Multi-Targets Tracking Based on Bipartite Graph Matching

Jinqin Zhong[1], *Jieqing Tan*[2], *Yingying Li*[1], *Lichuan Gu*[3],
Guolong Chen[4]

[1] *School of Computer and Information, Hefei University of Technology, Hefei 230009 Anhui, China*
[2] *School of Mathematics, Hefei University of Technology, Hefei 230009 Anhui, China*
[3] *School of Computer and Information, Anhui Agriculture University, Hefei 230036 Anhui, China*
[4] *Suzhou University, Suzhou 234000 Anhui, China*
Emails: jinqinzhong@163.com   jieqingtan@126.com   gulichuan@ahau.edu.cn

Abstract: Multi-target tracking is a challenge due to the variable number of targets and the frequent interaction between targets in complex dynamic environments. This paper presents a multi-target tracking algorithm based on bipartite graph matching. Unlike previous approaches, the method proposed considers the target tracking as a bipartite graph matching problem where the nodes of the bipartite graph correspond to the targets in two neighboring frames, and the edges correspond to the degree of the similarity measure between the targets in different frames. Finding correspondence between the targets is formulated as a maximal matching problem which can be solved by the dynamic Hungarian algorithm. Then, merging and splitting of the targets detection is proposed, the candidate occlusion region is predicted according to the overlapping between the bounding boxes of the interacting targets to handle the mutual occlusion problem. The extensive experimental results show that the algorithm proposed can achieve good performance on dynamic target interactions compared to state-of-the-art methods.

Keywords: Multi-target tracking, bipartite graph optimal matching, target interaction, merging and splitting.

1. Introduction

Multi-target tracking is a fundamental problem in computer vision and has broad applications in many areas, including intelligent surveillance, visual human-computer interfaces, autonomous robotics, augmented reality and video compression. It can be viewed as locating targets of interest, inferring their

trajectories and maintaining their identities through a video sequence. Despite the huge amount of excellent research in the field [1-4], developing an efficient and robust target detection and correspondence method is still challenging, especially in crowded and unstructured environments where targets can appear, disappear, be wrongly detected and frequently interact in the image frame.

Multi-target tracking can be viewed as a correspondence problem in sequential image frames. Traditional feature-based tracking methods, such as those based on colour [5], salient points [6], or motion blobs [7], do not have a discriminative model for target detection and thus suffer from poor detection of targets. Over the past two decades, the kinds of data association-based tracking methods have become popular because of the significant improvement in target detection techniques, such as [8-12]. Multiple Hypothesis Tracking method (MHT) [8] and Joint Probabilistic Data Association Filters (JPDAF) [9] are two widely used methods for data association. However, when the multiple targets and echo increase, JPDAF will result in the fact that the computational complexity of the data association increases exponentially, and MHT will waste much time in the large number of "iteration" processes. In recent years the Random Finite Set (RFS) approach [10] to multi-target tracking in a radar system has received considerable attention. Compared with the traditional association-based techniques, the difficulty caused by the data association is avoided in RFS formulation. One of its representatives is the Probability Hypothesis Density filter (PHD), proposed by Mahler [11], which aims to propagate recursively the first order moment or the intensity function associated with the multi-target posterior density, to obtain the state of target estimation. Although the PHD filter based tracking approaches avoid the data association problem and are widely used in visual tracking [13, 14], they fail to track the individual targets when mutual occlusion occurs in the interacting targets.

The graph-model based approach is qualitatively different from most previous approaches, formulating the correspondence task as a maximal graph matching problem. Here the nodes of the graphs correspond to local features of the image and graph matching is to find a correspondence between the nodes of the two graphs, such that they "look most similar" when the nodes are labeled according to such a correspondence. The graph matching model is very efficient and robust to feature point matching [15, 16], image jigsaw puzzle [17, 18], and video object segmentation [19, 20]. We adopt the maximal graph matching model to solve a multi-target tracking problem in this paper.

