Detection and tracking of multiple moving targets in complex conditions

Xu Tao[1], Huang Chaobing[1], Zhu Mengqiang[1]

[1] Key Laboratory of Fiber Optic Sensing Technology and Information Processing, Ministry of Education, Wuhan University of Technology, Wuhan, Hubei, 430070, China

E-mail: whutxutao@163.com

Keywords: scene constraint, motion estimation, particle filter, MCMC, multiple target tracking.

Abstract. Detection and tracking of multiple moving targets in complex conditions has always been a hot and difficult problem, which mainly includes that the tracked target is occluded, the characteristics of target and background are similar, illumination and background are changing over time, etc. In order to solve this problem, a method of multiple moving targets detection and tracking is proposed in this paper, which can be applied to complex conditions. In the aspect of target detection, on the basis of mixed multi-scales and deformable model, we have proposed scene constraint and the target own property to reduce the potential areas of targets and search scope and to filter interference object, which has improved the efficiency of detection algorithm greatly. In the aspect of target tracking, we use global motion estimation to modify motion model of tracking algorithm, at the same time, propose an improved particle filter algorithm which has introduced reversibility-jump mechanism and joined MCMC improvement strategy. By building the observation model with the color histogram of target RGB, it can track the moving targets in different conditions accurately. The experimental results show that the method proposed in this paper can accurately track the multiple moving targets in complex conditions, and the effect is significant.

Introduction

Target tracking is an important research topic in computer vision field. With the continuous progress of society, the target tracking technology is gradually applied to all areas, such as military, aerospace, industrial production, transportation, etc. and it promote the development of these areas greatly. Especially in video surveillance, nowadays, the video surveillance system has been found in shopping malls, streets and other places, which has become an important part of protecting social order and people's daily work. Multiple moving targets detection and tracking is the core part of video surveillance system, which has important research significance and application value. Target detection is the process of acquiring the position, attitude and size of the target in the image sequences which contains target. In the static background, background subtraction [1], interframe difference [2] and optical flow [3] are the common algorithms of target detection. This method can be used to the target detection under static or simple background, but it seems to be powerless about the camera motion and the messy environment. In this paper, this environment is usually referred to complex background; and in the face of complex background, we usually use motion compensation or training classifier. The target detection results are usually used as the initialization state of target tracking. The accuracy of the target detection determines the accuracy of target tracking results in a certain extent. In the process of target tracking, the feature description of target is a very important problem. Now, for a simple scene and a fixed camera, selecting visual information as the target description of tracking algorithm can achieve good results basically. Dorin Comaniciu et al convert the target color feature to the LUV color space, which makes the feature more robust [4]. However, when faced with complex environment (such as cluttered background, crowded scene), it will encounter tracking degradation problem. At this point, another method of target detection by a sample training classifier has a very good performance. Felzenszwalb et al proposed a discriminative training component model to detect target and a graph structure framework to divide the object into several parts, which makes each part connected with a spring, and with a series of relations between components and parts to represent the object [5]. The system by training is not

only efficient but also accurate. However, in addition to using a variety of information cues, we can also consider the interaction between targets. Therefore, the interaction model is proposed. Zia Khan Z and Tucker Balch et al applied the thought of interaction model to realize the multiple object tracking by using MCMC particle filter [6].Then, Wongun Choi and Silvio Savarese et al integrated the information like camera motion estimation, scene geometry and so on into interaction model, and constructed a complete multiple object tracking framework based on interaction model [7, 8].

This paper mainly researchs detection and tracking of multiple moving targets in complex conditions. In order to reduce the harmful effects of the moving background to target tracking, this paper estimates the global motion state, which is used to modify the motion model of tracking algorithm. In the aspect of detection, the traditional detection and tracking system often ignores scene information, and focuses on the target itself. In this paper, we consider the influence of scene information on target detection; on the basis of mixed multi-scales and deformable model, we have proposed scene constraint and target own property (such as height limit) to reduce the potential areas of targets and search scope and to filter interference object, which not only has improved the efficiency of detection algorithm greatly, but also has good detection effect in complex environment. In the aspect of tracking, this paper has proposed joining MCMC improvement strategy based on the basic particle filter to increase the diversity of particles and reduce the calculation of algorithm; and by building the observation model with the color histogram of target RGB, the robustness of target description has improved. Meanwhile, in order to solve the problem of random variation of target number, we introduce reversibility-jump mechanism and establish five kinds of target states: Add、Delete、Leave、Stay、Update, which can track the moving targets in different conditions accurately.

