Journal of Southeast University (English Edition) Vol. 24, No. 4, pp. 468 –472 Dec. 2008 ISSN 1003—7985

Multi-object tracking based on behaviour and partial observation

Lu Hong[1,2] Fei Shumin[1] Zheng Jianyong[3] Zhang Tao[1]

([1] School of Automation, Southeast University, Nanjing 210096, China)
([2] School of Automation, Nanjing Institute of Technology, Nanjing 211167, China)
([3] School of Electrical Engineering, Southeast University, Nanjing 210096, China)

Abstract: To cope with multi-object tracking under real-world complex situations, a new video-based method is proposed. In the detecting step, the moving objects are segmented with the third level DWT (discrete wavelet transform) and background difference. In the tracking step, the Kalman filter and scale parameter are used first to estimate the object position and bounding box. Then, the center-association-based projection ratio and region-association-based occlusion ratio are defined and combined to judge object behaviours. Finally, the tracking scheme and Kalman parameters are adaptively adjusted according to object behaviour. Under occlusion, partial observability is utilized to obtain the object measurements and optimum box dimensions. This method is robust in tracking mobile objects under such situations as occlusion, new appearing and stablization, etc. Experimental results show that the proposed method is efficient.
Key words: multi-object tracking; projection ratio; occlusion ratio; behaviour; partial observation; Kalman filter

The researches for detecting and tracking multi-object have received great attention. Although many tracking algorithms have been proposed in the literature, the problem of multiple objects tracking in a real complex scene is still far from being completely solved. There are two main difficulties. The one is occlusion. The other is the change in scale, shape and illumination. State estimation and data association are main facets of multi-object tracking and identification. For state estimation, the Kalman filter is a commonly used method, which can reduce the research space in data association and is more efficient than those algorithms which match objects throughout whole images[1]. But the Kalman filter cannot be applied to solve occlusions[2], because the object measurement cannot be accurately obtained in this case. Refs. [3 –9] proposed different algorithms to deal with this problem, but they only cope with one kind of occlusion.

The contribution of our work is: first, the observability of bounding edges and fragments is used to obtain object measurement under occlusion, which makes the track more robust than those measurements roughly deduced. Secondly, the projection ratio and the occlusion ratio are defined based on a scene model and combined to judge object behaviours. Finally, the tracking scheme and Kalman parameters are adaptively adjusted according to object behaviours. Experimental results demonstrate that this method is effective.

Received 2008-03-28.

Biographies: Lu Hong(1973—), female, graduate; Fei Shumin(corresponding author), male, doctor, professor, smfei@ seu. edu. cn.

Foundation item: The National Natural Science Foundation of China(No. 60574006, 60804017).

Citation: Lu Hong, Fei Shumin, Zheng Jianyong, et al. Multi-object tracking based on behaviour and partial observation[J]. Journal of Southeast University (English Edition), 2008, 24(4): 468 –472.

1 Proposed Methods

1.1 Motion detection and representation

1.1.1 Motion detection

Here, the traditional background difference and the three-level DWT are adopted. Since the low-resolution sub-images bear analogy to the original images and reduce high frequency noise in varied backgrounds[1], here, the third-level wavelet sub-images of background and image sequences are applied to background difference so as to detect mobile objects with decreased computing costs and image noises[10].

1.1.2 Motion representation

1) Bounding box parameters

A 2-dimensional bounding box is used to represent the mobile object. The center coordinates $(x^i(t), y^i(t))$, length $L^i(t)$ and width $W^i(t)$ are the main parameters of the i-th object bounding box in frame t.

2) Velocity

Center velocity $(v_x^i(t), v_y^i(t))$ is used to represent the motion features of the object. The horizontal velocity $v_x^i(t)$ is computed as

$$v_x^i(t) = \lambda v_x^i(t-1) + (1-\lambda)\frac{x^i(t) - x^i(t-1)}{\Delta t} \qquad v_x^i(1) = 0 \tag{1}$$

where λ is the weight, and Δt is the time interval between frames.

3) Scale parameters $S(t)$ [10]

$S(t)$ is used to estimate and rectify the box change in dimension caused by the motion component along the optical axes away from or towards the camera.

1.2 State estimation and behaviour reasoning

The flowchart of the proposed tracking algorithm is shown in Fig. 1. The bounding box of a moving object in the next frame is first estimated with the Kalman filter and scale parameter $S(t)$. Then the object's behaviour is judged with the association process. Finally, the track scheme and the Kalman gain are adaptively adjusted to obtain object measurement and optimum box dimensions under different behaviours.