In this work we provide a method for efficiently incorporating the correspondences into the commonly used integer programming formulations of graph matching. Therefore, the multi-target tracking problem is actually considered as the maximal weighted bipartite matching problem, which can be solved by the dynamic Hungarian algorithm. Besides, simplified merging and splitting of targets detection is proposed, the candidate occlusion region is predicted according to the overlapping between the bounding boxes of the interacting targets, which can handle the occlusion problem in complicated dynamic environments.

The remainder of this paper is organized as follows. Section 2 presents a new multi-target tracking algorithm based on bipartite graph optimal matching, which is the main contribution of this paper, and Section 3 presents the experiments performed and a qualitative and quantitative assessment of the tracker performance, followed by concluding remarks in Section 4.

2. Multi-target tracking based on bipartite graph matching

For the target tracking we consider a simple bipartite matching method to associate with the targets in two neighboring frames. We generate a weighted complete bipartite graph where a node represents a target and the weight between two nodes is the similarity between two targets in neighboring frames (Fig. 1). Thus the multi-target tracking problem is formulated as the weighted bipartite matching problem; we can construct the maximal weight matching model to determine the best matching results.
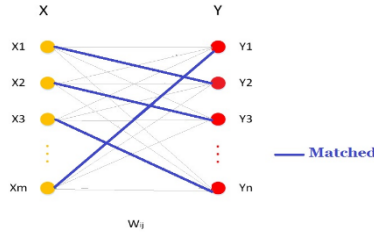

Fig. 1. Bipartite graph and matching diagram

2.1. Determination of weight

In the weighted bipartite graph matching problem, the similarity between two targets must be determined to get the weight ω_{ij} of the edge, then we need to describe the node first. We denote a node by $N_i=\{p_i,\ s_i,\ f_i\}$, in which p_i is the position, s_i is the size, f_i are the node's visual features. To build a discriminative appearance model, we adopt a colour HSV histogram and CS-LBP (Center Symmetric Local Binary Pattern) histogram to describe a tracked target.

Let $S=\{S_{ij}\}_{mn}$ denotes the similarity matrix. The similarity of a pair of nodes is computed by

(1) $\qquad s_{ij}(N_i, N_j) = s_{\mathrm{pos}}(N_i, N_j)\, s_{\mathrm{size}}(N_i, N_j)\, s_{\mathrm{feature}}(N_i, N_j).$

The three elementary similarities are:

(2) $\qquad s_{\mathrm{pos}}(N_i, N_j) = \gamma_{\mathrm{pos}} \exp\left[\dfrac{(x_i - x_j)^2}{\sigma_x^2}\right] \exp\left[\dfrac{(y_i - y_j)^2}{\sigma_y^2}\right],$

(3) $\qquad s_{\mathrm{size}}(N_i, N_j) = \gamma_{\mathrm{size}} \exp\left[\dfrac{(s_i - s_j)^2}{\sigma_x^2}\right],$

(4) $\qquad s_{\mathrm{feature}}(N_i, N_j) = B_{\mathrm{HSV}}(N_i, N_j) B_{\mathrm{CS\text{-}LBP}}(N_i, N_j),$

Where $B(\bullet)$ is the Bhattacharyya distance between two histograms γ_{pos}, and γ_{size} are normalization factors. Then we get the edge weight $\omega_{ij} = s_{ij}(N_i, N_j)$.