Scene constraint and target detection

Scene information has a very important role in target detection, in which the influence of scene on target detection is often overlooked. Most detection algorithms are only interested in the information contained in the target, and don't take into account the relationship between target and scene. On the other hand, the detection algorithm generally regards the scene or the background as the interference to target detection. In fact, the recognition rate of the detection algorithm can be greatly improved by using the scene information of the target.

Scene constraint

The content of scene constraint is: (1) in the situation of camera parallel to the horizon mobile, target detection at the bottom of the box position should be in a half below the height of image; (2) through estimating the imaging height range of pedestrian targets, the height of target detection frame should be in a certain range. On the basis of target detection algorithm, the two constraints are used to filter the non-target items to improve the accuracy of detection results. The application effect is shown in Figure 1.

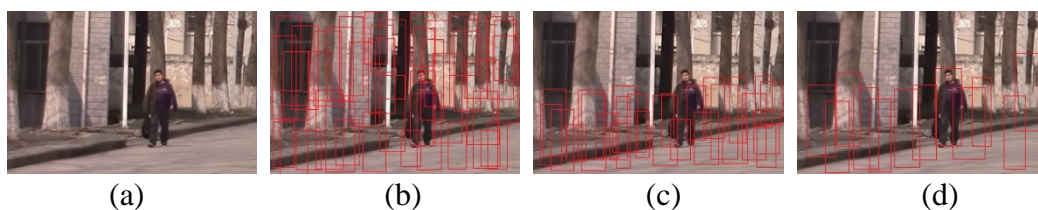| (a) | (b) | (c) | (d) |

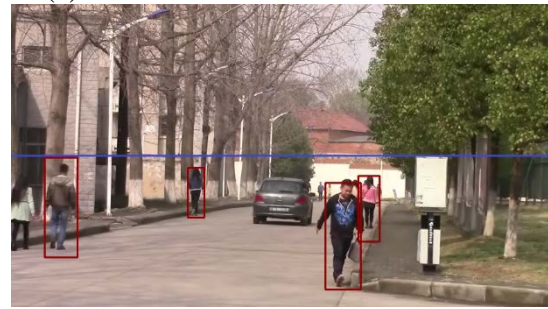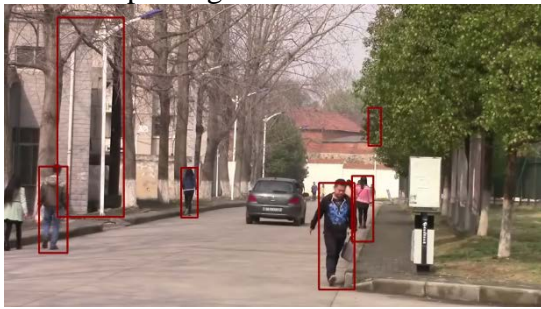Figure 1　Application effect of scene constraint in target detection

In Figure 1 (a) to find the possible location of the target (pedestrian), without considering the scene structure information, the target has the same possibility existing in any position and size of the image. In Figure 1 (b), the red box in the picture is the sampling frame. This is also the assumption of most detection algorithms. But if we use the scene information of geometry, we can eliminate a lot sampling frames which do not meet the constraint. The results are shown in Figure 1 (c). We use the camera pose information and the true height range of the target to estimate the size

range of the target in image, and filter out the sampling frame which does not meet the size range. The results are shown in Figure 1 (d). Before using detection algorithm, we can use the two constraints above to achieve the effect in Figure 1 (d). The result shows that the bottom detection algorithm can be improved in essence by considering the potential of the scene structure. Taking into account the scene constraint can reduce the search range, filter interference, and effectively improve the efficiency of detection algorithm.

Target detection and result

For the target detection algorithm, the reliability of detection algorithm is related to the success of target tracking. Therefore, the accuracy of target detection is very important. In this paper, we proposed to join scene constraint on the basis of mixed multi-scales and deformable model, which can not only improve the accuracy of target detection, but also greatly improve the efficiency of detection algorithm.

