Journal of Southeast University (English Edition) Vol. 31, No. 2, pp. 266 - 271 June 2015 ISSN 1003—7985

# Video-based vehicle tracking considering occlusion

Zhu Zhou[1,3] Lu Xiaobo[2,3]

([1]School of Transportation, Southeast University, Nanjing 210096, China)
([2]School of Automation, Southeast University, Nanjing 210096, China)
([3]Key Laboratory of Measurement and Control of Complex Systems of Engineering of Ministry of Education, Southeast University, Nanjing 210096, China)

**Abstract:** To track the vehicles under occlusion, a vehicle tracking algorithm based on blocks is proposed. The target vehicle is divided into several blocks of uniform size, in which the edge block can overlap its neighboring blocks. All the blocks' motion vectors are estimated, and the noise motion vectors are detected and adjusted to decrease the error of motion vector estimation. Then, by moving the blocks based on the adjusted motion vectors, the vehicle is tracked. Aiming at the occlusion between vehicles, a Markov random field is established to describe the relationship between the blocks in the blocked regions. The neighborhood of blocks is defined using the Euclidean distance. An energy function is defined based on the blocks' histograms and optimized by the simulated annealing algorithm to segment the occlusion region. Experimental results demonstrate that the proposed algorithm can track vehicles under occlusion accurately.

**Key words:** vehicle tracking; occlusion processing; motion vector; Markov random field

**doi:** 10. 3969/j. issn. 1003 - 7985. 2015. 02. 019

With the development of the economy, more cameras have been installed on roads to monitor the traffic. How to make the most out of videos to gather traffic information is a problem. In theory, video-based vehicle tracking technology can automatically collect traffic information such as traffic volume, speed, density and traffic incidents[1], so it has received much attention in recent years.

The main challenge of video-based vehicle tracking is to track vehicles effectively in complicated environments such as the occlusion phenomenon between vehicles, during changing illumination and so on[2-4]. There are several object tracking algorithms considering the occlusion. Harguess et al.[5-6] tracked objects under occlusion using binocular video, but these binocular vision methods are unsuitable for practical applications. In the field of monocular vision, Zhang et al.[7] segmented the occlusion region in a geometrical way and when the outline of vehicles is unbroken, it works well. In addition, the weighted histogram is integrated with a mean shift[8-11] or a particle filter[12-13] to track objects under occlusion. In general, they can effectively track the objects whose size changes little, but in the traffic scene, the vehicle's scale may undergo a large change. It is difficult for these algorithms to update the size of the tracking window adaptively. In Ref. [14], Kamijo et al. presented a block-based vehicle tracking method and used the Markov random field to segment the occlusion region. This method can process occlusion and adapt to the change of vehicle's size simultaneously. However, it also has limitations. First, when a vehicle is close to but not occluded by another vehicle, it may be considered to be mistakenly occluded by the latter in tracking. Secondly, it fails to segment the vehicles whose motion vectors are similar. Furthermore, the two negative factors may appear simultaneously to make the tracking unsuccessful.

To deal with the above problems, a novel vehicle tracking algorithm is proposed. First, the target vehicle is divided in a different way to reduce the influence of its neighboring vehicle. Then, the block's histogram is used to define the Markov random field's energy function and make the occlusion segmentation insensitive to the vehicles' motion vectors. Experimental results demonstrate the effectiveness of the proposed method.

## 1 Regular Block-Based Vehicle Tracking Method

In Kamijo's vehicle tracking method, the image is divided into regular blocks, each of which consists of $8 \times 8$ pixels and has a single label. All of the blocks' labels constitute a label map, and vehicle tracking is equivalent to updating the label map continually.