1.2.1 State estimation with Kalman filter

A Kalman filter based on a first-order motion model is used to estimate object states in the next frame. Since our system mainly aims to monitor vehicles, each target is assumed to be along a linear trajectory with constant velocity and constant size. The state vector used is $\boldsymbol{X} = \{x, y, v_x, v_y\}^{\mathrm{T}}$ and measurement vector is $\boldsymbol{Z} = \{x, y\}^{\mathrm{T}}$.

1.2.2 Defining object's behaviour

Because the camera is fixed, an object's behaviour can be

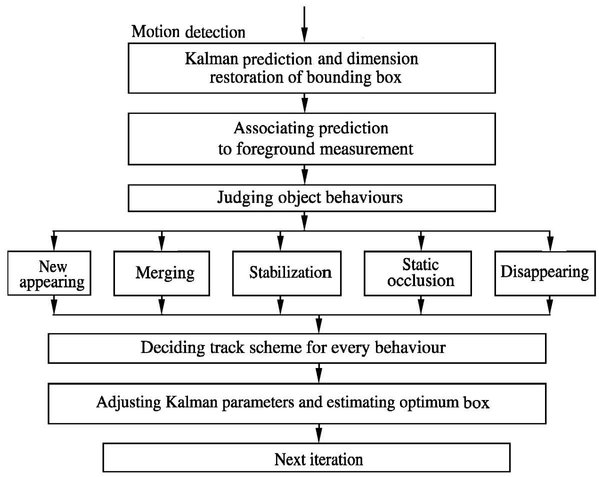


Fig. 1 The flowchart of the tracking algorithm

defined for a specific camera position. Five types of basic behaviours in a public transport environment are defined as follows:

1) New appearing. It represents that objects enter into the field-of-view.

2) Merging. Where more than one dynamic object group together because of being too near. There are three types of merging, i. e., PDO (partially dynamic occlusion), TDO(totally dynamic occlusion) and grouping(before or after PDO).

3) Static occlusion. It represents that objects are occluded by static background. There are two types of static occlusion, i. e., PSO(partially static occlusion) and TSO(totally static occlusion).

4) Disappearing. There are two types of disappearing: One is outside the limit of the field-of-view; the other is outside the time limit of TSO.

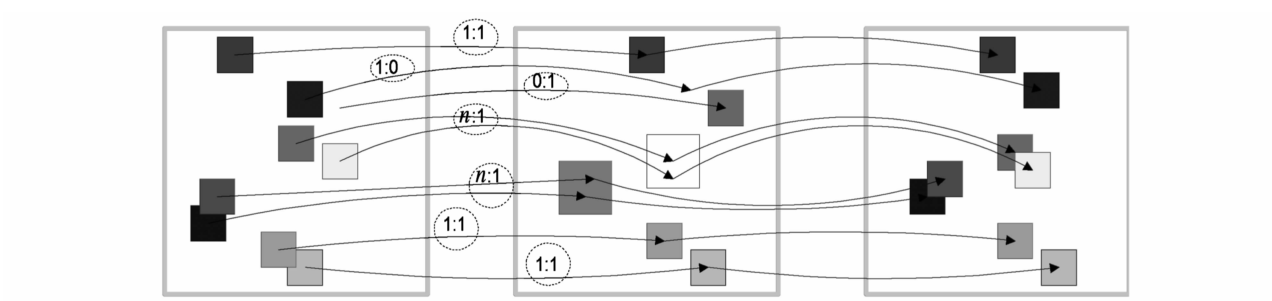
5) Stabilization. It represents that objects move individually without any of the above-mentioned behaviours.

The complex behaviour of objects can be viewed as the synthesization of the five basic behaviours.

1.2.3 Reasoning and judging object behaviour

1) Center-based association and projection ratio

To associate the same object in adjacent frames, we first estimate the object center in the next frame with the Kalman filter, then project this center toward detected bounding boxes, and finally label the projection ratio. In Fig. 2, the same intensity box denotes the same object. The projection ratio value R_P on every arrow represents the quantitative relationship between Figs. 2(a) and (b). In Fig. 2, 0, 1 and n are the numbers of associated objects. $R_P = 1:1$ denotes stablization or PSO, $R_P = 1:0$ denotes PSO, TSO or disappearing, $R_P = 0:1$ denotes new appearing, and $R_P = n:1$ denotes merging or continuing merging.



