International Journal of Signal Processing, Image Processing and Pattern Recognition
Vol.9, No.1 (2016), pp.311-322
http://dx.doi.org/10.14257/ijsip.2016.9.1.30

Improved Algorithm of Association Matching for Pedestrian Tracking in Occlusion Scene

xiaofeng ZHANG[1] , jianlan LIAO[2], dixing LI[3] and guowei YANG[4]

[1,2,3,4]*School of Information Engineering, Nanchang Hangkong University, No.696, Fenghe Nan Avenue, Nanchang, Jiangxi Province, 330063, China*
E-mail: jianlan3373@163.com

Abstract

In order to improve the association matching of multiple pedestrians in occlusion scene, an improved algorithm is proposed in this paper that combines the similarity matching of pedestrian object's major color spectrum histogram combination with joint probabilistic data association. The similarity to major color spectrum histogram of pedestrian's object is calculated between the previous and the frame that current frames in this algorithm. And take the centroid location of pedestrian's target which is with the most largest similarity to correct the state vector of the JPDA algorithm in the current frame. The experimental results show that the proposed algorithm could make the number of the targets' label adhesion reduced to 40%, the number of the targets' label confusion decreased about 33%, the recall rate of targets is increased by 6.8%, and the tracking accuracy is improved by 9.2% compared with the classical JPDA algorithm when the occlusion rate between pedestrians and intensity are not very high.

Keywords*: pedestrian tracking, occlusion, part model combination, similarity matching, joint probabilistic data association*

1. Introduction

Video pedestrian tracking is one of the most important research directions in computer vision[1], and it is widely used in security monitoring, human computer interaction and other fields [2-3]. In order to solve this tracking problem, researchers have proposed many tracking algorithms. However, there are still many difficulties in tracking pedestrians; one of the most important challenges is occlusion, including the occlusion between pedestrians as well as the pedestrians between pedestrians and background.

To achieve multi-target tracking, the recognition of different moving targets in successive frames, is needed to carry on data association according to the detection result of the surveillance video which exists multiple pedestrians. Joint probabilistic data association (JPDA) is a data association algorithm that is commonly used in multi-targets tracking. This is an association matching algorithm based on probability which is commonly used for point targets in clutter environment. This kind of methods which are based on probability and applied to the video images tracking system of pedestrian, will easy make targets' label adhere and confuse when there exists occlusion between the pedestrians[4]. So-called adhesion, means that certain pedestrian targets' label are associated with a target in the algorithm, and misses the other targets' trace. Confusion is that some targets' label happen to exchange error. This is because that the classic JPDA algorithm[5] only uses status information of targets(*e.g.,* target position, velocity, *etc*.) when we're taking measurement with the state of multiple targets. so that if the target tracking gate have some overlap, then error association which lead to the emergence of adhesion and confusion is prone to happen.

In order to overcome the impact of occlusion effectively, a number of anti-occlusion tracking algorithm has been proposed in recent years. The occluded situation is

ISSN: 2005-4254 IJSIP

considered as a state component of the target in reference [4], and observations confidence of the target is calculated on the basis of the estimation of occlusion. This method can deal with the adhesion problem when the population density is not high, but the targets' label processed is not good. Objects are distinguished between the prior object and the occluded object by the method based on color histogram matching in reference [6].This method has a fast matching speed, is prone to emerge error matching however when different objects have similar colors, and re-division after target occlusion emerging. The tracking method which relies on detection is used in reference [7], detect pedestrian objects by using a variable part model (DPM), and train classifier for multiple pedestrians occluded each other by using support vector machine method. Although it improves the missing rate of pedestrian targets, but still haven't address the problems of adhesion and confusion for targets' label effectively. Andriyenko etc believed that multi target tracking problem can be modeled as a discrete continuous optimization problem in reference [8-9], in Which discrete optimization and labels cost is used to execute data association, resulting in a near optimal solution, its tracking accuracy is better, but the adhesive number of targets' label is still high. For the questions of literature [8-9] that target's label will be adhesion, Andriyenko has proposed an improved algorithm that combined Conditional random field (CRF) model in the literature [10], which greatly improves the issues of the targets' label adhesion, but for the question of label's confusion also can not get better improvement.