$L_n$ is the label map of the $n$-th image. When the $(n+1)$-th image comes, the motion vectors of the vehicle's blocks are estimated. In them, the most frequent one is considered as the vehicle's motion vector, and based on it, the vehicle's blocks in $L_n$ are moved to new places. Then, the moved blocks are expanded as candidate blocks whose labels will be checked next. If the difference between intensities at a candidate block and those at the background image is greater than the threshold value, the block's label is set to be the vehicle's label, or else to be

**Received** 2014-10-18.

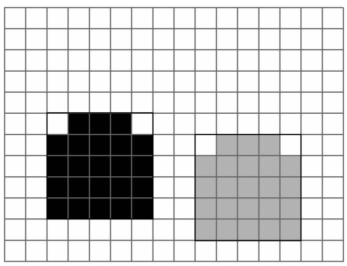
**Biographies:** Zhu Zhou(1984—), male, graduate; Lu Xiaobo(corresponding author), male, doctor, professor, xblu@seu.edu.cn.

**Foundation item:** The National Natural Science Foundation of China (No. 60972001, 61374194).

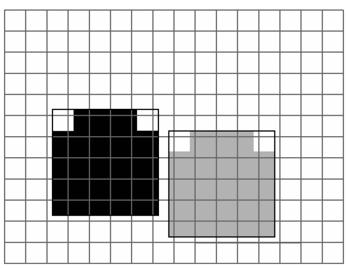
**Citation:** . Zhu Zhou, Lu Xiaobo. Video-based vehicle tracking considering occlusion[J]. Journal of Southeast University (English Edition), 2015, 31(2): 266 - 271. [doi: 10. 3969/j. issn. 1003 - 7985. 2015. 02. 019]

0. After checking the candidate blocks, a new label map $L_{n+1}$ can be obtained.

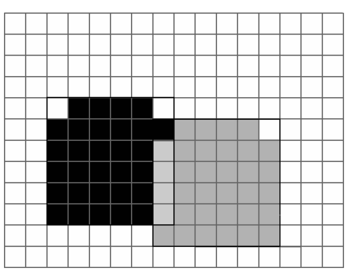
In general, the aforementioned algorithm works well. An exception is that when the moved blocks of a vehicle and that of its neighboring vehicle are less than $16\sqrt{2}$ pixels, the candidate blocks of one vehicle may overlap with that of another vehicle and a redundant occlusion may be created (As shown in Fig. 1(c), the gray blocks constitute an occlusion region). It is the first shortcoming of the algorithm in Ref. [14]. The second one will be analyzed as below.



(a)



(b)

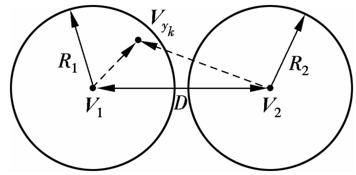


(c)

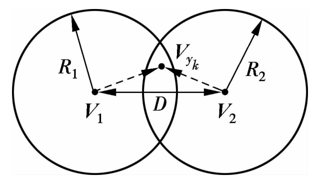
**Fig. 1** The redundant occlusion. (a) The blocks in $L_k$; (b) The moved blocks; (c) The candidate blocks

When a vehicle occludes another vehicle, the labels of the blocks in the occlusion region needs to be optimized. Each block's label is a discrete random variable and the combination of all block's labels is assumed to be a Markov random field. To block $y_k$ in the occlusion region, its label is optimized by minimizing the energy function $U(y_k)$ defined in Ref. [14]. If $U(O_1) \leqslant U(O_2)$, then block $y_k$ is considered to belong to vehicle $O_1$ or else belongs to vehicle $O_2$.

$\boldsymbol{V}_{y_k}$, $\boldsymbol{V}_1$, $\boldsymbol{V}_2$ are the motion vectors of the block $y_k$, vehicle $O_1$ and vehicle $O_2$, respectively. $D = |\boldsymbol{V}_1 - \boldsymbol{V}_2|$ is the distance between $\boldsymbol{V}_1$ and $\boldsymbol{V}_2$. We can assume that the block $y_k$ is part of vehicle $O_1$, so $\boldsymbol{V}_{y_k}$ locates in the range of the left circle in Figs. 2(a) and (b). If $D$ is relatively large ($D \geqslant R_1 + R_2$, as shown in Fig. 2(a)), $|\boldsymbol{V}_1 - \boldsymbol{V}_{y_k}| < |\boldsymbol{V}_2 - \boldsymbol{V}_{y_k}|$ and $U(O_1) < U(O_2)$, the block $y_k$ is labeled as $O_1$. But if $D$ is small ($D < R_1 + R_2$, as shown in Fig. 2(b)), $|\boldsymbol{V}_1 - \boldsymbol{V}_{y_k}| > |\boldsymbol{V}_2 - \boldsymbol{V}_{y_k}|$ and $U(O_1) > U(O_2)$, the block $y_k$ may be labeled mistakenly as $O_2$. In other words, when two vehicles' motion vectors are similar, the occlusion region may be segmented incorrectly. This is the second shortcoming of the method in Ref. [14]. To deal with the aforementioned problems, a novel vehicle tracking method is proposed in the next section.




