MULTI-TARGET TRACKING VIA PARATACTIC-SERIAL TRACKLET GRAPH

Hao Sheng[1], *Jiahui Chen*[1], *Jiangjian Xiao*[2], *Chao Li*[1], *Zhang Xiong*[1]

[1]School of Computer Science and Engineering, Beihang University, Beijing 100191, P.R.China
[2] Ningbo Institute of Industrial Technology, Chinese Academy of Sciences, Ningbo 315201, P.R.China

ABSTRACT

Many recent advances in multi-target tracking have grown concern over latent corresponding relation among observations, *e.g.* social connection. To handle long-term occlusion within group and tracking failure caused by interaction of targets, various correlations among tracklets need to be exploited. In this work, we propose a paratactic-serial tracklet graph(PSTG)-based framework for multi-target tracking. Contrary to recent approaches, we define a novel PSTG-based method describing the correlation among all tracklets in spatio-temporal domain to model the mutual influence among trajectories. Paratactic-tracklet graph extends the potential relationship among tracklets, and serial-tracklet graph enhances the integrity and continuity of trajectories. In addition, we design a PSTG energy function to make the trajectory estimation more accurate. Moreover, multi-label optimization is presented to embody constraints on group, integrity and spatio-temporal domain. We demonstrate the validity of our approach on several public datasets and our dataset, and achieve very competitive results by quantitative evaluation.

Index Terms— Paratactic-Serial Tracklet Graph, Social Group, Multi-label, Multi-target Tracking, Correlation

1. INTRODUCTION

Concurrently keeping track of multiple targets from a video sequence remains a challenging task in computer vision. Recovering the spatio-temporal trajectories accurately is one of the key technologies in many applications. In recent years, state-of-the-art algorithms for multi-target tracking are no match for human abilities of tracking, both in terms of precision and accuracy, though it has made great progress in recent years. After detections are given, appearance, direction, social group, spatio-temporal continuity and interaction among targets should be taken into account for tracking.

In the last few years, tracking-by-detection framework has grown more and more popular[1][2][3][4]. Berclaz *et al.* [5] capture the dependencies between video through network using Markovian dependencies. Pirsiavash *et al.* [6] introduce a greedy optimization scheme globally by inserting target hypotheses for tracking a variable number of objects. However, the targets in the scene need global data association. Min-cost network flow algorithms [7][8] are used to reformulate the task as a network flow problem, which can be solved in polynomial time. To avoid ID switches, an online CRF model is proposed to learn appearance features that discriminate among close targets for tracklet association [9]. Similarly, a more effective dynamic model leveraging nearby target positions is proposed in [10]. Label cost based tracker [11] addresses both data association and trajectory estimation by minimizing energy.However, the correlation among targets which is important in visual tracking is not modeled into the framework. To overcome the drawback, Milan *et al.* [12] propose a mixed discrete-continuous CRF to fix the weakness. The above approaches still have some limitations.

In this paper, we discuss a paratactic-serial tracklet graph(PSTG)-based multi-label optimization method for multi-target tracking. Compared with the previous works, PSTG optimization is proposed to exploit various correlations among tracklets and minimize the energy of PSTG. In PSTG optimization algorithm, the correlation among tracklets can be paratactic or serial. The correlation among tracklets embodies group, integrity and spatio-temporal constraints. Group constraint mainly reflects the group relations of moving targets, which means the targets in same group should have similar trajectories. Using our paratactic-tracklet graph, such group relationship among tracklets can be effectively recovered, which can partly settle detection errors or trajectory fragments caused by occlusion. Integrity constraint means any trajectory that is continuous and has unique start and end points, which can be implemented by our serial-tracklet graph. Spatio-temporal constraint shows the relationship of tracklets in temporal and spatial domain. To demonstrate our PSTG optimization algorithm, we run extensive experiments on various datasets and achieve very competitive results.