For the problems the defects that the targets' label is prone to emerge adhesion and confusion in multiple pedestrian targets tracking where above reference mentioned, an improved joint probabilistic data association algorithm combined with pedestrian objects similarity matching method is proposed in this paper. It is took centroid location of pedestrian target which has maximum similarity largest to correct the state vector of the JPDA algorithm in current frame by calculating the major color histogram similarity spectrum of pedestrian object which is detected in successive frames.

2. Introduction of Classic JPDA Algorithm

We assume that T-targets were tracked in the crowded and clutter environment, and measured in discrete time. The state equation and measurement equation goal are shown as follows:

$$\begin{cases} X^t\left(k+1\right) = F^t(k)X^t(k) + V^t(k) \\ Z^t\left(k\right) = H\left(k\right)X^t\left(k\right) + W\left(k\right) \end{cases} \tag{1}$$

Where $k = 0,1,2,\cdots, t = 1,2,\cdots,T$; $F^t(k)$ is the state transition matrix of target t in the moment k. $X^t(k)$ is the state vector of target t in the moment k. $V^t(k)$ is the zero mean and white Gaussian processing noise sequence of target t in the moment k. $H\left(k\right)$ is the measurement matrix. $W\left(k\right)$ is the zero mean and white Gaussian measurement noise sequence of target t in the moment k.

When the measurement falls in the confirmation gate of a certain target, the measurement will be considered to be a valid measurement. We assume that there are m_k valid measurements. Then the associated probability of the i-th valid measurement and the target t is shown as formula (2):

$$\beta_{it}\left(k\right) = \sum_{j=1}^{N_k} \Pr\left\{\theta_{it}\left(k\right) | Z^k\right\} \hat{\omega}_{it}^{j}\left(\theta_j\left(k\right)\right) \tag{2}$$

The state estimation of target t in the moment k is shown as formula (3):

$$\begin{aligned}\hat{X}^t(k|k) &= \sum_{i=0}^{m_k} \beta_{it}(k)\hat{X}_i^t(k|k) \\ &= \hat{X}^t(k|k-1) + K^t(k)\sum_{i=0}^{m_k}\beta_{it}(k)v_i^t(k)\end{aligned} \tag{3}$$

Where $\theta_{it}(k)$ represents the event that measurement i and target t are associated at time k. $\theta_j(k) = \bigcap_{i=1}^{m_k} \theta_{it_i}^j(k)$, where $\theta_j(k)$ represents j-th joint event. N_k represents the number of joint events. Z^k represents the set of all valid echo. $\hat{X}_i^t(k|k)$ represents the state estimation which is gotten with Kalman filtering for target t by the i-th valid measurement of target i.. $K^t(k)$ is the filter gain. The error covariance updated formula which is corresponding to the updated state estimate is shown as follow:

$$\begin{aligned}&P^t(k|k) = P^t(k|k-1) - (1-\beta_{0t}(k))K^t(k)S^t(k)K^t(k)^T \\ &+\sum_{i=0}^{m_k}\beta_{it}(k)\left(X_i^t(k|k)\hat{X}_i^t(k|k)^T - \hat{X}^t(k|k)\hat{X}^t(k|k)^T\right)\end{aligned} \tag{4}$$

3. Improved JPDA Algorithm Combined with the Major Color Spectrum Histogram Similarity Matching Method

3.1. Pedestrians Object Major Color Spectrum Histogram Similarity Matching Method

3.1.1 Parts of Pedestrian with Major Color Clustering: The major color spectrum histogram method which Piccardi *etc* proposed [12] Meanwhile, because this paper has considered the situation of occlusion, and uses the method of reference [14] to modeling those pedestrians in the target detection stage. The target is divided into 10 parts, including the head, torso, two thighs, two legs, two upper arms, two arm *etc*, the model of the human body is shown in Figure 1.