(a)



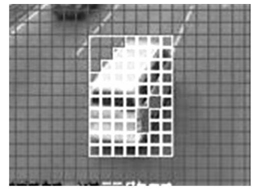

(b)

**Fig. 2** The relationship between $\boldsymbol{V}_1$, $\boldsymbol{V}_2$ and $\boldsymbol{V}_{y_k}$. (a) With large $D$; (b) With small $D$

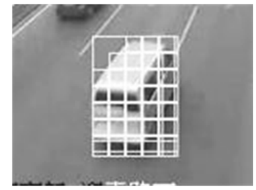
## 2 Novel Block-Based Vehicle Tracking Method

### 2.1 Vehicle division

The method in Ref. [14] divides the whole image and the vehicle simultaneously into several blocks, whereas the proposed method only divides the vehicle into blocks, each of which consists of $N \times N$ pixels. $N$ is chosen according to the vehicles' resolutions. When the vehicles contain more pixels, $N$ is set to be larger and vice versa. Because the vehicle's length and width are not always the integral multiples of $N$ pixels, the blocks on the vehicle's edge may overlap with its adjacent blocks. The two ways of division are compared in Fig. 3.



(a)



(b)

**Fig. 3** Two ways of division. (a) The division in Ref. [14]; (b) The proposed division ($N = 11$)

### 2.2 Detection and adjustment of noise motion vectors

After division, all the blocks' motion vectors are estimated as shown in Fig. 4(a). To avoid the expansion of blocks described in Section 1, the proposed method moves each block based on its own motion vector, but some motion vectors may be incorrect. So, it is necessary to detect and adjust noise motion vectors before moving the blocks.

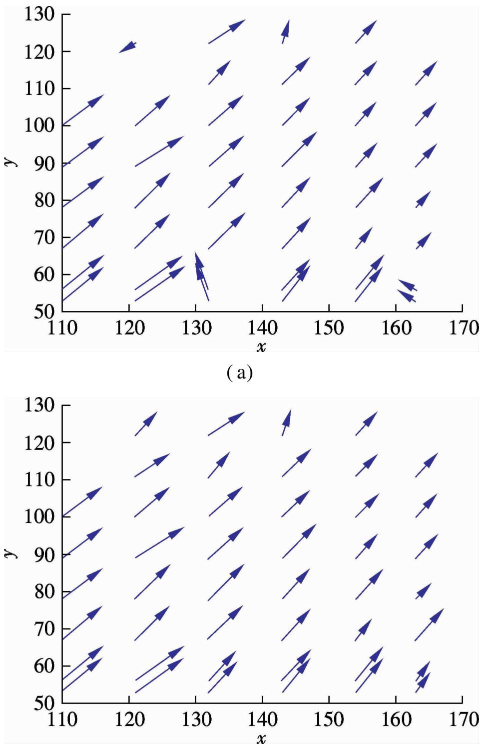



**Fig. 4** The adjustment of the motion vectors. (a) The original motion vectors; (b) The adjusted motion vectors

It is observed that the differences between the noise motion vector and its surrounding motion vectors are relatively large while the difference between the normal motion vector and its surrounding motion vectors are relatively small. Based on the smoothness of the motion vectors' spatial distribution, the noise motion vectors are detected and adjusted as below.

$\boldsymbol{V}_i$ is the motion vector of block $B_i$, and the neighborhood of $B_i$ is defined as

$$N_{B_i} = \{B_j \mid \mathrm{dis}(B_i, B_j) \leqslant T_1 \text{ and } i \neq j\} \tag{1}$$

where $\mathrm{dis}(B_i, B_j)$ is the distance between $B_i$ and $B_j$.