(a) (b) (c)

Fig. 2 Judging object behaviours with R_P. (a) Tracked boxes in frame $t-1$; (b) Detected boxes in frame t; (c) Expectantly tracked boxes in frame t

Under PSO, the un-occluded fractional object can be detected. If the projecting process succeeds (The estimated center hits a detected box), $R_P = 1:1$, which will confuse stablization and PSO. If the projecting process fails, $R_P = 1:0$, which will confuse PSO and TSO. So it is not adequate only using center-based association to classify object behaviours.

2) Region-based association and occlusion ratio

To avoid confusion, occlusion ratio $R_O(t)$ is proposed for further classifying object behaviours. $R_O(t)$ is the ratio of $A_{dete}(t)$ to $A_{esti}(t)$, and computed by Eqs. (2) and (3). Where $A_{esti}(t)$ is the area of the estimated box with $S(t)$, $A_{dete}(t)$ is the area of the detected one, and $A_O(t)$ is the overlapped region of the estimated box and the detected one. To n merged objects, $A_{esti}(t)$ is the area of a minimum box enclosing n estimated boxes. β and γ are given threshold values. $R_O(t) = 1$ represents TSO and $R_O(t) = 0$ represents no static occlusion.

$$R_O(t) = \begin{cases} 0 & \text{if } 1 - \dfrac{A_{dete}(t)}{A_{esti}(t)} \leqslant \beta \\ 1 - \dfrac{A_{dete}(t)}{A_{esti}(t)} & \text{if } 1 - \dfrac{A_{dete}(t)}{A_{esti}(t)} > \beta \end{cases} \tag{2}$$

$$A_{dete}(t) = \begin{cases} 0 & \text{if } A_O(t) < \gamma A_{dete}(t) \\ A_{dete}(t) & \text{if } \gamma A_{dete}(t) \leqslant A_O(t) \leqslant A_{dete}(t) \end{cases} \tag{3}$$

3) Judging object behaviour

According to the projection ratio and the occlusion ratio, the object behaviours can be judged as follows (see Tab. 1). Under $R_P = 1:1$, if $R_O = 0$, the object is in stabilization, or else PSO. Under $R_P = 1:0$ and $R_O = 1$, the object is in TSO. If TSO continues and goes beyond the time limit, the object disappears. In addition, if the estimated object center is out of the field-of-view, the object is directly judged as disappearing. Under $R_P = n:1$ and $0 < R_O < 1$, the object is in merging and PSO synchronously (complex behaviour).

Tab. 1 Object behaviours judgement with projection ratio and occlusion ratio

Projection ratio	Occlusion ratio		
	0	(0, 1)	1
0:1	New appearing	—	—
1:0	—	PSO	TSO or disappearing
1:1	Stabilization	PSO	—
n:1	Merging	* Merging and PSO	—

Notes: — denotes non-existence; * denotes complex behaviour.

1.3 Tracking algorithm

Object behaviour is changeable. In special situations (e.g. static occlusion), the object measurement $\boldsymbol{Z}(t)$ is partly or completely unavailable. To adapt to these, our tracking scheme is adjusted adaptively.

1.3.1 Initialization and quitting track

Initialization: if an object first appears in the field-of-view, we record its position but do not track it in the current frame. Once this object appears in next frame, we initialize its velocity and scale parameter and update the projection ratio with 1:1.

Quitting track: If an object is judged to be disappearing, the track quits.

1.3.2 Stabilization tracking

The center and dimensions of the detected box are regarded as $\boldsymbol{Z}(t)$ and the optimum dimensions of the tracked box, respectively. $\boldsymbol{R}(t)$ and $\boldsymbol{Q}(t-1)$ are initialized with constants.