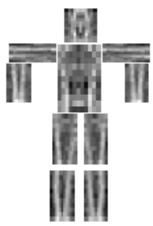

Figure 1. Human Body Model Learned From the Reference [14] Example Configuration

We can know by the prior knowledge, the major difference between pedestrian objects is the cloth color difference of the object. Therefore, upper body parts of the target(head, torso, upper arms, forearm) and lower body parts of the target (thighs, calves) is used the normalized Euclidean distance in the RGB color space. And the pixels of target parts are clustered by setting the threshold between classes. It is shown as follows:

$$d(C_1, C_2) = \frac{\|C_1 - C_2\|}{\|C_1\| + \|C_2\|} = \frac{\sqrt{(R_1 - R_2)^2 + (G_1 - G_2)^2 + (B_1 - B_2)^2}}{\sqrt{R_1^2 + G_1^2 + B_1^2} + \sqrt{R_2^2 + G_2^2 + B_2^2}} \tag{5}$$

Where C_1 and C_2 are two color vectors in RGB color space; R_i, G_i and B_i represent gray values of each color channel for the pixel i in RGB color space respectively. We assume that there exist M major colors in color spectrum of part i for target A, then the major color spectrum of part i for target A can be expressed as formula (6):

$$MCSHR(A_i) = \left\{C_{A_{i1}}, C_{A_{i2}}, \cdots, C_{A_{it}}, \cdots, C_{A_{iM_i}}\right\} \tag{6}$$

Where $C_{A_{it}}, i = 1, 2, \cdots, 10; t = 1, 2, \cdots, M_i$ is the major T-th color class of the part i in target A, The major color histogram statistics of part i in target A can be expressed as formula (7)

$$p(A_i) = \left\{p(A_{i1}), p(A_{i2}), \cdots, p(A_{it}), \cdots, p(A_{iM_i})\right\} \tag{7}$$

3.1.2. MCSHR Spatial Distribution Model for Parts of Pedestrian: According to the method of reference [14], combination of pedestrian body parts can be expressed as $L = \{l_0, l_1, \cdots, l_N\}$.

Where $l_i = (x_i, y_i, s_{li}, s_{hi}, \theta_i)$, x_i and y_i are the center position coordinates of part i in l_i; s_{li} and s_{hi} are the sizes of bounding box for part i; θ_i is the absolute orientation of part i. This paper learned from the thought of reference [13], and extended it to the parts model of pedestrian, calculated major color spectrum histogram combined with spatial information for each part of the pedestrian.

It is assumed that the center coordinates of M_i major color classes of part i for pedestrian target can be expressed as:

$$\left\{O_{i1}, O_{i2}, \cdots, O_{iM_i}\right\} = \left\{(x_{i1}, y_{i1}), (x_{i2}, y_{i2}), \cdots, (x_{iM_i}, y_{iM_i})\right\} \tag{8}$$

The distance-L_{it} between the center of each color class to the center of human body part is calculated as follow:

$$L_{it} = \sqrt{(x_{it} - x_i)^2 + (y_{it} - y_i)^2} \tag{9}$$

The position weight-l_t of each color class is defined as formula (10):

$$l_t = \frac{\max(s_{li}, s_{hi}) - l_{it}}{\max(s_{li}, s_{hi})} \tag{10}$$

The relative weight coefficient of each color class is gotten by normalizing l_t:

$$V_{it} = \frac{l_t}{\sum_{t=1}^{M} l_t} \tag{11}$$

The frequency statistics of major color for the human part is multiplied by relative weight coefficients before normalization processing, then can get frequency statistics value combined with the position weight of color class space distribution $p_{sp}(A_{it})$:

$$p_{sp}(A_{it}) = \frac{p(A_{it}) \times V_{it}}{\sum_{t=1}^{M_i} p(A_{it}) \times V_{it}} \tag{12}$$

3.1.3 Pedestrian Similarity Measurement and Match: In the occlusion scene, the pedestrian object similarity matching algorithm is shown in Figure 2.