2. ALGORITHM OVERVIEW

To solve the problems – interaction of targets, integrity of trajectories and spatio-temporal constraint, we propose a novel framework called PSTG-based tracking, shown in Fig.1. In this approach, paratactic-tracklet graph is generated to model the interaction among targets and serial-tracklet graph reflects the integrity of tracklets. Both two graphs embody spatio-temporal constraints.

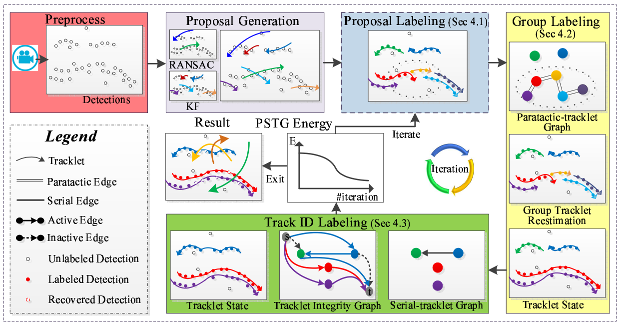

Fig. 1. PSTG-based Tracking Framework. Each iteration consists mainly of 3 steps : Proposal labeling, group labeling and track ID labeling. Proposal labeling can be any algorithm processing tracklet proposals.

In Sec.3, graph theory for inter-tracklet analysis is introduced, and details of the PSTG-based framework are described in Sec.4. Our framework is verified in Sec.5.

3. GRAPH THEORY FOR INTER-TRACKLET ANALYSIS

3.1. Paratactic-tracklet graph(PTG)

Paratactic-tracklet. Two tracklets which have similar motion model in spatio-temporal neighbor are defined as paratactic tracklets. Two paratactic tracklets cannot represent the same target. Paratactic-tracklet graph is used to illustrate this kind of relation between two paratactic tracklets.

We generate a undirected graph $G_{pa} = (V_{pa}, E_{pa})$, shown in Fig.2(top). V_{pa} represents tracklet set $\mathcal{T}$. The edge set E_{pa} is defined as

$$\begin{aligned} E_{pa} = &\{(\mathcal{T}_i, \mathcal{T}_j) | s(\mathcal{T}_i) \cap s(\mathcal{T}_j) \neq \emptyset, \\ &\|p_t(\mathcal{T}_i) - p_t(\mathcal{T}_j)\| > \alpha, cos(\mathcal{T}_i, \mathcal{T}_j) > 0\} \end{aligned} \tag{1}$$

where $s(\mathcal{T}), p_t(\mathcal{T}), \alpha, \cos(\mathcal{T}_1, \mathcal{T}_2)$ are respectively the time span of tracklet $\mathcal{T}$, the position of tracklet $\mathcal{T}$ at time t, the threshold of distance and the cosine of angle between $\mathcal{T}_1$ and $\mathcal{T}_2$.

3.2. Serial-tracklet graph(STG)

Serial-tracklet. Two tracklets which may represent two trajectory fragments of a certain target in different periods are defined as serial tracklets. Serial-tracklet graph is used to show this kind of relation between two serial tracklets.

We generate a directed graph $G_{ser} = (V_{ser}, E_{ser})$, shown in Fig.2(bottom). V_{ser} represents tracklet set $\mathcal{T}$. Edge set E_{ser} is defined as

$$\begin{aligned} E_{ser} = &\{(\mathcal{T}_i, \mathcal{T}_j) | s(\mathcal{T}_i) \cap s(\mathcal{T}_j) = \emptyset, \\ &\|p_{en}(\mathcal{T}_i) - p_{st}(\mathcal{T}_j)\| < \alpha, \|v\| < \beta\} \end{aligned} \tag{2}$$

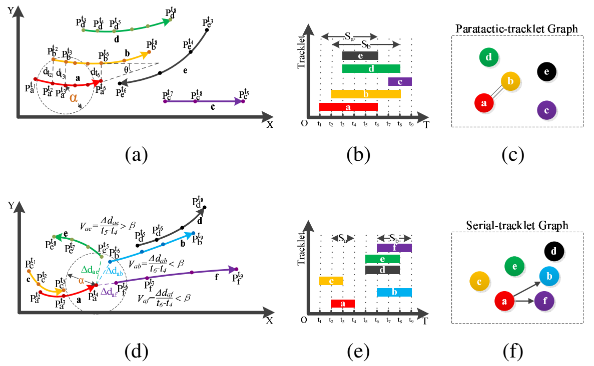

Fig. 2. PTG and STG. (a)(d) illustrates the constraints in spatio domain and (b)(e) shows the constraints in temporal domain. (c)(f) is the PTG or STG of tracklets.