1.3.3 Occlusion tracking

The occlusion tracking steps are as follows:

1) Obtaining prior estimating center by the Kalman filter;

2) Estimating box dimensions by Eq. (4);

3) Adjusting the estimated box to obtain $\boldsymbol{Z}(t)$ and $(L_{\mathrm{opti}}(t), W_{\mathrm{opti}}(t))$ according to partial observability;

4) Adjusting the Kalman parameters to obtain the posterior estimate center.

$$\begin{aligned} L_{\mathrm{esti}}(t) &= S(t) L_{\mathrm{opti}}(t-1) \\ W_{\mathrm{esti}}(t) &= S(t) W_{\mathrm{opti}}(t-1) \end{aligned} \tag{4}$$

Assume that $(x_{\mathrm{esti}}, y_{\mathrm{esti}})$ is a prior estimate center, and $(x_{\mathrm{adj\_esti}}, y_{\mathrm{adj\_esti}})$ is an adjusted one. $(x_{\mathrm{min\_esti}}, y_{\mathrm{min\_esti}})$ and $(x_{\mathrm{max\_esti}}, y_{\mathrm{max\_esti}})$ are the corner positions at the upper left and the lower right of the estimated box, respectively. $(x_{\mathrm{min\_dete}}, y_{\mathrm{min\_dete}})$ and $(x_{\mathrm{max\_dete}}, y_{\mathrm{max\_dete}})$ are the corner positions at the upper left and the lower right of the detected box.

The adjustments of the estimated box and the Kalman gain are different for every type of occlusion.

1) PSO tracking

The detected partial region usually overlaps with the estimated bounding box, so we can use the former to rectify the latter. $\boldsymbol{R}(t)$ is set as $R_{\mathrm{O}}(t)$ and $\boldsymbol{Q}(t-1)$ as $1-R_{\mathrm{O}}(t)$. The track scheme is as follows:

If $\dfrac{A_{\mathrm{O}}(t)}{A_{\mathrm{dete}}(t)} = 1$, then $\boldsymbol{Z}(t) = \{x_{\mathrm{esti}}(t), y_{\mathrm{esti}}(t)\}^{\mathrm{T}}$ and $\{L_{\mathrm{opti}}(t), W_{\mathrm{opti}}(t)\} = \{L_{\mathrm{esti}}(t), W_{\mathrm{esti}}(t)\}$.

If $\dfrac{A_{\mathrm{O}}(t)}{A_{\mathrm{dete}}(t)} \in [\gamma, 1)$, then $\boldsymbol{Z}(t) = \{x_{\mathrm{adj\_esti}}(t), y_{\mathrm{adj\_esti}}(t)\}^{\mathrm{T}}$ and $\{L_{\mathrm{opti}}(t), W_{\mathrm{opti}}(t)\} = \{L_{\mathrm{adj\_esti}}(t), W_{\mathrm{adj\_esti}}(t)\}$.

$x_{\mathrm{adj\_esti}}(t)$ and $L_{\mathrm{adj\_esti}}(t)$ can be obtained by moving the estimated bounding box to enclose the detected box. The moving rules are as follows:

$$x_{\mathrm{adj\_esti}}(t) = \begin{cases} x_{\mathrm{esti}}(t) - \Delta x_{\min} & \text{if } |\Delta x_{\min}| < |\Delta x_{\max}| \text{ and } \Delta x_{\min}\Delta x_{\max} > 0 \\ x_{\mathrm{esti}}(t) - \Delta x_{\max} & \text{if } |\Delta x_{\min}| > |\Delta x_{\max}| \text{ and } \Delta x_{\min}\Delta x_{\max} > 0 \\ 0.5(x_{\mathrm{min\_dete}}(t) + x_{\mathrm{max\_dete}}(t)) & \text{if } \Delta x_{\min}\Delta x_{\max} < 0 \end{cases} \tag{5}$$

$$\begin{aligned} L_{\mathrm{adj\_esti}}(t) = \max\{ &|x_{\mathrm{max\_esti}}(t) - x_{\mathrm{min\_esti}}(t)|, \\ &|x_{\mathrm{max\_dete}}(t) - x_{\mathrm{min\_dete}}(t)|\} \end{aligned} \tag{6}$$

where $\Delta x_{\min} = x_{\mathrm{min\_esti}}(t) - x_{\mathrm{min\_dete}}(t)$ and $\Delta x_{\max} = x_{\mathrm{max\_esti}}(t) - x_{\mathrm{max\_dete}}(t)$.

For an object being always in PSO, our algorithm can track it according to the partially observable region.

PSO tracking can also be utilized to track the complex behaviour mixed by merging and PSO.