$S_{B_i} = \{\boldsymbol{V}_j \mid B_j \in N_{B_i}\}$ is the set of motion vectors of blocks in $N_{B_i}$, and the distance between motion vector $\boldsymbol{V}_i$ and motion vectors in $S_{B_i}$ is calculated by

$$D_{B_i} = \{\|\boldsymbol{V}_i - \boldsymbol{V}_j\| \mid \boldsymbol{V}_j \in S_{B_i}\} \tag{2}$$

$M_{B_i}$ is the mid-value of $D_{B_i}$. If $M_{B_i}$ is smaller than a specific threshold value $T_2$, $\boldsymbol{V}_{\mathrm{i}}$ is considered as a correct motion vector; otherwise, it is considered as a noise motion vector and will be replaced by the motion vector $\boldsymbol{V}'_i$, which corresponds to $M_{B_i}$. Fig. 4(b) shows the adjusted motion vectors. Compared Fig. 4(b) with Fig. 4(a), we can find that the noise motion vectors have adjusted well.

## 2.3 Blocks moving

After the adjustment of motion vectors, each block is moved using its own motion vector. If two moved blocks have the same coordinates, one of them is deleted from the set of blocks to reduce computational time. Finally, the rest of blocks constitute the vehicle's new location in the current frame.

Since we only divide the vehicle into a group of blocks and the moved blocks can be located at any position in the image, it is unnecessary to expand the moved blocks as described in Section 1. It means that the vehicle will be influenced little by its neighboring vehicle in tracking. The first shortcoming of the method in Ref. [14] has been avoided until now.

## 2.4 Occlusion segmentation

In the proposed algorithm, each vehicle's position is represented by the minimum rectangle which contains all the blocks belonging to the vehicle. If a rectangle overlaps another rectangle, one vehicle is considered to occlude another. The overlapping region is defined as the occlusion region, and occlusion segmentation is equivalent to determining the labels of the blocks in the occlusion region. It is assumed that each block's label is a discrete random variable whose distribution relies only on the distribution of its neighborhood, and all the blocks' labels compose a Markov random field. The neighborhood is defined as

$$N_i = \{i' \in S \mid |i - i'| \leqslant r, i \neq i'\} \tag{3}$$

where $|i - i'|$ is the Euclidean distance between block $i$ and block $i'$. For example, in Fig. 5, block $e$, $f$ and $g$ constitute the neighborhood of block $i$.

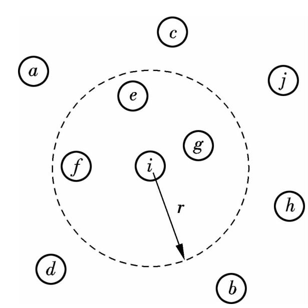



**Fig. 5** The neighborhood of block $i$

Based on the above assumption, the occlusion segmentation is equivalent to maximizing the probability $P(Z/Q)$, where $Z$ is the combination of labels and $Q$ is the current image. According to the Bayes rule, the optimal combination of labels $Z^*$ is

$$Z^* = \mathrm{argmax}(P(Q/Z)P(Z)) \tag{4}$$

The Hammersley-Clifford theorem proves that a Markov random filed can be expressed by a Gibbs distribution. So, we can obtain

$$P(Q/Z) = \frac{1}{M_1}\mathrm{e}^{-U(Q/Z)} \tag{5}$$

$$P(Z) = \frac{1}{M_2} \mathrm{e}^{-U(Z)} \tag{6}$$

where $M_1$ and $M_2$ are the normalization constants.

Eq. (4) is equivalent to $Z^* = \arg\min(U(Q/Z) + U(Z))$, and the energy function $U(Z)$ is defined as

$$U(Z) = \sum_{c \in C} F_c(z_i, z_j) \tag{7}$$

$$F_c(z_i, z_j) = \begin{cases} 0 & z_i = z_j \\ 1 & \text{else} \end{cases} \tag{8}$$

where $c = \{z_i, z_j\}$ is a clique and $C$ is the set of cliques.