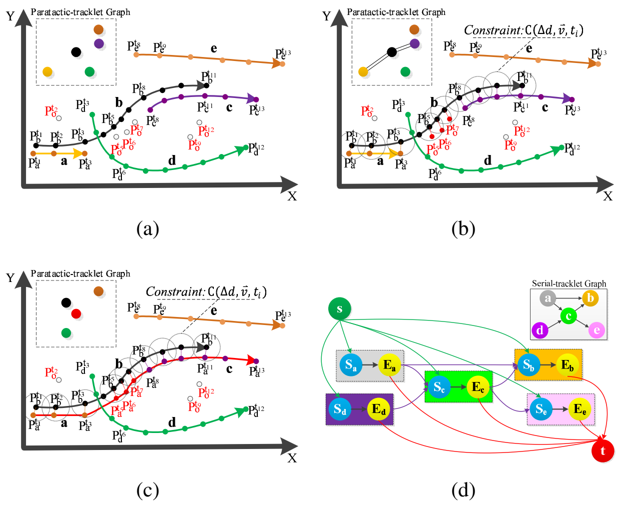

Fig. 3. Group Tracklet Reestimation & Tracklet Integrity Graph. (a) is tracklet state before reestimation. Outlier points around are considered in (b). (c) is the reestimation result.(d) is Tracklet Integrity Graph.

where $s(\mathcal{T}), p_{st}(\mathcal{T}), p_{en}(\mathcal{T}), \alpha, v, \beta$ are respectively the time span of tracklet $\mathcal{T}$, the start point position of tracklet $\mathcal{T}$, the end point position of tracklet $\mathcal{T}$, the threshold of distance, the estimated velocity of the target over two tracklets' gap and the threshold of velocity.

3.3. Tracklet group

The integrity of correlation information among the tracklets has a great influence on the quality of tracking result[10][13][14]. Tracklet group is introduced to improve correlation information and solve the problem of locating individual targets position in the case of missing measurements.

Correlation information is usually formulated based on only tracklets with spatio-temporal proximity in previous works. To gain more correlation information to recover tracklets, we extend tracklets' correlation through paratactic-tracklet graph. With tracklet graph, we enlarge propagation range of tracklet correlative information in time dimension. Fig.3((a), (b) and (c)) illustrates the main processes of group tracklet reestimation.

3.4. Trajectory integrity

Trajectory integrity is proposed to gain longer trajectories. We generate a graph $G_{tr} = (V_{tr}, E_{tr})$, when serial-tracklet graph of $\mathcal{T}$ is available. V_{tr} in addition to tracklet set $\mathcal{T}$, also contains two virtual nodes – source node s and target node t. Edge set E_{tr} is defined as

$$E_{tr} = \{(i, j) | i \in V, j \in follow(V_i)\} \tag{3}$$

where $follow(V_i)$ is the node set in which tracklet $\mathcal{T}$ may be the subsequent tracklet of V_i.

In Fig.3(d), tracklet integrity graph is shown. The edges of the graph can be divided into three categories: s connected edge set E_s(green), t connected edge set E_t(red) and all the other edges set E_c(purple).

4. MULTI-LABEL OPTIMIZATION FOR MULTI-TARGET TRACKING

4.1. Proposal labeling

As shown in Fig.1, proposal set $\mathcal{H}$ need be generated in proposal generation. In our experiment, the proposals fall into two categories – proposals generated through RANdom SAmple Consensus (RANSAC) and proposals generated through Kalman Filter (KF). The former are tracklets considering spatio-temporal proximity, and the latter are tracklets considering motion model.