2) TSO tracking

$\boldsymbol{R}(t)$ and $\boldsymbol{Q}(t-1)$ are set as infinity and zero, respectively; thus, the Kalman gain $\boldsymbol{K}_{\mathrm{g}}(t)$ is a zero value. That is to say, if an object is totally occluded, the system will trust the predicted results completely; i. e., $\boldsymbol{Z}(t) = \{x_{\mathrm{esti}}(t), y_{\mathrm{esti}}(t)\}^{\mathrm{T}}$ and $\{L_{\mathrm{opti}}(t), W_{\mathrm{opti}}(t)\} = \{L_{\mathrm{esti}}(t), W_{\mathrm{esti}}(t)\}$.

3) PDO tracking (or group tracking)

$\boldsymbol{R}(t)$ and $\boldsymbol{Q}(t-1)$ are initialized with constants. The adjustment and tracking rule[10] is as follows:

Step 1 Judging observable bounding edge and adjusting box.

Step 2 Judging and adjusting unobservable bounding edge.

Step 3 The adjusted box center is regarded as object measurement $\boldsymbol{Z}(t)$ to feed back to the Kalman filter. The dimensions of the adjusted bounding box are the optimum ones.

4) TDO tracking

Under TDO, one estimated box is embedded in another box, so the track scheme of the former is the same as TSO. But the latter box need be adjusted to fit in the detected box, and $\boldsymbol{R}(t)$ and $\boldsymbol{Q}(t-1)$ are initialized with constants.

2 Experimental Results and Discussion

Testing sequence 1 (see Fig. 3) is the Nibelungen-Platz sequence. Testing sequence 2 (see Figs. 4 and 5) is from video captured by us. These videos simulate several cases of condition for moving objects, such as new appearing, merging and static occlusion etc. For both sequences, we use the following parameters: the state transition matrix

$$\boldsymbol{A} = \begin{bmatrix} 1 & 0 & \Delta t & 0 \\ 0 & 1 & 0 & \Delta t \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

and the measurement matrix

$$\boldsymbol{H} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix}$$

When $\boldsymbol{R}(t)$ and $\boldsymbol{Q}(t-1)$ are initialized, $\boldsymbol{Q} = 0.1\boldsymbol{I}$ ($\boldsymbol{I}$ is a 4×4 identity matrix) and $\boldsymbol{R} = 0.1\boldsymbol{I}$ ($\boldsymbol{I}$ is a 2×2 identity matrix). $\beta = 0.1$, and $\gamma = 0.6$. In Figs. 3 to 5, the gray dashed box represents foreground detection. We process testing sequences at a sampled rate of 30 frame/s.

In experiment 1, the proposed method is mainly evaluated by testing the case of new appearing, TSO, PSO and stablization. We process sequence 1 at an image size of 268×201. Figs. 3(a) to (d) are the detected results of frames 315, 331, 335, and 343, and Figs. 3(e) to (h) are corresponding tracked results with the proposed method, where two moving cars are moving across a clumpy road. The object tracked with the black box undergoes stablization(see Fig. 3(e)) , TSO(see Fig. 3(f)) and PSO(see Fig. 3(g)) and comes back to stablization (see Fig. 3 (h)) . The object, tracked with the white box, appears in Fig. 3 (e) , and then maintains stablization(see Figs. 3 (f) to (h)) . The tracked results show that our method can accurately track the objects in these cases.