In the target detection experiment of actual scene, we choose the image sequence of the campus image took by ourselves to experimental tests. When the scene constraint is not considered, the effect of multiple targets detection is shown in Figure 2 (a).

(a) Target detection effect of original algorithm   (b) Target detection effect in scene constraint

Figure 2   The detection effect contrast before and after joining scene constraint

From Figure 2 (a) we can see that, there has a similar background in the presence of target (the boys in the left side are in a similar color to the background), and the distant target is small, and the target is close and multiple targets etc. In these cases, the target detection algorithm has a good detection effect. However, there are two false detection boxes. One is that the height of detection frame is too large, which does not conform to the height of pedestrian; the other is that the bottom of detection frame position is above the terrain, or not in the image below a half height line. Next, we use scene constraint to filter interference target and to correct this error. The method is as follows.

1）   The target detection algorithm is used to obtain the initial set of targets.

2）   Filter error target through the bottom position of target detection frame.

3）   Estimate target imaging height, and filter non-target item.

In the experiment, the height range of pedestrian target is $y \in (1.2，2.2)$ meter, camera height is $y_c = 1$ meter, image size is $w \times h = 640 \times 480$. The target imaging height can be estimated by using the relationship between camera pose and target [9].

$$h_i \approx \frac{y\left(\frac{h}{2} - v_i\right)}{y_c} = \frac{y(h - 2v_i)}{2y_c} \tag{1}$$

Among them, h is the image height, v_i is the bottom position of detection frame, y is true height of pedestrian, y_c is the camera height, h_i is the imaging height of pedestrian. Comparing the estimate of detection frame height with the detection of target frame, we will eliminate the targets outside the range of estimate. Considering scene constraint, the detection effect is shown in Figure 2 (b). In the Figure the blue line is a half of image height line with removing the detection frame whose bottom position is not in this line. The experimental results show that scene constraint can effectively improve the accuracy of detection algorithm.

Tracking the moving targets after detection, then entering next detection tracking processing cycle.

Moving target tracking

Aimed at Tracking of multiple moving targets in complex conditions, the moving background has a large interference to target tracking. To solve this problem, in this paper, we use global motion estimation to modify motion model of tracking algorithm, and propose an improved particle filter algorithm, and join MCMC improvement strategy to optimize particle distribution, and build the observation model with the color histogram of target RGB to improve the robustness of target description, and introduce reversibility-jump mechanism to solve the problem that the target number is uncertain. By combining with each other it can track the multiple moving targets accurately. The whole process of multiple moving targets tracking is shown in Figure 3.
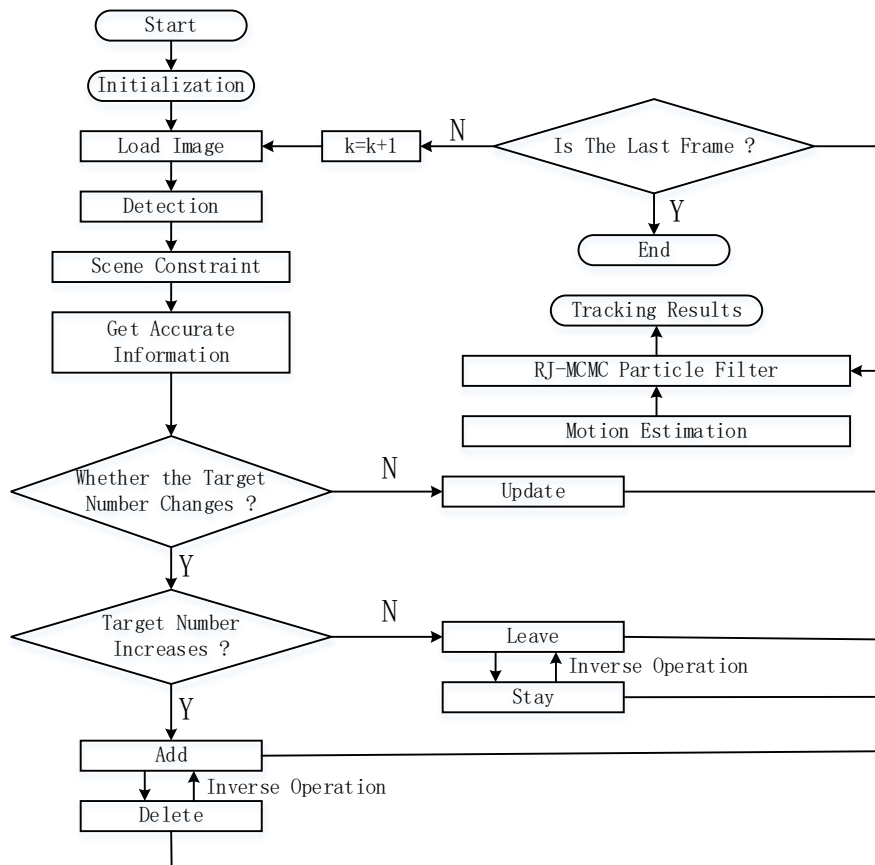

