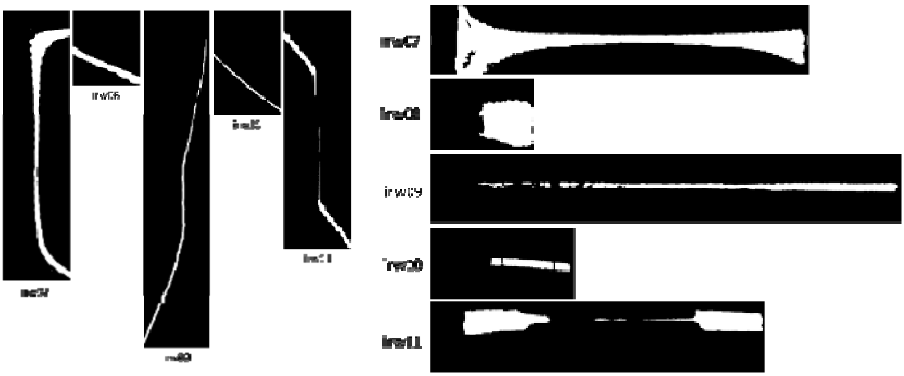

Fig. 6. Trajectories of all sequences (a) left: horizontal (b) right: vertical

The figures below are part of the tracking results of irw07, irw09, irw10 and irw11.

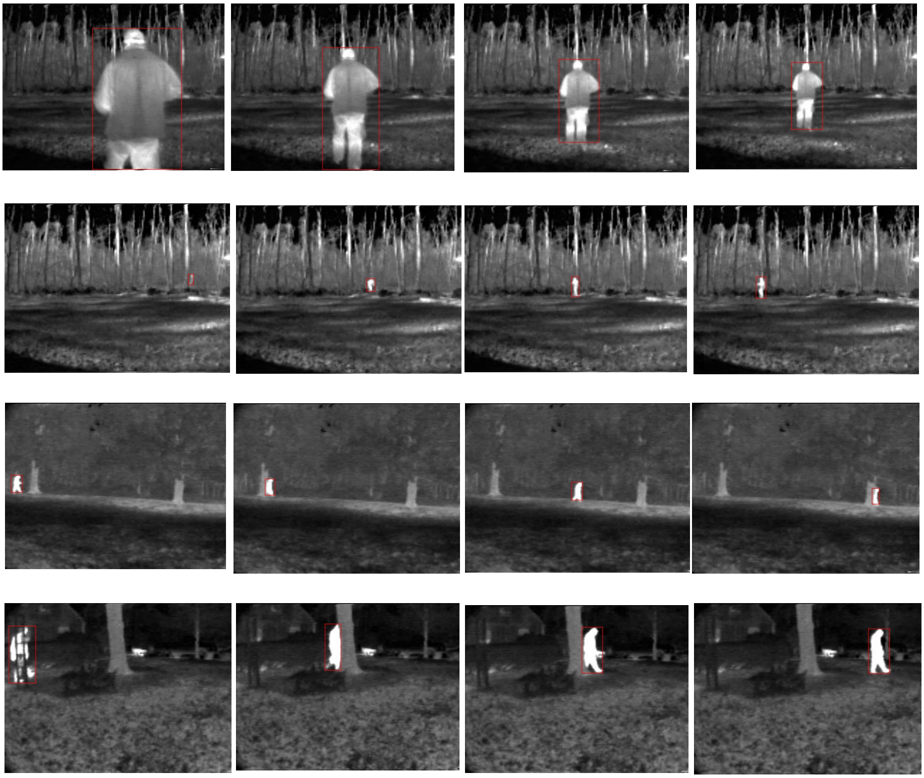

Fig. 7. Part of the tracking results from OTCBVS (1) Row1: irw07 (2) Row 2: irw09 (3)Row3:irw10 (4) Row4:irw11

If the foreground (pedestrian) is regarded as positive and background is negative, then we can use the following evaluation index to validate the efficiency of the proposed algorithm.