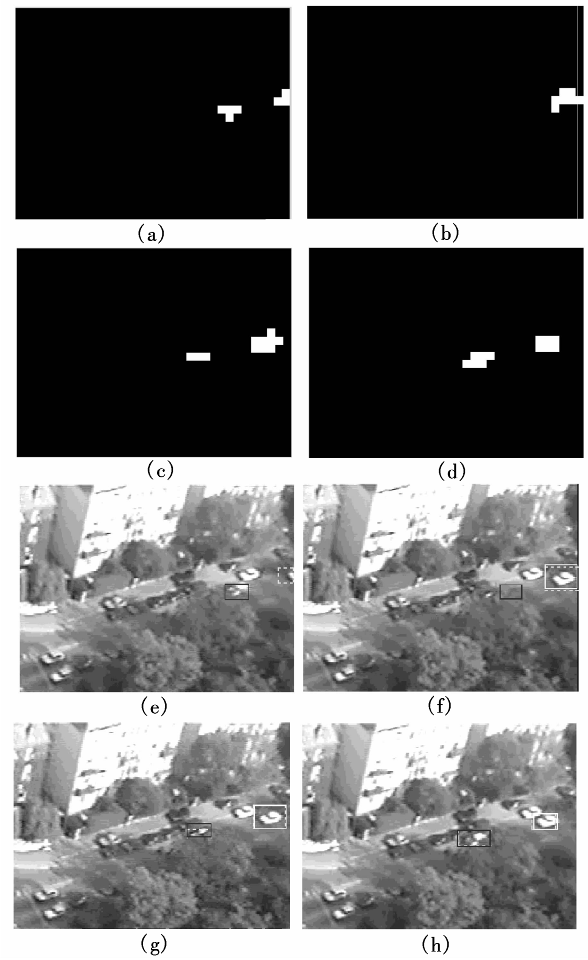


Fig. 3 Region detection and the tracked results of experiment 1 with proposed method

In experiment 2, merging(PDO, TDO and grouping) and unmerging are mainly tested. Blind tracking under the typical Kalman filter is used for comparison with our method, in which uniform scale parameters and velocity rectification are adopted. We process frames 4302 to 4541 of sequence 2 at an image size of 320 × 240. Figs. 4 (a) to (d) are the tracked results of frames 4482, 4488, 4492 and 4502 with blind tracking, and Figs. 5(a) to(d) are ones with the proposed method. The car tracked with black box groups with the one tracked with the white box in frame 4482, then is partially occluded by the latter(see frame 4488) , and is in TDO from frame 4492. Two cars depart from each other in frame 4502. The motorcycle tracked with the gray solid box keeps stablization. Since the position and velocity of occluded objects cannot be rectified in time under blind tracking, the tracked box is incorrect(see Figs. 4(b) to (d)) . In Fig. 4(d) , the object formerly tracked with the black box is now tracked with the white one and no object is now tracked with the black one. These errors do not occur in Fig. 5. Obviously, our method can effectually track objects under these cases.

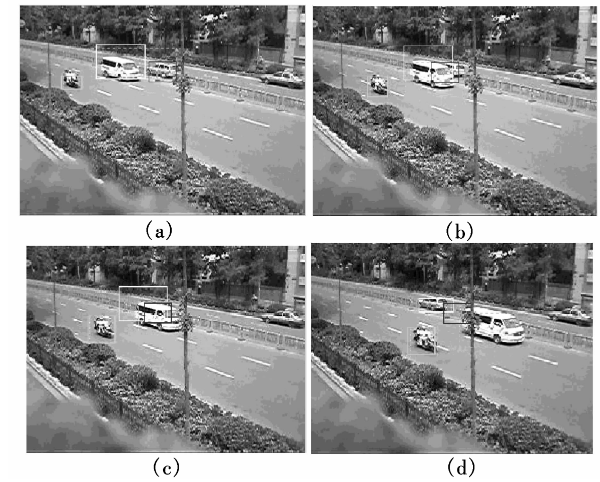


Fig. 4 The tracked results of experiment 2 with blind tracking

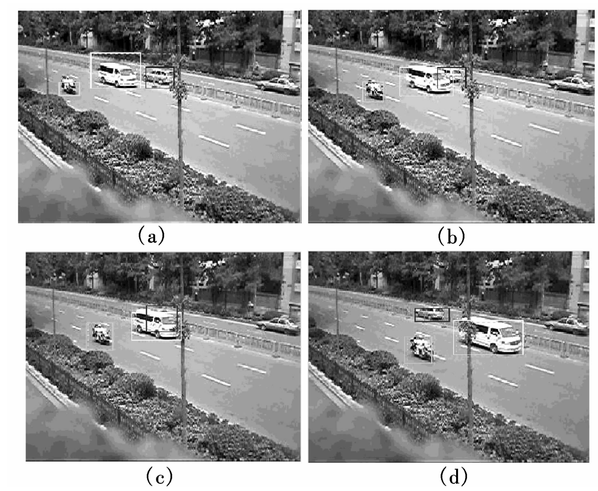


Fig. 5 The tracked results of experiment 2 with proposed method

In experiment 3, frames 895 to 997 of sequence 2 are processed, and merging among foreground objects and dynamic background is mainly tested.

Tab. 2 shows the tracking ratios. In experiments 1 and 2, our tracking ratios are up to 100% . In experiment 3, our tracking ratio is 88. 7% . The erroneous tracking in experiment 3 is mainly from the deep disturbance in the background, which leads to imperfect foreground detection in several frames. But this error can automatically disappear in the case that the occlusion is over.