Figure 3   The process of multiple moving targets tracking

Global motion estimation to modify motion model

Moving background often has a very bad influence on target tracking, such as target position prediction error, etc. The general processing method is to carry on image stabilization processing and tracking, but it will cause the image sequence information loss, and can't get the tracking effect in the original image. According to this we estimates the global motion state, which is used to modify the motion model of tracking algorithm to achieve good tracking results under moving background.

Building the motion model is to model the dynamic information of target, whether the motion model is good or bad which is related to the accuracy of predict target's location in the next moment, having great influence on the performance of tracking algorithm. In the concrete realization of particle filter, the two auto regressive model is generally used. However, in order to reduce the impact of moving background on target tracking, in this paper, the global motion estimation is introduced based on the traditional two auto regressive motion model to modify the motion model and improve the reliability of target tracking. The two auto regressive model is introduced as follows.

Suppose $\{x_t\}$ is the stationary process of zero average, appropriate model of $\{x_t\}$ is

$$x_t = \varphi_1 x_{t-1} + \varphi_2 x_{t-2} + a_t \tag{2}$$

Among them, a_t is the white noise sequence.

If meet $E(x_t, a_t) = 0, (t < s)$, we deem x_t follows a two auto regressive stochastic process, the model is the two order auto regressive model. If the formula (2) is changed to its equivalent form, that is

$$x_t - \varphi_1 x_{t-1} - \varphi_2 x_{t-2} = a_t \qquad (3)$$

Through formula (3) we can see that, the two order auto regressive model remove the part which depend on x_{t-1}、x_{t-2} in x_t , and make the data into independent data a_t. From this we can see, the past state of $t - 1$ and $t - 2$ can be used to predict the state of t by using the auto regression model.

In this paper, the position of target in the image is used as the state variable [10], then the state variable of k moment can be expressed as $X_k^i = [u_i(k), v_i(k)]^T$. Among them, $(u_i(k), v_i(k))$ is the pixel coordinates of target in the image. According to the third chapter of the rotation matrix R_k and translation vector t_k which are obtained from global motion estimation, we can modify the state variable, and get the correction Δ_k , as follows

$$\Delta_k = R_k X_k^i + t_k \qquad (4)$$

Among them, R_k、t_k are rotation matrix and translation vector of k moment respectively.

Using the correction Δ_k to modify motion model

$$X_{k+1}^i - \hat{X}_k^i = A_1 \left(X_k^i + \Delta_k - \hat{X}_k^i \right) + A_2 \left(X_{k-1}^i + \Delta_k + \Delta_{k-1} - \hat{X}_k^i \right) + B_w \qquad (5)$$

The formula (5) is the final motion model used in this paper. Among them, $\hat{X}$ is the estimated value of particle, A_1、A_2 is the coefficient of two auto regressive model, B_w is the system random noise.

RJ-MCMC improvement strategy

In the field of target tracking, Particle filter algorithm (called PF for short) has become the mainstream algorithm because of its ability to deal with the non-linear and non-gauss state [11]. However, the particle degradation and the random variation of target number make the tracking process not stable, and cause that the target can't be tracked accurately. In order to solve this problem, we introduce MCMC improvement strategy to increase the diversity of particles and use reversibility-jump mechanism to solve the problem of uncertainty about the number of targets.