Table 2. Tracking results for the thermal video sequences

Name	Total Frames	TP	FP	FN	TN	TR	ACC	Comp.Time (s/frame)
irw07	1300	1206	0	0	94	1	1	0.05117
irw08	360	184	0	0	176	1	1	0.05071
irw09	1620	1354	0	48	218	0.9658	0.9704	0.05358
irw10	500	263	0	4	233	0.9850	0.9920	0.04972
irw11	1150	818	0	27	305	0.9680	0.9765	0.04898
Avg.	-	-	-	-	-	0.9838	0.9878	0.05083

(TP: True Positive; FP: False Positive; FN: False Negative; TN: True Negative; TR: Tracking Rate, TR = TP/ (TP + FN); ACC: ACC= (TP+TN)/ (P+N))

It can be observed from Table 2 that the proposed method has average 98.38% tracking rate at a relatively fast computation speed and zero false alarm rate. For sequence like irw09, irw10 and irw11, all the one pedestrian in motion are under the condition of target partially or completely occlusion for a period, so it was a little difficult to track the indetectable target and some false negative result (which means the target was regarded as background) occurred.

5 Conclusions

A novel tracking algorithm based on spatio-temporal slice trajectory analysis for single pedestrian tracking is proposed. Trajectories both in horizontal and vertical direction are used for determining the information of target boundary and position. Experiments show that the proposed tracking algorithm could perform well in different infrared image sequence, especially that it tolerates the situation that subregion of the pedestrian was with occlusion. In conclusion, it can provide relatively accurate pedestrian bounding boxes with low computational cost, and it has robustness to some extent.

References

1. Leichter, I., Lindenbaum, M., Rivlin, E.: Mean shift tracking with multiple reference color histograms. Computer Vision and Image Understanding **114**(3), 400–408 (2010)
2. Hsia, K.H., Lien, S.F., Su, J.P.: Moving target tracking based on CamShift approach and Kalman filter. Int. J. Appl. Math. Inf. Sci. **7**(1), 193–200 (2013)
3. Corvee, E., Bremond, F.: Body parts detection for people tracking using trees of histogram of oriented gradient descriptors. In: 2010 Seventh IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS), pp. 469–475. IEEE (2010)
4. Bai, Y., Tang, M.: Robust tracking via weakly supervised ranking svm. In: 2012 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1854–1861. IEEE (2012)
5. Ning, J., Zhang, L., Zhang, D., et al.: Robust object tracking using joint color-texture histogram. International Journal of Pattern Recognition and Artificial Intelligence **23**(07), 1245–1263 (2009)
6. Wang, X., Tang, Z.: Modified particle filter-based infrared pedestrian tracking. Infrared Physics & Technology **53**(4), 280–287 (2010)
7. Ngo, C.W., Pong, T.C., Zhang, H.J.: Motion analysis and segmentation through spatio-temporal slices processing. IEEE Transactions on Image Processing **12**(3), 341–355 (2003)
8. Sato, K., Aggarwal, J.K.: Temporal spatio-velocity transform and its application to tracking and interaction. Computer Vision and Image Understanding **96**(2), 100–128 (2004)
9. Yang, J., Su, X., Ma, P.: Fast pedestrian detection using slice-based motion analysis. In: 2010 First International Conference on Pervasive Computing Signal Processing and Applications (PCSPA), pp. 74–77. IEEE (2010)
10. Adelson, E.H., Bergen, J.R.: Spatiotemporal energy models for the perception of motion. JOSA A **2**(2), 284–299 (1985)

11. Zhang, R., Ge, P., Zhou, X., et al.: An method for vehicle-flow detection and tracking in real-time based on Gaussian mixture distribution. Advances in Mechanical Engineering (2013)
12. Guo, J.M., Hsia, C.H., Liu, Y.F., et al.: Fast background subtraction based on a multilayer codebook model for moving object detection. IEEE Transactions on Circuits and Systems for Video Technology **23**(10), 1809–1821 (2013)
13. IEEE OTCBVS WS Series Bench; Roland Miezianko, Terravic Research Infrared Database. http://www.vcipl.okstate.edu/otcbvs/bench/Data/05/download.html