2.2. The bipartite graph optimal matching

In this section our goal is to find the maximal weighted matching in bipartite graphs. In a bipartite graph $G = (X, Y, E)$, when $m \neq n$ (propose $m<n$), the maximal weighted matching of a bipartite graph is complete matching, this problem can be solved by the following integer programming model:

$$(5) \qquad \max g = \sum_{i=1}^{m} \sum_{j=1}^{n} \omega_{ij} x_{ij},$$

$$(6) \qquad \text{subject to } \sum_{j=1}^{n} x_{ij} = 1, \quad i = 1, 2, \ldots, m,$$

$$(7) \qquad \sum_{i=1}^{m} x_{ij} \leq 1, \quad j = 1, 2, \ldots, n,$$

$$(8) \qquad x_{ij} = 0 \text{ or } 1, \quad i = 1, 2, \ldots, m; \quad j = 1, 2, \ldots, n.$$

In the above model, (5) is the objective function, the solution of which is the maximal weighted matching M, typically subject to the constraint (6) and (7). The solution of this optimal model can be found using the dynamic Hungarian algorithm.

2.3. Merging and splitting processing

From the solution of the optimal model, the existing tracks and foreground measures are classified into three parts: non-matched existing track, non-matched measure, and matched track and measure.

For the non-matched existing tracks we must judge whether the existing track is merged by another measure or is missed. Merging might occur due to a non-matched existing track overlapped by a measure. For two adjacent frames, the target position change is very small, as soon as the targets touch each other at time $k+1$, a large bounding box containing all the merged targets will be created and it has large overlapping areas with merged targets at time k, as shown in Fig. 2b. If merging occurs, a new group is generated, the feature of the occluded track will be added into the group, after that, the entire group will be tracked as one target, the feature of the occluded track and the measure are not updated. Otherwise the existing track is missed and it increases the deactivation count value. If the value is bigger than a user-defined count, the tracked target is deleted from the tracked target list.

For those non-matched measures, a splitting detection algorithm is used to decide whether the measure is split from an existing group track or is an added new target. Similar to the merging procedure, splitting is detected due to a non-matched measure overlapped with a track (Fig. 2c). If splitting happens, the feature information of the occluded target in the group will be used for correspondence. Otherwise the measure is a new target, the attribute of it is computed, and the deactivation count value is set to zero.

81

(a) Tracking two people

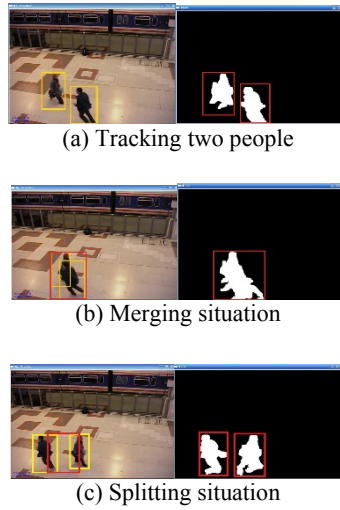(b) Merging situation

(c) Splitting situation

Fig. 2. The characteristics of a detection region in merging and splitting situation: detected and tracked target region with no marginal (a); merging situation and overlapping areas (b); splitting situation (c)

3. Experimental results

To validate the efficiency of the algorithm, the proposed multi-target tracking based on bipartite graph matching method has been applied to the video data sets coming from PETS2006, AVSS2007 and PETS2009 S2.L2. All the videos are captured by a static camera.

We compare our proposed tracking algorithm to GM-PHD (Gaussian Mixture Probability Hypothesis Density filter) and CP1 (single-template coupling tracker). GM-PHD tracker [13] involves modeling of the respective collections of targets and measurements as random finite sets and applying the PHD recursion to propagate the posterior intensity. CP1 tracker [12], in which the problems of object detection and data association are expressed by a single objective function and through dual decomposition, the objective function can be optimized iteratively with off-the-shelf efficient algorithms for each sub problem.