The particle produced by importance distribution function $q(x_{0:k}|z_{1:k})$ through MCMC strategy and a Markov chain transition nuclear $K(x_{0:k}\hat{x}_{0:k})$ makes it to a place where is closer to the real posterior probability density of system [12], namely

$$\int K(x_{0:k}\hat{x}_{0:k})q(x_{0:k}|z_{1:k})\, d\hat{x}_{0:k} = p(x_{0:k}|z_{1:k}) \qquad (6)$$
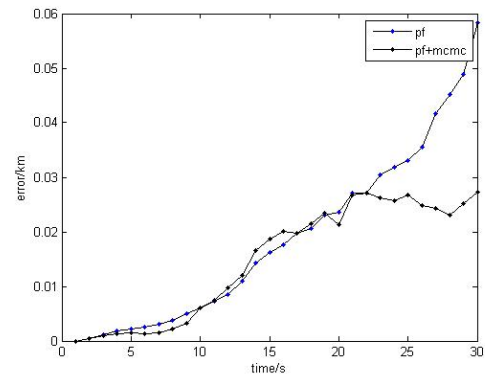
Among them, $\hat{x}$ is optimal estimation of x, $p(x_{0:k}|z_{1:k})$ is posterior probability density.

The new particles have a good convergence, which make the distribution of the sample more reasonable, and thereby avoid the degradation phenomenon of particle effectiveness and diversity.

The performance difference between basic particle filter and MCMC particle filter is compared by particle tracking simulation.
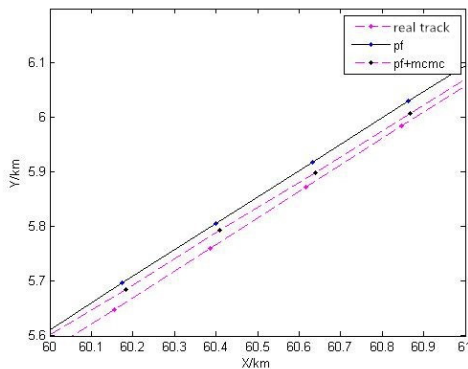
(a) Particle track and observation value   (b) Distance total error of observation points

Figure 4　Performance comparison simulation between PF and PF- MCMC

The simulation of basic particle filter algorithm and MCMC particle filter algorithm of the distance error of each observation point is shown in Figure 4. The blue points represent basic particle filter and black points represent MCMC particle filter. Figure 4 (a) is the real track of the particle and the observation values of two algorithms. Figure 4 (b) is the total observation error of the two algorithms in XY direction. The estimation error of each point is represented by $err = \sqrt{(x - \hat{x})^2 + (y - \hat{y})^2}$, the (x, y) and $(\hat{x}, \hat{y})$ are real location and observation position of particle respectively. In the first 20 sampling points, two algorithms can achieve smaller errors, and the performance of two algorithms have little difference, but there is obvious difference after 20 sampling points in performance. Because MCMC makes the particle gain more reasonable distribution, which is more excellent than particle filter.

The essence of the reversibility-jump mechanism is to apply the Metropolis Hasting sample to a string Markov chain whose dimension of state variables is variable. The particle state is represented by $S_t^i = \{k_t^i, X_{k_t^i}\}$ in i times iteration, and selects one of the five movements of Add, Delete, Leave, Stay and Update with the random choice of probability, according to the proposed distribution $q(\cdot)$ to produce particles of candidate states S_t' to carry out the i+1 times iteration[13]. It is assumed that we has obtained N unweight particles $\{X_i\}_{i=1}^N$ to approach the posterior probability $P(X_{t-1}|Z_{t-1})$ of $t - 1$ moments. According to the acceptance rate of status transfer by calculation, we can judge the situation of status transfer, and use the N unweight particles produced by RJ-MCMC sampling to approach the posterior probability $P(X_t|Z_t)$ of t moments, calculate the mean of particle and get the state of each target of t moments, then complete the whole process of RJ-MCMC sampling.