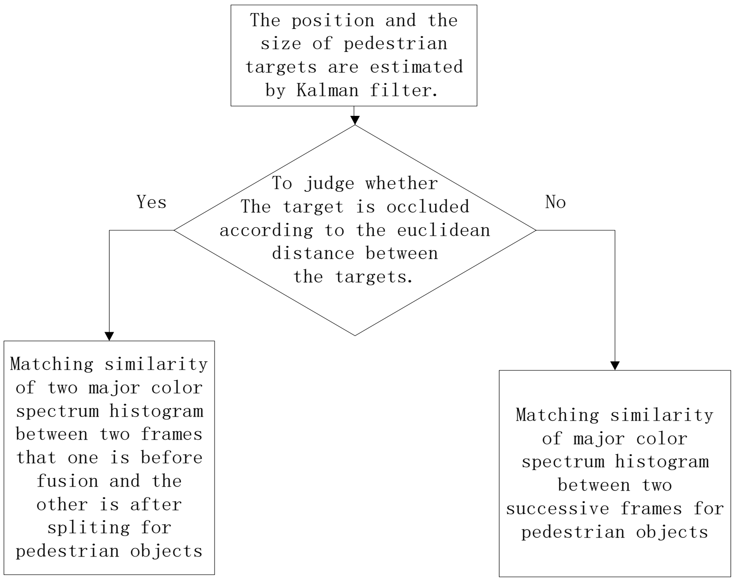

Figure 2. The Pedestrian Object Similarity Matching Algorithm in this Paper

In Figure 2, the occlusion between pedestrian targets means the serious occlusion between targets. *i.e.*, occluded object is more than half of the body occluded. Due to we consider the occlusion between target may exist directionality that is shown in Figure 3, where euclidean distance threshold is selected adaptively by using the following formula:

$$d_{threshhold} = \begin{cases} \frac{1}{2}w, D(a,b)\cos\theta > D(a,b)\sin\theta \\ \frac{1}{2}h, D(a,b)\cos\theta < D(a,b)\sin\theta \end{cases} \tag{13}$$

Where $D(a,b)$ is the Euclidean distance of pedestrian target centroid between a and b. θ is the included angle of two pedestrian target centroid line and a horizontal direction. And where h and w are respectively the height and width of bounding box for the occluded object.

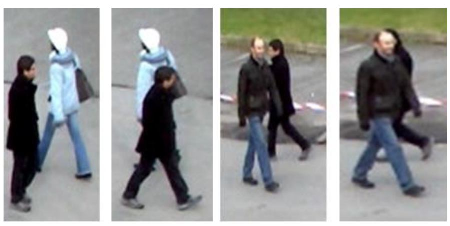

Figure 3. The Occlusion of Vertical and Horizontal Direction

We assume that the object of pedestrians in a frame for a goal. A, its matching a candidate target in another frame for B, similarly, in the part i' of the candidate target B, which the spatial distribution location weighting of main color spectrum histogram can be represented as $MCSHR'$ and the spatial distribution location weighting of the fusion color can be represented as $\mathrm{P_{sp}}(B_{i'})$, to define a subset of $MCSHR'$ as follows:

$$MCSHR'\left(B_i \mid C_{A_{it}}, \sigma\right) = \left\{C_{B'_{i1}}, C_{B'_{i2}}, \cdots, C_{B'_{iL}}\right\} \tag{14}$$

Where $C_{B'_{ij}}, j = 1, 2, \cdots, L$. And the euclidean distance between $C_{B'_{ij}}$ and $C_{A_{it}}$ is less than a given Cluster distance threshold σ.

$C_{B_{ij}|A_{it}}$ is defined the most similar color with $C_{A_{it}}$ in the subset $MCSHR'\left(B_{i'}\right)$. satisfies the formula (15):

$$C_{B_{i'j}|A_{it}} : j = \arg\min\left\{d\left(C_{B'_{i'k}}, C_{A_{it}}\right)\right\} \tag{15}$$

The frequency statistics of color class $C_{A_{it}}$ can be calculated simply by the formula (16) in the part i of target A:

$$p_{norm}\left(A_{it}\right) = \frac{p_{sp}\left(A_{it}\right)}{\sum\limits_{t=1,2,\cdots,M_i} p_{sp}\left(A_{it}\right)} \tag{16}$$