The object of proposal labeling is to map elements in detection set D with elements in proposal set $\mathcal{H}$. Proposal energy which is used to measure the performance of the proposal labeling $\mathcal{L}_{pr}$, is defined as

$$\mathcal{E}_{pr} = \mathcal{E}_{det} + \mathcal{E}_{tra} \tag{4}$$

where $\mathcal{E}_{det}$ is defined as

$$\mathcal{E}_{det}(h_l, d_t^i) = \begin{cases} s_{d_t^i} \cdot \|h_l(t) - p_t^i\|^2 & l \in N_h \\ s_{d_t^i} \cdot O & l = 0 \end{cases} \tag{5}$$

where $s_{d_t^i}, O$ are respectively the detection confidence of detection d_t^i and the constant cost of the outlier point.

And $\mathcal{E}_{tra}$ is defined as

$$\mathcal{E}_{tra}(h_l) = \gamma + score(h_l) \tag{6}$$

where γ is the punishment for the number of trajectories, and $score(h_l)$ is the goodness of proposal $\mathcal{H}_l$ including punishment for occlusion , persistence, speed and length.

Proposal energy $\mathcal{E}_{pr}$ is minimized to obtain a suitable proposal labeling result $\mathcal{L}_{pr}$. To apply fitting method to these detections in different groups, we can gain tracklet set $\mathcal{T}_{pr}$.

4.2. Group labeling

The object of group labeling is to map elements in $\mathcal{T}_{pr}$ with elements in group set $\mathcal{K}$. Paratactic-tracklet graph is generated with the information of $\mathcal{T}_{pr}$. Then group set $\mathcal{K}$ and number set M are calculated by the method described in Sec.3.3.

In the process of trajectory reestimation, two strong constraints are considered: the number of targets in a certain spatio-temporal area and motion similarity. In addition, the outlier points in proposal labelling are taken into. The trajectory reestimation in a group is shown as Alg.1, then tracklet set $\mathcal{T}_{gr}$ is gained.

Algorithm 1 Group Tracklet Reestimation

Require: Target Number M_i, Detection Set D_i, Start Time T_s, End Time T_e
Ensure: $\mathcal{T}$
$p \leftarrow initial\,position, list \leftarrow \emptyset, t \leftarrow T_s$
while $t \leq T_e$ **do**
  $det \leftarrow getDetection(t)$
  mapping target with det
  if no target maps a suitable det **then**
    Ignore this frame
  else if Not all target map a suitable det **then**
    for i has a suitable det **do**
      $p^i \leftarrow det$
      $list^i \leftarrow list^i + \{(t, p^i)\}$
    end for
    $s \leftarrow average\,speed$
    for i has no suitable det **do**
      $p^i \leftarrow p^i + s$
      $list^i \leftarrow list^i + \{(t, p^i)\}$
    end for
  else
    for $i = 1 \rightarrow M_i$ **do**
      $p^i \leftarrow det$
      $list^i \leftarrow list^i + \{(t, p^i)\}$
    end for
  end if
  $t \leftarrow t + 1$
end while
$\mathcal{T} \leftarrow BSpline(list^i)$

Alg.1 initializes the states of all the targets firstly. Then trajectories are recovered in all three situations : (1) No suitable detection are found. (2) Partial detections are found. (3) All detections are found.

4.3. Track ID labeling

The object of track ID labeling is to map elements in $\mathcal{T}_{gr}$ with elements in track ID set $\mathcal{T}_{in}$. Serial-tracklet graph $G_{in} = (V_{in}, E_{in})$ is generated with the information of $\mathcal{T}_{gr}$. w_{st}^i and w_{en}^i are defined as

$$w_{st}^{(i)} = \min_{p \in P_{ex}} \|p_{st}^{(i)} - p\|, w_{en}^{(i)} = \min_{p \in P_{ex}} \|p_{en}^{(i)} - p\| \tag{7}$$

where P_{ex} is the set of exit positions in tracking area. $p_{st}^{(i)}$ and $p_{en}^{(i)}$ are respectively the positions of start point and end point of tracklet $\mathcal{T}^i$. $w_{co}^{(i,j)}$ in serial-tracklet graph is defined as

$$w_{co}^{(i,j)} = \begin{cases} \|p_{st}^{(j)} - p_{en}^{(i)}\| & (i, j) \in E_{ser} \\ \lambda & (i, j) \notin E_{ser} \end{cases} \tag{8}$$