Observation model

Because of the complexity of moving background and the change of target appearance, the segmentation of moving objects is uncertain. After the modified target motion model, we can predict where the target may appear in the next moment. By using observation model to measure the similarity between the feature distribution in candidate target region and the feature distribution of real target, we can improve the robustness of target description. In this paper, we use an apparent model to describe the target by color histogram information of RGB color space. The channel value is divided into $m = 64 \times 64 \times 64$ grades, and in a certain extent, the calculation of the subsequent processing is reduced, meanwhile it also ensures the robustness of the model. The similarity of RGB color histogram distribution between the target region and the candidate target region is needed to establish the observation model. In the comparison of histogram similarity, Bhattacharyya distance is a common used method. In this paper, the similarity of the two color histogram distribution models is compared with the normalized Bhattacharyya distance [14]. Bhattacharyya coefficients can be defined as

$$\rho[\hat{p}(y), \hat{q}] = \sum_{u=1}^m \sqrt{\hat{p}_u(y)\hat{q}_u} \tag{7}$$

Therefore, the two region histogram similarity can be expressed as

$$d = \sqrt{1 - \rho[\hat{p}(y), \hat{q}]} \tag{8}$$

Finally, we use the normal distribution model to establish the observation likelihood model, as follows

$$p(Z_k|X_k) = \frac{1}{\sqrt{2\pi\sigma^2}} exp\left(-\frac{d^2}{2\sigma^2}\right) \tag{9}$$

Take target 1 and target 2 as an example to carry out the statistics experiment of RGB color histogram. And the test result is shown in Figure 5.

Table 1   Test results of Bhattacharyya distance

Bhattacharyya distance	Target 1（1062th frame）	Target 2（1064th frame）
Target 1（1064th frame）	0.121155	0.406824
Target 2（1062th frame）	0.431246	0.208038

From Figure 5 we can see that the RGB color histogram distribution is close to the same target in different frames. It shows that the model has good robustness in complex environment. Table 1

shows the Bhattacharyya distance between target 1 and target 2. Smaller the value is, closer the similarity is. The results show that the Bhattacharyya distance of the same target in different frames is smaller and the Bhattacharyya distance of different target in the same frame is larger, which are conducive to distinguishing target, and it shows that the reliability of the model is high to greatly improve the robustness of target description.
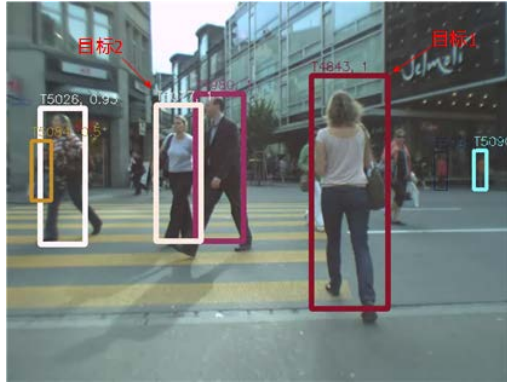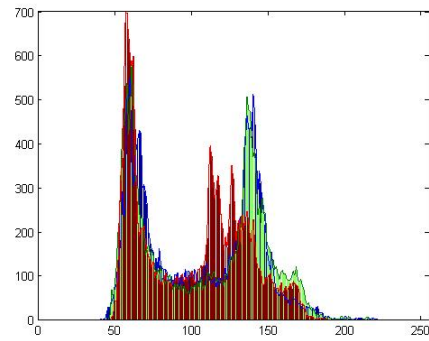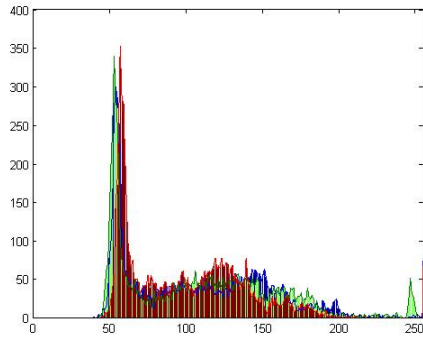