Similarly, in the part i' of the candidate target B, the normalized value of frequency statistics for color class $C_{B_{i'j}|A_{it}}$ is calculated as formula (17):

$$p_{norm}^{[A_{it}]}\left(B_{i'j}\right) = \frac{p^{[A_{it}]}\left(B_{i'j}\right)}{\sum\limits_{t=1,2,\cdots,N_{i'}} p\left(B_{i'j}\right)} \tag{17}$$

According to the reference [12], the major color class similarity for corresponding parts of target A and B is calculated as follow:

$$sim\left(C_{A_{it}}, C_{B_{i'j}|A_{it}}\right) = \min\left\{p_{norm}\left(A_{it}\right), p_{norm}^{[A_{it}]}\left(B_{i'j}\right)\right\} \tag{18}$$

The similarity from part i of target A to part of the candidate target B is calculated as follow:

$$sim\left(A_i, B_{i'}\right) = \sum_{t=1}^{M_i} sim\left(C_{A_{it}}, C_{B_{i'j}|A_{it}}\right) \tag{19}$$

Since the uniqueness of pedestrian objects it is in adjacent two frames, so the symmetry of matching is considered in calculation of similarity. Similarly the similarity which is from the part of candidate target B to the part i of the target A can be gotten.

A symmetrical similarity matching formula finally can be gotten as follows:

$$similarity\left(A_i, B_{i'}\right) = \begin{cases} Sim_{\min}\left(A_i, B_{i'}\right) \\ , Sim_{\min}\left(A_i, B_{i'}\right) < \eta_{discrim} \\ 1 - \dfrac{Sim_{\max}\left(A_i, B_{i'}\right) - Sim_{\min}\left(A_i, B_{i'}\right)}{Sim_{\max}\left(A_i, B_{i'}\right) + Sim_{\min}\left(A_i, B_{i'}\right)} \\ , Sim_{\min}\left(A_i, B_{i'}\right) \geq \eta_{discrim} \end{cases} \tag{20}$$

Where $Sim_{\min}\left(A_i, B_{i'}\right) = \min\left\{sim\left(A_i, B_{i'}\right), sim\left(B_{i'}, A_i\right)\right\}$, $Sim_{\max}\left(A_i, B_{i'}\right) = \max\left\{sim\left(A_i, B_{i'}\right), sim\left(B_{i'}, A_i\right)\right\}$,

$\eta_{discrim}$ is a given difference threshold value.

Due to this paper took into account the possible that local limb occlusion of pedestrian target may exist, Actually it only need to successfully match five parts which are pedestrians body, a thigh, a leg, an upper arm, a forearm respectively. Then it can be considered as the same pedestrian.

3.2. Revise the State Vector of JPDA Algorithm

After completing the similarity matching of the target, the state vector of JPDA algorithm is corrected by the measurement of targets' position in current frame as which is made by the position of targets' centroid that is matched with the one of previous frame. Association matching algorithm is shown in Figure 3.

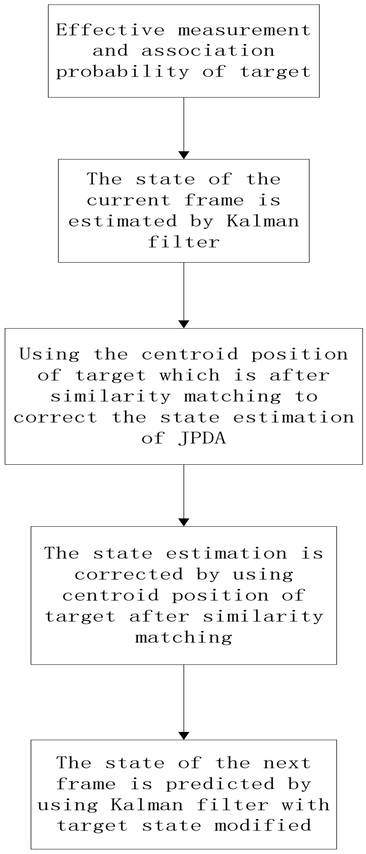

Figure 3. The Improved Association Matching Algorithm

4. Experimental Results and Analysis

The experimental environment is AMD Athlon dual-core processor, 1.7GHz for dominant frequency, 1GB RAM. The software development platform are Matlab 2010b and VC + +6.0 which are used as mixed programming, operating system is Windows XP Professional SP3, PETS 2009 data set[15] is used in this paper, its total number of frames is 795, we only process the prior 300 frames. The detection figure of pedestrian's parts model is shown as Figure 4.