Tab. 2 The results of objects tracking

Experiment	Frames	Moving objects	Correctly tracking	Correction ratio/%	Error ratio/%
1	347	933	933	100	0
2	240	720	720	100	0
3	103	309	303	88. 7	11. 3

3 Conclusion

In this paper, an adaptive method is proposed to track multiple moving objects in transportation. The projection ratio and occlusion ratio (based on center and region association, respectively) are used to judge object behaviours so as to adjust tracking schemes and the error covariances of the Kalman filter adaptively. For occlusion tracking, partial observabililty and adjustment of the bounding box are utilized. The proposed method can track the moving object in some kinds of real-world situations such as objects merging, static occlusion, stablization, disappearing and new appearing. The new algorithm has advantages over traditional blind tracking and group tracking schemes and has a high tracking ratio. It is noted that the correct identification rate highly depends on the performance of motion detection. This problem can be improved by better background reconstruction and other improved detection methods in the future.

References

[1] Cheng F H, Chen Y L. Real time multiple objects tracking and identification based on discretewavelet transform [J]. *Pattern Recognition*, 2006, **39**(6): 1126 − 1139.

[2] Jang D S, Choi H I. Active models for tracking moving objects [J]. *Pattern Recognition*, 2000, **33**(7): 1135 − 1146.

[3] Letang J M, Rebuffel V, Bouthemy P. Motion detection robust to perturbations: a statistical regularization and temporal integration framework[C]//*Proceedings of the International Conference on Computer Vision*. Berlin, Germany, 1993: 21 − 30.

[4] Weng S K, Kuo C M, Tu S K. Video object tracking using adaptive Kalman filter [J]. *Journal of Visual Communication and Image Representation*, 2006, **17**(6): 1190 − 1208.

[5] Xu M, Ellis T. Tracking occluded objects using partial observation [J]. *Acta Automatica Sinica*, 2003, **29**(3): 370 − 380.

[6] Ge Jiaqi, Li Bo, Chen Qimei. A region-based vehicle tracking algorithm under occlusion [J]. *Journal of Nanjing University*: *Natural Sciences*, 2007, **43**(1): 66 − 72. (in Chinese)

[7] Chang Faliang, Liu Xue, Wang Huajie. Target tracking algorithm based on meanshift and Kalman filter [J]. *Computer Engineering and Applications*, 2007, **43**(12): 50 − 52. (in Chinese)

[8] Sun Zhongsen, Sun Junxi, Song Jianzhong, et. al. Anti-occlusion arithmetic for moving object tracking [J]. *Optics and Precision Engineering*, 2007, **15**(2): 267 − 271. (in Chinese)

[9] Chang Faliang, Ma Li, Qiao Yizheng. Human oriented multi-target tracking algorithm in video sequence [J]. *Control and Decision*, 2007, **22**(4): 418 − 422. (in Chinese)

[10] Lu Hong, Fei Shumin, Zheng Jianyong, et al. An occlusion tolerant method for multi-object tracking [C]//*Proceedings of the 7th World Congress on Intelligent Control and Automation*. Chongqing, China, 2008: 5105 − 5110.

基于行为和部分观测的多目标跟踪

路　红[1,2]　费树岷[1]　郑建勇[3]　张　涛[1]

([1]东南大学自动化学院,南京 210096)
([2]南京工程学院自动化学院,南京 211167)
([3]东南大学电气工程学院,南京 210096)

摘要:针对复杂环境下的多目标跟踪问题提出了一种新的跟踪方法. 在检测部分,采用第三级离散小波变换和背景差分进行了目标分割. 在跟踪部分,首先利用卡尔曼滤波器和缩放因子估计目标在下一帧中的中心位置和矩形框尺寸,然后在中心关联和区域关联的基础上提出了投射率和遮挡率的概念,并结合投射率和遮挡率推断目标行为. 最后针对具体目标行为,自适应地调整跟踪方案和卡尔曼参数实现了多目标跟踪. 在遮挡情况下利用部分观测调整估计框以获得目标测量值和最优框尺寸. 提出的方法对处于遮挡、新出现以及稳定等情况下的运动目标均具有鲁棒的跟踪性能. 实验结果表明提出的方法是有效的.

关键词:多目标跟踪;投射率;遮挡率;行为;部分观测;卡尔曼滤波

中图分类号:TP391. 41