(a) The 1062th frame
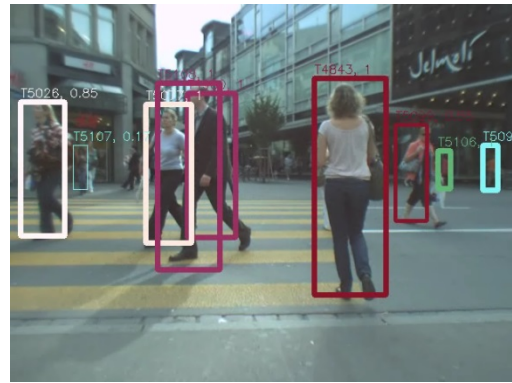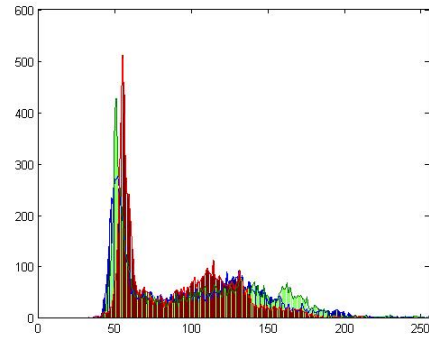
(b) The 1064th frame
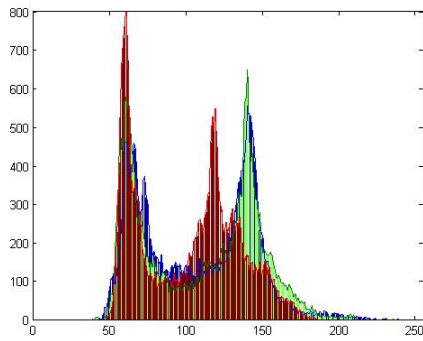
(c) RGB color histogram of target 1 in 1062th frame

(d) RGB color histogram of target 1 in 1064th frame

(e) RGB color histogram of target 2 in 1062th frame

(f) RGB color histogram of target 2 in 1064th frame
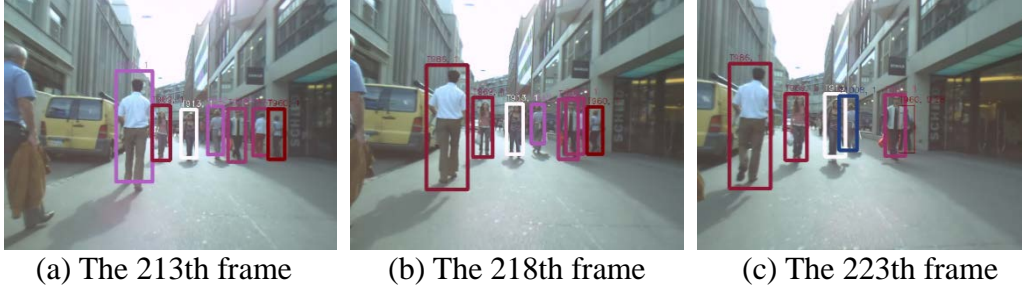
Figure 5   RGB color histogram

Experimental results

In this paper, we propose a new method to introduce the scene constraint, and adopt global motion estimation to modify motion model, and use the modified RJ-MCMC particle filter tracking algorithm to carry out multiple targets tracking. Then we test the method by experiment.

In the experiment, two segment image sequences of ETH pedestrian data set were tested. The data set is based on the mobile platform, including the complex conditions of occlusion, illumination change, background clutter and so on which can meet the requirements of this experiment. The specific properties of test image sequence is shown in table 2.

Table 2　The properties of test image sequence

Sequence number	1	2
Sequence name	BAHNHOF	LINTHESCHER
Frame	1000	1209
Size	640×480	640×480
Format	png	png
Bit depth	24	24

The test results of sequence 1

(a) The 213th frame　　(b) The 218th frame　　(c) The 223th frame

The test results of sequence 2

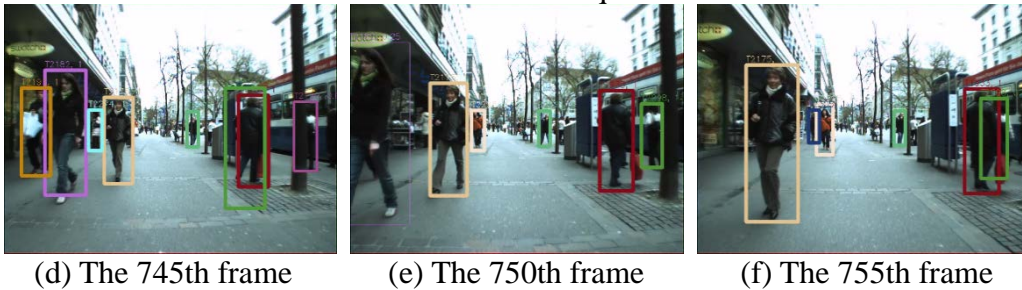(d) The 745th frame　　(e) The 750th frame　　(f) The 755th frame

Figure 6　The experimental results of tracking

For each image sequence, we select 3 images with a phase difference of 5 frames. These images include illumination changes, pedestrian occlusion, pedestrian appearance changes, background clutter and so on. In order to facilitate observation, we set up a unique label for each tracking target. In the case of some targets, in Figure 6 (b) and (c), the target T960 (the most right in the image) can be successfully tracked after being occluded by other targets partially. In sequence 2, the most right target (the woman in white shirt) in the three images of Figure 6 (d), (e), and (f) is still successfully tracked after being occluded by the tree and other targets partially. Generally, in the six frame images, the pedestrian in the scene can be tracked accurately. The test results show that the tracking algorithm we proposed can track target accurately in complex conditions of illumination changes, target occlusion, background clutter and multiple moving targets, which also has good robustness and accuracy.