We assume that the distance between the histogram of block $i$ and that of the non-occlusion region has the same label which follows that Gaussian distribution. So the energy function $U(Q/Z)$ can be defined as

$$U(Q/Z) = \sum_{i=1}^{n} \frac{(|\boldsymbol{h}_i - \boldsymbol{h}_{O_i}| - u_i)^2}{2\sigma_i^2} \tag{9}$$

where $\boldsymbol{h}_i$, $\boldsymbol{h}_{O_i}$ are the color histograms of block $i$ and the non-occlusion region of vehicle $O_i$ (The number of each histogram's bins is $64(4^3)$). If the label of block $i$ is $O_1$, $O_i = O_1$; otherwise, $O_i = O_2$. $u_i = 0$, and $\sigma_i$ is estimated empirically.

The minimization of the energy function is a combinatorial optimization problem. We use the simulated annealing algorithm[15] to minimize it. Finally, the labels $Z^*$ are given to the corresponding blocks and the occlusion region is segmented.

Fig. 6 shows an example of occlusion segmentation. In Fig. 6(a), the black blocks constitute the occlusion region, and in Fig. 6(b) they are segmented accurately by minimizing the energy function.


**Fig. 6** Occlusion segmentation. (a) Before segmentation; (b) After segmentation

It can be seen from Eqs. (7) to (9) that using the spatial color information instead of the motion vector to segment the occlusion region, the proposed method is insensitive to the motion vectors of vehicles in occlusion.

## 3 Experiments

To demonstrate the effectiveness of the proposed method, we apply it and the method in Ref. [14] to two traffic videos and define the tracking error of two adjacent vehicles as

$$E = |\boldsymbol{C}_1 - \boldsymbol{C}'_1| + |\boldsymbol{C}_2 - \boldsymbol{C}'_2| \tag{10}$$

where $\boldsymbol{C}_1$ and $\boldsymbol{C}_2$ are the two vehicles' true center coordinates which are determined manually; $\boldsymbol{C}'_1$ and $\boldsymbol{C}'_2$ are center coordinates which are obtained by tracking.

In the first video, no occlusion occurred but a couple of vehicles were close to each other. The tracking results are shown in Fig. 7. Due to the closeness, they were connected together and a false occlusion was created by the method in Ref. [14]. By contrast, they were still separated and tracked well by the proposed method.

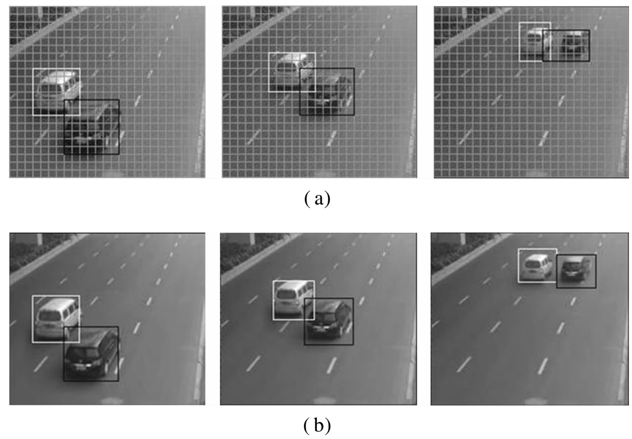



**Fig. 7** Experimental results of the first couple of vehicles. (a) The tracking results of the method in Ref. [14]; (b) The tracking results of the proposed method

In the second video, there are 19 pairs of vehicles in occlusion. If the tracking error exceeds 20 pixels, we consider it a failed tracking. As a result, the method in Ref. [14] and the proposed method can respectively track 14 and 17 pairs of vehicles. Three pairs of vehicles cannot be tracked well by the method in Ref. [14] because of their similar motion vectors; and two pairs of vehicles cannot be tracked by both methods because the occlusions occur at the beginning of tracking and the vehicles' initial locations are difficult to obtain. Fig. 8 shows the tracking results of a pair of vehicles which have similar motion vectors. $\boldsymbol{V}_1$ and $\boldsymbol{V}_2$ are motion vectors of the upper vehicle