Figure 4. The Detection Figure of Pedestrian's Parts Model

This paper refers to the reference [12], the color clustering distance threshold and the difference threshold are taken the value of 0.05 and 0.4 respectively in similarity matching module of improvement algorithm.

Classic JPDA algorithm will appear the phenomenon of label adhesion and label confusion in experiment. Four frames which are took as an example in this paper are shown in Figure 5 and Figure 6 respectively. It can be seen from Figure 5, two target's label which are labeled No. 7 and No. 8 respectively adhered on a target because the label of No. 8 target is probability associated to No. 7 target. The targets' label appear confusion in two frames of Figure 6, The label of yellow No. 6 target and the label of red No. 1 target are exchanged. It is because that the error probability association which is generated by the using of single target motion information.

(24th Frame) (32th Frame)

Figure 5. Classical JPDA Algorithm Appears Label Adhesion

(104th Frame) (110th Frame)

Figure 6. Classical JPDA Algorithm Appears Label Confusion

The effects of association matching algorithm which is improved are shown in Figure 7 and Figure 8. Two targets whose label are No. 5 and No. 6 in Figure 7 has overcame the problem of adhesion corresponding to the targets whose labels are No. 7 and No. 8 in Figure 5. The labels of two targets in Figure 8 has also solved the problem of confusion corresponding to the targets whose labels are No. 6 and No. 1 in Figure 5.

(24th Frame) (32th Frame)

Figure 7. Improved Association Matching Algorithm

(104th Frame) (110th Frame)

Figure 8. Improved Association Matching Algorithm

The targets' trajectory of 50 images for two algorithms are shown as Figure 9 and Figure 10. It can be seen from Figure 10, the targets which label No. 7 and No. 8 have achieved trajectory separated that they are adhered together in Figure 5, as yellow-green target trajectory in Figure 10. Meanwhile, because of using association matching, the trajectory of target whcih labels No. 6 and yellow didn't regenerate as the red and green discontinuous trajectory in Figure 9.

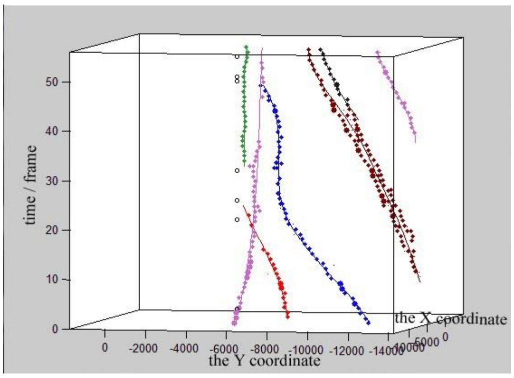

Figure 9. The Trajectory Figure of Classical JPDA Algorithm

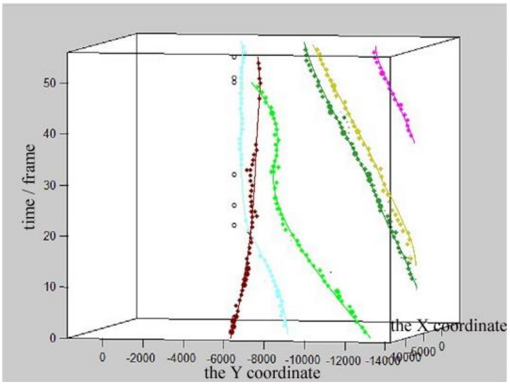

Figure 10. The Trajectory Figure of Improved Association Matching Algorithm

This paper adopts literature [4, 10] indicators, using targets' recall rate (Rcll) and precision (Prn), the number of adhesion and confusion as the indicators to compare the performance of literature [5, 10] with the tracking algorithm in this paper. And it is shown in Table 1.