Conclusion

In this paper, we mainly research detection and tracking of multiple moving targets in complex conditions. Firstly, we introduce scene constraint on the basis of mixed multi-scales and deformable model which has improved the efficiency and accuracy rate of detection algorithm greatly; at the same time, we use global motion estimation to modify motion model of tracking algorithm. To overcome the tracking disturbance caused by moving background, we estimate the motion state of a camera through two consecutive frames. Then we improve the degradation phenomenon of basic particle filter by using improved RJ-MCMC particle filter tracking algorithm and MCMC strategy. Secondly, we build the observation model with the color histogram of target RGB to improve the feature of target description, and join reversibility-jump mechanism to deal with the random problem of target motion. Finally, we can track the multiple moving targets accurately, and it have good robustness and accuracy in complex conditions. Although the target detection algorithm proposed in this paper can achieve better results under moving background, the calculation is

complex and real-time effect is not satisfactory. These problems need to be further studied in the following work.

Acknowledgment

References

[1] CHEN SH Y, HUANG Y W, HSIEH B Y. Fast video segmentation algorithm with shadow cancellation global motion compensation and adaptive threshold techniques [J]. IEEE Trans on Multimedia, 2004, 6(5): 732-748.

[2] COLLINS R. A system for video surveillance and monitoring VSAM final report[R]. Carnegie Mellon University. Technical Report: CMU2RI2TR200212, 2000.

[3] BARRON J, FLEETD, BEAUCHEM IN S. Performance of optical flow techniques [J]. International Journal of Computer Vision, 1994, 12(1): 42-77.

[4] Dorin Comaniciu, Peter Meer. Mean shift: A robust approach toward feature space analysis [J]. IEEE transactions on pattern analysis and machine intelligence, 2002, Vol. 24(5).

[5] Pedro F Felzenszwalb, Ross B Girshick, David McAllester, et al. Object detection with discriminatively trained part based models[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2009, Vol. 32: 1627-1645.

[6] Zia Khan, Tucker Balch, Frank Dellaert, et al. MCMC-Based Particle Filtering for Tracking a Variable Number of Interacting Targets [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2005, Vol. 27: 1805-1819.

[7] Wongun Choi, Caroline Pantofaru, Silvio Savarese. A General Framework for Tracking Multiple People from a Moving Camera [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2012, Vol. 35: 1577-1591.

[8] Wongun Choi, Caroline Pantofaru, Silvio Savarese. Detecting and Tracking People using an RGB-D Camera via Multiple Detector Fusion [J]. IEEE International Conference on Computer Vision Workshops, 2011: 1076-1083.

[9] Derek Hoiem, Alexei A Efros, Martial Hebert. Putting Objects in Perspective[C]. International Journal of Computer Vision, 2008, Vol. 80: 3-15.

[10] Jiang Mingxin. Real-time Object Tracking Algorithm with Cameras Mounted on Moving Platforms [J]. Journal of Dalian Nationalities University, 2014, Vol. 16(1).

[11] Wang Xin, Xu Lizhong. Image target tracking technology [M]. Beijing: The People's Posts and Telecommunications Press, 2012.

[12] Gao Jing, Li Shangji, Shao kuijun. Application and research of MCMC particle filter algorithm [J]. Electronic Test, 2009(12): 19-22.

[13] Zhou Wei, Xu Haixia, Zheng Jinhua. Multi-object Visual Tracking Based on Reversible Jump Markov Chain Monte Carlo [J]. Computer science, 2012, Vol. 39(7): 270-275.

[14] Chen Weibing. Comparison of matching capabilities in similarity measurements [J]. Computer application, 2010, Vol. 30(1): 98-100.