Table 1. Quantitative Results

Sequence	Method	MOTA	MOTP	Rcll	MT	PT	FP	FN	IDs	FM
S2L1	[11]	89.3	56.4	-	-	-	-	-	-	-
	[12]	90.3	74.3	**96.8**	**18**	**1**	282	**148**	**22**	**15**
	proposal	88.1	77.7	91.3	17	2	112	405	37	30
	+group	91.1	**79.3**	94.2	**18**	**1**	113	268	34	27
	Ours	**91.9**	79.1	94.9	**18**	**1**	**110**	239	26	20
Stadtmitte	[12]	56.2	61.6	**69.1**	**4**	**6**	134	**357**	15	**13**
	Ours	**58.7**	**62.6**	68.5	**4**	**6**	**100**	364	**14**	14
Kindergarten	[11]	83.1	80.5	85.4	5	3	4	57	5	5
	Ours	**91.6**	**80.8**	**92.3**	**6**	**2**	**3**	**30**	**0**	**3**
Crossing	[11]	81.4	81.3	83.4	4	2	**0**	48	6	6
	Ours	**97.9**	**82.4**	**98.3**	**6**	**0**	1	**5**	**0**	**1**

Fig. 4. Results on Kindergarten. The results of [11] (top). Our results (bottom).

where λ is the constant cost between two unlinked nodes.

Because tracklets in $\mathcal{T}_{gr}$ are highly confident, we assume that each node should be in an active trail. We can formulate the problem as binary linear programming problem.

5. EXPERIMENTS

Dataset. We have already tested our approach on various datasets and achieved very competitive results. For a fair comparison, we use the same detection set as [11] and [12]. So all tracking approaches are based on the same input.

To further verify the effectiveness, we evaluate our approach on the surveillance sequence with two main characteristics: 1) Pedestrians walk in groups. 2) Targets blocked by another targets. In addition, the targets are sometimes blocked by the obstacles in the scene.

5.1. Framework verification

As shown in Tab.1, two intermediate results(only using proposal labeling and adding only group labeling) and final result on PETS2009-S2L1 are given. From the tracking results, we find that the partial method (proposal labeling only) can generate low confidence tracklets, while one of the trajectories is tracked partially, due to the detector errors caused by two people close to each other. After adding group labeling, MT increases from 17 to 18, so the problem caused by close people is solved. The missing detections of targets in groups are recovered by group analysis and reestimation, so FN plunges from 405 to 268. Compared with the result of proposal labeling, MOTA and MOTP both increase. Then track ID labeling is also added. Compared with results above, the number of fragments falls from 27 to 20, as a result of improving trajectory integrity. The indicators – MOTA, FN and IDs are all better than before. However, MOTP decreases slightly from 79.3 to 79.1, because the position estimation of missing detections in the gap between two trajectory fragments is not the actual value. The results prove each process (proposal labeling, group labeling and track ID labeling) of our framework is effective.

5.2. Quantitative evaluation

Tab.1 illustrates the results compared with some state-of-the-art approaches on four databases. [11] and [12] converts data association into labeling problem as we do. The multiple object tracking precision (MOTA) combines all errors (false positives, false negatives, ID switches) into a percentage, and multiple object tracking precision (MOTP) measures the precision of tracking result. In terms of MOTA and MOTP we have performed the best, so it has proven that our tracking is the best in both error measurement and precision measurement.