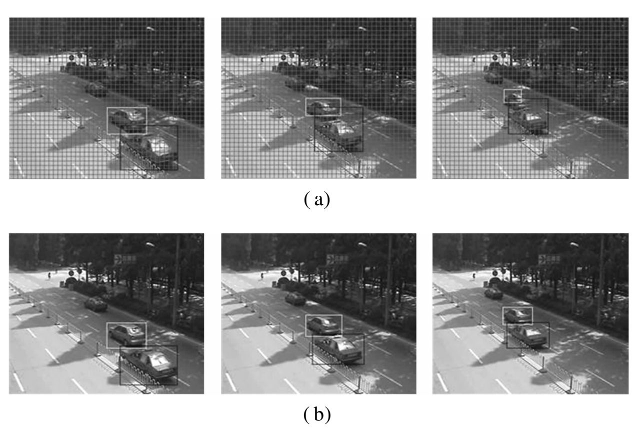



**Fig. 8** Experimental results of the second couple of vehicles. (a) The tracking results of the method in Ref. [14]; (b) The tracking results of the proposed method

and the lower vehicle, respectively, and in the tracking, the distance between them is less than 4 pixels. In the left images, $\boldsymbol{V}_1 = \{-2, -4\}$ and $\boldsymbol{V}_2 = \{-3, -6\}$. In the middle images, $\boldsymbol{V}_1 = \{-2, -4\}$ and $\boldsymbol{V}_2 = \{-3, -5\}$. In the right images, $\boldsymbol{V}_1 = \{-1, -3\}$ and $\boldsymbol{V}_2 = \{-2, -4\}$.

The tracking errors of two pairs of vehicles are shown in Fig. 9 and Fig. 10, respectively. The above experiments demonstrated that the proposed method can track vehicles under occlusion more accurately than the method in Ref. [14].

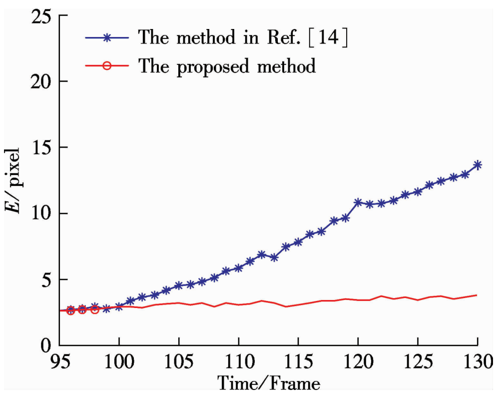



**Fig. 9** Tracking errors of the first couple of vehicles

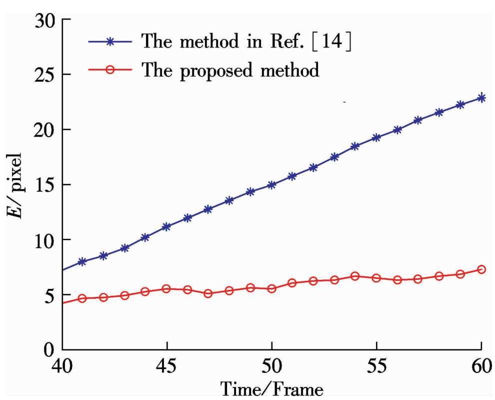



**Fig. 10** Tracking errors of the second couple of vehicles

## 4 Conclusion

In this paper, a novel vehicle tracking algorithm based on blocks is proposed. The vehicle is divided into a group of blocks, in which the edge block can overlap its neighboring block. The advantage of this division method is that the target vehicle is influenced little by its neighboring vehicle in tracking. Then the noise motion vectors of blocks are detected and adjusted. Based on the adjusted motion vectors, the blocks are moved to update the vehicle's location. When a vehicle is occluded by another vehicle, we establish a Markov random field to segment the occlusion region. In the Markov random field, the neighborhood system is defined using the Euclidean distance and the energy function is built based on the block's histogram rather than the vehicle's motion vector. This allows vehicles under occlusion which have similar motion vectors to also be segmented accurately. Experimental results demonstrate the effectiveness of the proposed method.

## References


[1] Mallikarjuna C, Phanindra A, Ramachandra K R. Traffic data collection under mixed traffic conditions sing video image processing[J]. *Journal of Transportation Engineering*, 2009, **135**(4): 174 – 182.