3.1. Qualitative analysis

Figs 3-5 show the tracking results of GM-PHD (the first row in Figs 3-5), of the CP1 (the second row in Figs 3-5), and of the proposed tracker (the third row in Figs 3-5). All the videos are concerned with the interacting target in complicated dynamic environments. From the experimental results, by using the bipartite graph optimal matching method, our tracker can perform robustly to interacting the target in complicated environments, especially to frequent dynamic interactions between the targets.
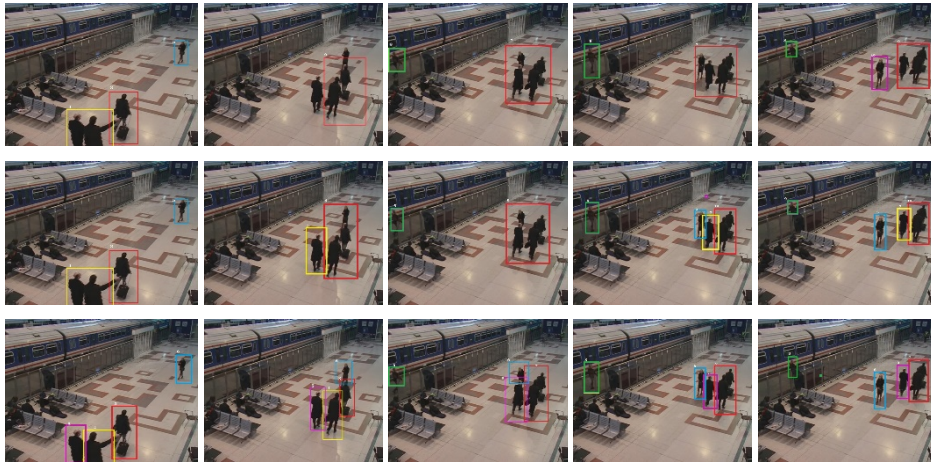
Fig. 3. Tracking results comparison for "PETS2006": first row – tracking results of GM-PHD; second row – tracking results of CP1; third row – tracking results of our tracker
(Frame: 103, 131, 142, 148, 162)

In Fig. 3 (PETS2006), the first three people come nearer from frame 103 to 130. Then one of them is parted occluded by another one and merges with the fourth target from frame 131 to 147. Finally the fourth target splits out while the other two targets remain merging from frame 148. The main challenge is that the partial occlusions occurred in the above four interacting targets which have similar appearance. The GM-PHD cannot handle this challenging issue while CP1 can tracker better and our method tracks well (especially from frame103 to 130). Moreover, as shown in frame 162, target 3 is totally occluded by target 4, CP1 and our method track the two targets as a new target and this will be explored in our future work.

Fig. 4. Tracking results comparison for "AVSS2007": first row – tracking results of GM-PHD; second row – tracking results of CP1; third row – tracking results of our tracker
(Frame: 362, 402, 431, 528, 580)

In Fig. 4 (AVSS2007) we further evaluate the robustness of three trackers to handle vehicles interactions and hybrid difficulties including illumination change. As shown in frame 362, when multiple vehicles occlude each other, they are often detected as a single vehicle by GM-PHD and CP1, while our tracker can successfully track all the interaction vehicles. In the video, from frame 431 to frame 580, the targets go through a dark region to a bright region, the illumination has a great change. GM-PHD mis-tracks the existing vehicle as a newborn one, while CP1 and our method can track the targets correctly.

Fig. 5. Tracking results comparison for "PETS2009 S2.L2": first row – tracking results of GM-PHD; second row – tracking results of CP1; third row – tracking results of our tracker
(Frame: 225, 234, 246, 282, 309)

In Fig. 5 (PETS2009 S2.L2), the pedestrians are walking across an intersection in various directions at variable speed and the number of people varies from frame 225 to frame 309. Most of them move closely, sometimes occlude between targets, which are challenging issues for frequent dynamic interactions handling. For example, in the left region of frame 225, there are four pedestrians walking closely, GM-PHD tracks them as a target and CP1 tracks them as two targets, while our method can track correctly. Up to frame 234, two of the four pedestrians have parted occlusion; GM-PHD and CP1 lose the occlusion target, while our method can track the occlusion target successfully.