6. CONCLUSION

We have presented a PSTG-based framework for multi-target tracking, which includes a novel PSTG-based method describing correlation among tracklets and a multi-label optimization. Contrary to previous label cost tracking methods, we avoid the tracking failure caused by interaction of targets through modelling the mutual influence among trajectories. All components of the correlation among tracklets are considered including motion, appearance, group similarity, trajectory integrity and spatio-temporal continuity. The PSTG energy is minimized by multi-label optimization, including proposal labeling, group labeling and track ID labeling to make the trajectory estimation more accurate. We demonstrate the validity of our approach on various public datasets and our dataset, achieving very competitive results, both visually and in terms of quantitative evaluation with respect to ground truth are also given.

7. ACKNOWLEDGEMENT

This study was partially supported by the National Natural Science Foundation of China (No.61472019), the National High Technology Research and Development Program of China (No.2013AA01A603), the National Aerospace Science Foundation of China (No.2013ZC51), the Programme of Introducing Talents of Discipline to Universities and the Open Fund of the State Key Laboratory of Software Development Environment under grant #SKLSDE-2015ZX-21.

8. REFERENCES

[1] S. Oron, A. Bar-Hillel, D. Levi, and S. Avidan, "Locally orderless tracking," in *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*, June 2012, pp. 1940–1947.

[2] Dong Wang, Huchuan Lu, and Ming-Hsuan Yang, "Online object tracking with sparse prototypes," *Image Processing, IEEE Transactions on*, vol. 22, no. 1, pp. 314–325, Jan 2013.

[3] D. Hall and P. Perona, "From categories to individuals in real time – a unified boosting approach," in *Computer Vision and Pattern Recognition (CVPR), 2014 IEEE Conference on*, June 2014, pp. 176–183.

[4] H. Possegger, T. Mauthner, P.M. Roth, and H. Bischof, "Occlusion geodesics for online multi-object tracking," in *Computer Vision and Pattern Recognition (CVPR), 2014 IEEE Conference on*, June 2014, pp. 1306–1313.

[5] Jerome Berclaz, Francois Fleuret, Engin Turetken, and Pascal Fua, "Multiple object tracking using k-shortest paths optimization," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 33, no. 9, pp. 1806–1819, 2011.

[6] Hamed Pirsiavash, Deva Ramanan, and Charless C Fowlkes, "Globally-optimal greedy algorithms for tracking a variable number of objects," in *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on*. IEEE, 2011, pp. 1201–1208.

[7] Asad A Butt and Robert T Collins, "Multi-target tracking by lagrangian relaxation to min-cost network flow," in *Computer Vision and Pattern Recognition (CVPR), 2013 IEEE Conference on*. IEEE, 2013, pp. 1846–1853.

[8] Zheng Wu, Ashwin Thangali, Stan Sclaroff, and Margrit Betke, "Coupling detection and data association for multiple object tracking," in *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*. IEEE, 2012, pp. 1948–1955.

[9] Bo Yang and Ram Nevatia, "An online learned crf model for multi-target tracking," in *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*. IEEE, 2012, pp. 2034–2041.

[10] Stefano Pellegrini, Andreas Ess, Konrad Schindler, and Luc Van Gool, "You'll never walk alone: Modeling social behavior for multi-target tracking," in *Computer Vision, 2009 IEEE 12th International Conference on*. IEEE, 2009, pp. 261–268.

[11] Anton Andriyenko, Konrad Schindler, and Stefan Roth, "Discrete-continuous optimization for multi-target tracking," in *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*. IEEE, 2012, pp. 1926–1933.

[12] Anton Milan, Konrad Schindler, and Stefan Roth, "Detection-and trajectory-level exclusion in multiple object tracking," in *Computer Vision and Pattern Recognition (CVPR), 2013 IEEE Conference on*. IEEE, 2013, pp. 3682–3689.

[13] Zhen Qin and Christian R Shelton, "Improving multi-target tracking via social grouping," in *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*. IEEE, 2012, pp. 1972–1978.

[14] M. Feldmann, D. Franken, and W. Koch, "Tracking of extended objects and group targets using random matrices," *Signal Processing, IEEE Transactions on*, vol. 59, no. 4, pp. 1409–1420, April 2011.