[2] Hu W M, Tan T N, Wang L, et al. A survey on visual surveillance of object motion and behaviors [J]. *IEEE Transactions on Systems, Man, and Cybernetics: Applications and Reviews*, 2004, **34**(3): 334 – 352.

[3] Hou Z Q, Han C Z. A survey of visual tracking [J]. *Acta Automatica Sinica*, 2006, **32**(4): 603 – 617.

[4] Yilmaz A, Javad O, Shah M. Object tracking: a survey [J]. *ACM Computing Surveys*, 2006, **38**(4) : 1 – 45.

[5] Harguess J, Hu C B, Aggarwal J K. Occlusion robust multi-camera face tracking[C]//*Proceedings of the* 2011 *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*. Colorado Springs, CO, USA, 2011: 31 – 38.

[6] Qian Z, King N N. Segmentation and tracking multiple objects under occlusion from multi-view video[J]. *IEEE Transactions on Image Processing*, 2011, **20**(11): 3308 – 3313.

[7] Zhang W, Wu Q M J, Yang X K, et al. Multilevel framework to detect and handle vehicle occlusion [J]. *IEEE Transactions on Intelligent Transportation Systems*, 2008, **9**(1): 161 – 174.

[8] Li Z, Tang Q L, Sang N. Improved mean shift algorithm for occlusion pedestrian tracking[J]. *Electronics Letters*, 2008, **44**(10): 622 – 623.

[9] Yan J, Wu M Y. Anti-occlusion tracking algorithm based on mean shift and fragments[J]. *Optics and Precision Engineering*, 2010, **18**(6): 1413 – 1419.

[10] Panahi R, Gholampour I, Jamzad M. Real time occlusion handling using Kalman filter and mean-shift[C]//*The* 8*th Iranian Conference on Machine Vision and Image Processing*. Zanjan, Iran, 2013: 320 – 323.

[11] Khan B, Khan A K, Raja G, et al. Implementation of modified mean-shift tracking algorithm for occlusion handling[J]. *Life Science Journal*, 2013, **10**(11): 337 – 342.

[12] Wang Z W, Yang X K, et al. Camshift guided particle filter for visual tracking[J]. *Pattern Recognition Letters*, 2009, **30**(4): 407 – 413.

[13] Abramson H, Avidan S. Tracking through scattered occlusion[C]//*Proceedings of the* 2011 *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*. Colorado Springs, CO, USA, 2011: 1 – 8.

[14] Kamijo S, Sakauchi M. Segmentation of vehicles and pedestrians in traffic scene by spatio-temporal Markov random field model [C]//*Proceedings of the* 21*st International Conference on Data Engineering Workshops*. Tokyo, Japan, 2005: 1 – 8.

[15] Tang L S, Xie Y, You S Y. *Non-numeric parallel algorithm-simulated annealing algorithm*[M]. Beijing: Science Press, 2000: 22 – 55. (in Chinese)

# 考虑遮挡的视频车辆跟踪

朱　周[1,3]　　路小波[2,3]

([1] 东南大学交通学院,南京 210096)
([2] 东南大学自动化学院,南京 210096)
([3] 东南大学复杂工程系统测量与控制教育部重点实验室,南京 210096)

**摘要:**为了对遮挡情况下的运动车辆进行跟踪,提出一种基于分块的车辆跟踪算法. 该算法将目标车辆以可重叠的方式划分为若干大小一致的子块. 在分块的基础上估计所有子块的运动矢量,检测噪声运动矢量并进行调整,以减少运动矢量估计的误差,然后对子块进行移位以实现车辆跟踪. 为了处理车辆间的遮挡现象建立了马尔可夫随机场描述子块之间的关系,利用欧氏距离定义块的邻域,并基于块的直方图构建能量函数,最后利用模拟退火法对能量函数进行优化,以对遮挡区域进行分割. 实验结果表明,该算法能够对遮挡车辆进行准确跟踪.

**关键词:**车辆跟踪;遮挡处理;运动矢量;马尔可夫随机场

**中图分类号:**U491.1