3.2. Quantitative analysis

For evaluation, we compute the standard CLEAR multiple target tracking performance metrics, i.e., MOTA↑ (Multiple Target Tracking Accuracy) and MOTP↓ (Multiple Target Tracking Precision). To better assess the quality, we additionally report the MT↑(Mostly tracked trajectories), ML↓ (Mostly lost trajectories), FM↓ (Fragments) and ID↓ (ID switches). Those are implemented a program to compute these metrics automatically. ↑ means higher is better, as opposed to ↓.

Table 1. Quantitative comparison with the state-of-the-art methods

Dataset	Algorithm	MOTP[m]	MOTA	MT	ML	FM	IDS
PETS2006	GM-PHD	0.168	71.6%	72.6%	7.0%	241	7
	CP1	0.102	92.5%	78.3%	6.4%	21	4
	OUR	0.102	98.3%	86.7%	0.7%	6	1
AVSS2007	GM-PHD	0.201	55.5%	50.2%	9.9%	31	17
	CP1	0.128	82.2%	58.4%	8.0%	44	3
	OUR	0.133	90.4%	68.0%	7.2%	14	0
PETS2009S2.L2	GM-PHD	0.231	49.0%	57.0%	14.9%	220	46
	CP1	0.211	59.7%	60.6%	9.4%	202	10
	OUR	0.215	67.5%	75.2%	5.1%	172	9

To show the tracking performance of the proposed algorithm, a fair quantitative comparison of multiple target tracking methods is challenging. The results are listed in Table 1. As it can be seen from the overall scores on the above three sequences, our proposed algorithm outperforms GM-PHD and CP1 trackers at complex scenarios. For PETS2009 S2.L2, in situations where people exhibit challenging poses, suddenly change the motion directions, intersection occurs frequently, GM-PHD and CP1 are often not able to link the true positive detections correctly or drifting starts after several frames of missed detections. However, our approach is able to handle such complex poses and frequent intersections by using colour and texture features and merging or splitting processing. Considering the high number of identity switches, GM-PHD obviously suffers from the missing colour information, especially in illumination changing. For example, in AVSS2007, when the vehicle enters or is out of the dark region, GM-PHD loses the target and re-tracks it as a newborn target, while CP1 and our tracker are able to track well. Regarding the precision metrics, our tracker achieves almost the same performance as CP1.

4. Conclusions

We have proposed a bipartite optimal matching algorithm with highly discriminative attributes of nodes to simultaneously track multiple targets in complex environments. In the method proposed, the targets of two successive frames are considered as the node of bipartite, targets association is regarded as the maximal weighted bipartite matching problem which can be formulated as an integer programming model and solved by the dynamic Hungarian algorithm. We also proposed efficient merging and splitting processing procedure to deal with the mutual occlusion problem in multiple interacting targets. The experimental results demonstrate that the algorithm proposed enables tracking of targets in complex scenes with occlusions and varying interaction behaviours.

In the future work we consider investigating an online training appearance model for interaction handling. Furthermore, we would like to cope with the noise problem. The proposed method has the ability to filter those noises close to the survival targets. However, as the noises are far away from the moving targets, they may be tracked as newborn targets. All these issues are worth further studying.