Table 1. The Performance Comparison Chart of these Algorithms

algorithm	Rcll	Prn	The number of adhesion	The number of confusion
Classical JPDA algorithm[5]	69.5%	77.4%	5	9
DTLE[10]	77.3%	87.2%	1	8
Improved association matching algorithm	76.3%	86.6%	2	3

5. Conclusions

For the defects that the targets' label adhesion and targets' label confusion are prone to emerge in classical JPDA algorithm, a improved tracking algorithm that pedestrian object's major color spectrum histogram (MCSHR) similarity matching method combined with joint probabilistic data association (JPDA) is proposed in this paper. Experimental results show that the algorithm in this paper has achieved the following four improvements compared with the classic JPDA algorithm when the occlusion rate and intensity between pedestrians are not very high: first, to reduce the number of the adhesion for targets' label to the 40% of original; second, the number of confusion for the targets' label decreased about 33%; third, the recall rate of targets is increased by 6.8%, fourth, the tracking accuracy is improved by 9.2%. The parts model of pedestrian targets need to be matched by the major color spectrum similarity in the improved algorithm of this paper. Therefore, with the increasing number of people in the scene, it will make the amount of calculation double and affect the real-time of this algorithm. This problem will be studied as a key issue in the next step.

Acknowledgement

This work was financially supported by the Natural Science Foundation of China (No.61272077), Natural Foundation of Jiangxi Province (No. 20114BAB20142BAB207012), and Key Laboratory Open Foundation (No.TX201204 005). The authors thank the anonymous reviewers and the editors for their valuable comments to improve the presentation of the paper.

References

[1] W. Hu, T. Tan, L. Wang and S. Maybank, "A survey on visual surveillance of object motion and behaviors", IEEE Transactions on Systems, Man, and Cybernetics, PART C: Applications and Reviews, vol. 34, no. 3, **(2004)** August.

[2] M. BERTALMIO, G. SAPIRO and V. CASELLES, "Image inpainting", Proceedings of the 27th Annual Conference on Computer Graphics, **(2000)** July 23-28, New Orleans, Louisiana, USA, New York, USA: ACM Press, **(2000)**, pp. 417-424.

[3] T. CHAN and J. SHEN, "Mathematical models for local deterministic inpaintings", Los Angeles: Department of Mathematics, UCLA, **(2000)**.

[4] T. Wei, Y.-P. LUO, Cheng-ying, "New method of multiple object tracking under circumstance of occlusion", Computer Engineering and Applications, vol. 43, no. 13, **(2007)**, pp. 33-36.

[5] T. E. Fortmann, Y. Bar-Shalom and M. Scheffe, "Multi-target tracking using joint probabilistic data association", IEEE Conf. on Decision and Control, **(1980)**.

[6] J. Moonyong, J. Kiseon and P. Dong Sun, "Efficient Histogram based Occluded Object Segmentation in Tracking System", Proceedings of the 2012 international conference on image processing computer vision & pattern recognition, vol. 819.

[7] S. Tang, M. Andriluka and B. Schiele, "Detection and Tracking of Occluded People", In BMVC, **(2012)**.

[8] A. Andriyenko, S. Roth and K. Schindler, "An Analytical Formulation of Global Occlusion Reasoning for Multi-Target Tracking", In ICCV, **(2011)**.

[9] A. Andriyenko, K. Schindler and S. Roth, "Discrete-Continuous Optimization for Multi-Target Tracking", In CVPR, **(2012)**.

[10] A. Milan, K. Schindler and S. Roth, "Detection- and Trajectory-Level Exclusion in Multiple Object Tracking[C]", In CVPR, **(2013)**.

[11] C. MADDEN, E. D. CHENG and M. PICCARDI, "Tracking people across disjoint camera views by an illumination -tolerant appearance representation", Machine Vision and Applications, vol. 18, vol. 3, **(2007)**, pp. 233-247.

[12] LI Bo, LI Guo-hui and TU Dan, "Human matching approach based on color spatial distribution model across non-overlapping camera views", Application Research of Computers, vol. 29, Issue 2, **(2012)**.

[13] D. Ramanan, "Learning to parse images of articulated bodies", IN NIPS, **(2006)**.

[14] PETS: Eleventh IEEE International Workshop on Performance Evaluation of Tracking and Surveillance, 2009.Available:http://pets2009.net/, **(2009)**.