References

1. L i, Z h a n g, Y u a n L i, R. N e v a t i a. Global Data Association for Multi-Object Tracking Using Network Flows. – In: IEEE Conference on Computer Vision and Pattern Recognition, Anchorage, AK, 2008, 1-8.
2. B i b b y, C., I. R e i d. Real-Time Tracking of Multiple Occluding Targets Using Level Sets. – In: IEEE Conference on Computer Vision and Pattern Recognition, San Francisco, CA, USA, 2010, 1307-1314.
3. H u a n g, C h a n g, Y u a n L i, R. N e v a t i a. Multiple Target Tracking by Learning-Based Hierarchical Association of Detection Responses. – IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. **35**, 2013, No 4, 898-910.
4. A n d r i y e n k o, A., K. S c h i n d l e r. Multi-Target Tracking by Continuous Energy Minimization. – In: IEEE Conference on Computer Vision and Pattern Recognition, Providence, RI, 2011, 1265-1272.
5. C o m a n i c i u, D., V. R a m e s h, P. M e e r. The Variable Bandwidth Mean Shift and Data-Driven Scale Selection. – In: IEEE International Conference on Computer Vision,Vancouver, BC, 2001, 438-445.
6. T o m a s i, C., T. K a n a d e. Detection and Tracking of Point Features. Technical Report CMUCS-91-132, Carnegie Mellon University, Pittsburgh, PA, 1991.
7. Z h a o, T., R. N e v a t i a. Tracking Multiple Humans in Complex Situation. – IEEE Transactions on Analysis and Machine Intelligence, Vol. **26**, 2004, No 9, 1208-1221.
8. R e i d, D. An Algorithm for Tracking Multiple Targets. – IEEE Transactions on Automatic Control, Vol. **24**, 1979, No 6, 8430-854.
9. B a r s h a l o m, Y., T. F o r t m a n n. Tracking and Data Association. San Diego, Academic Press,1988.
10. G o o d m a n, I. R., R. M a h l e r, H. T. N g u y e n. Mathematics of Data Fusion. Norwell, MA, Kluwer Academic Press, 1997.
11. M a h l e r, R. P. S., M. L o c k h e e d. Multi Target Bayes Filter via First-Order Multi Target Moments. – IEEE Transactions on Aerospace and Electronic Systems, Vol. **39**, 2003, No 4, 1152-1178.
12. W u, Z h e n g, A. T h a n g a l i, S. S c l a r o f f, M. B e t k e. Coupling Detection and Data Association for Multiple Target Tracking. – In: IEEE Conference on Computer Vision and Pattern Recognition, Providence, RI, 2012, 1948-1955.
13. V o, B a-N g u, W i n g-K i n M a. The Gaussian Mixture Probability Hypothesis Density Filter. – IEEE Transaction on Signal Processing, Vol. **54**, 2006, No 11, 4091-4104.
14. W a n g, Y. D., J. K. W u, A. A. K a s s i m, W. M. H u a n g. Date-Driven Probability Hypothesis Density Filter for Visual Tracking. – IEEE Transaction on Circuits and Systems for Video Technology, Vol. **18**, 2008, No 8, 1085-1095.
15. C a e t a n o, T. S., J. J. M c a u l e y, L i C h e n g et al. Learning Graph Matching. – IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. **31**, 2009, No 6, 1048-1058.
16. D u c h e n n e, O., F. B a c h, K w e o n I n-S o, J. P o n c e. A Tensor-Based Algorithm for High-Order Graph Matching. – IEEE Transaction on Pattern Analysis and Machine Intelligence, Vol. **33,** 2011, No 12, 2383-2395.
17. T a e g, S a n g C h o, S. A v i d a n, W. T. F r e e m a n. A Probabilistic Image Jigsaw Puzzle Solver. – In: IEEE Conference on Computer Vision and Pattern Recognition, Providence, RI, 2011, 183-190.

18. Y a n g, X i n g w e i, N. A d l u r u, L. J. L a t e c h i. Particle Filter with State Permutations for Solving Image Jigsaw Puzzles. – In: IEEE Conference on Computer Vision and Pattern Recognition, Providence, RI, 2011, 2873-2880.
19. G r u n d m a n n, M., V. D w a t r a, H a n M e i, I. E s s a. Efficient Hierarchical Graph-Based Video Segmentation. – In: IEEE Conference on Computer Vision and Pattern Recognition, San Francisco, CA, USA, 2010, 2141-2148.
20. H u, N a n, R. M. R u s t a m o v, L. G u i b a s. Graph Matching with Anchor Nodes: A Learning Approach. – In: IEEE Conference on Computer Vision and Pattern Recognition, Portland, OR, 2013, 2096-2104.