Computer Vision and Image Understanding 144 (2016) 217–227

Contents lists available at ScienceDirect

Computer Vision and Image Understanding

journal homepage: www.elsevier.com/locate/cviu

PSTG-based multi-label optimization for multi-target tracking

Jiahui Chen[a], Hao Sheng[a,b,*], Chao Li[a,c], Zhang Xiong[a]

[a] State Key Laboratory of Software Development Environment, School of Computer Science and Engineering, Beihang University, Beijing, PR China
[b] Wenzhou Research Institute of Beihang University, Beihang University, Wenzhou, PR China
[c] Shenzhen Key Laboratory of Data Vitalization, Research Institute in Shenzhen, Beihang University, Shenzhen, PR China

ARTICLE INFO

Article history:
Received 24 December 2014
Accepted 13 June 2015

Keywords:
Group
Paratactic–serial tracklet graph (PSTG)
Multi-label
Multi-target tracking

ABSTRACT

Many recent advances in multi-target tracking have grown concern over latent corresponding relation among observations, *e.g.* social relationship. 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 paper, a paratactic–serial tracklet graph (PSTG) theory is proposed for inter-tracklet analysis in multi-target tracking to avoid tracking failure caused by long-term occlusion within group or crossing trajectories. Contrary to recent approaches, a novel PSTG is defined to describe 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 which show similar motion patterns in spatio-temporal neighbor. Serial–tracklet graph enhances the integrity and continuity of trajectories which represent two trajectory fragments of a certain target in different periods. Furthermore, a PSTG-based multi-label optimization algorithm is presented to make the trajectory estimation more accurate. A PSTG energy is minimized by multi-label optimization, including group, integrity and spatio-temporal constraints. Experiments demonstrate the anti-occlusion performance of the proposed approach on several public datasets and actual surveillance sequences, and achieve competitive results by quantitative evaluation.

1. Introduction

Tracking of multiple targets from a video sequence is a challenging problem and a research hot spot in computer vision. The aim of the task is to automatically locate targets, give a certain label to each, and recover their trajectories which mean the continuous position sequences for all targets in the camera scenes. In many applications, such as motion analysis and video content understanding, one of the key technologies is recovering the spatio-temporal trajectories accurately. Though there has been significant progress in tracking technology in recent years, current state-of-the-art multi-target tracking algorithms are no match for human abilities of tracking, both in terms of accuracy and precision. Invisible correlation among observations and unknown number of targets lead to more complicated problem in real scenario. Besides image evidence given by a target detector, direction, appearance, moving group, spatio-temporal continuity and interaction among targets should also be taken into consideration in tracking task.

Many of the perfect tracking approaches adopt the tracking-by-detection framework, in which detections are the target hypotheses extracted from the background by the detector (*e.g.* [1]) in all frames and then associate them into trajectories. Using a target detector can help to reduce the number of model drift when a target has been lost and convert tracking problem into association problem. Association based approaches are powerful at dealing with extended occlusions between targets. The detector produces the per-frame image evidence for each target. Hence, when tracking a single target, the problem is to estimate a trajectory which can fit those evidences well. Obviously, the task is significantly more difficult in the multi-target tracking, since all these data associations must be settled at the same time. Intuitively speaking, one has to establish a unique label for each target, and then simultaneously estimate the motion patterns of all targets and the assignment of detections to the targets.

Andriyenko et al. [2] follow the tracking-by-detection framework. a label cost based energy minimization is used to get local optimum for multi-target tracking. The energy in [2] falls into two categories: continuous part and discrete part. Continuous energy is used to illustrate the goodness of trajectories and discrete energy is mainly to describe the data association among image evidences. Because the optimization with two kinds of energy uses local terms, the tracking problem becomes more challenging. As a result, the consequences of detection errors caused by interaction among targets, such as ID

[*] Corresponding author at: State Key Laboratory of Software Development Environment, School of Computer Science and Engineering, Beihang University, New Building G937, #37 Xueyuan Road, Beijing 100191, PR China.

E-mail address: shenghao@buaa.edu.cn (H. Sheng).

http://dx.doi.org/10.1016/j.cviu.2015.06.002
1077-3142/© 2015 Elsevier Inc. All rights reserved.

switch, trajectory fragment, *etc.*, become more obvious when optimized with label cost only using these local terms.

A paratactic–serial tracklet graph (PSTG)-based multi-label optimization method for multi-target tracking is proposed in this work. Unlike other current approaches, PSTG optimization is proposed to exploit various correlations among tracklets and minimize the energy of PSTG. The correlation among tracklets describing in PSTG optimization algorithm can be paratactic or serial. Moving group, integrity and spatio-temporal constraints are reflected in the correlation among tracklets (track fragments, usually short [3]). Moving group constraint mainly illustrates the group relations of moving targets, which means the targets in same group should have similar trajectories. Using our paratactic–tracklet graph (PTG), such paratactic tracklet group information can be effectively found, which can partly settle detection errors or trajectory fragments caused by occlusion. Integrity constraint means any trajectory that is continuous and has only one pair of endpoints (start point and end point), which can be implemented by our serial–tracklet graph (STG). The relationship of tracklets in temporal and spatial domain is reflected in spatio-temporal constraint, *e.g.*, two concurrence tracklets respectively express two distinct targets at the same time. To solve the above problems, a PSTG-based multi-label optimization algorithm is presented to express the energy function including all these constraints (group, integrity and spatio-temporal). We run extensive experiments on various datasets and achieve very competitive results to demonstrate our PSTG optimization algorithm.

The main contribution of this work is a PSTG-based model for multi-target tracking that

(i) defines a novel PSTG-based describe method which can embody the correlation among all tracklets in spatio-temporal domain;
(ii) includes direction, appearance, moving group, spatio-temporal continuity and interaction among targets;
(iii) explicitly handles long-term partial or full occlusion within group;
(iv) can effectively avoid the tracking failures caused by close or crossing trajectories.

Furthermore, (v) a multi-label optimization framework for multi-target tracking is also proposed. And multi-label consists of proposal labeling, group labeling, track ID labeling in this paper. Through the multi-label optimization, correlation among trajectories is introduced to make multi-target tracking more accurate.

The rest of the paper is organized as follows: related work is discussed in Section 2; graph approach for inter-tracklet analysis is given in Section 3; Section 4 describes the multi-label optimization for multi-target tracking framework; experiments are shown in Section 5, followed by conclusion in Section 6.

2. Related works

Visual tracking has been a research hot spot in computer vision field, and these are abundant related literature. In the presence of a single target, the task is to calculate the target location in every frame and recovery the trajectory for a certain target [4]. In the presence of multiple targets, the challenges become significatively complex. In this research, we concentrate on recent advances related mainly to visual multi-target tracking.

Visual multiple target tracking approaches can usually fall into two categories [5]: online approaches [6–9] and offline approaches [10–13]. Online approaches are used in time-critical scene and it uses the information from the past frames to estimate the current state recursively. For example, an agent-based behavior model is proposed in Yamaguchi et al. [14]. Breitenstein et al. [15] adopt particle filtering to approximate more complex multi-modal posteriors. The approach uses the information of social and environmental factors to predict endpoints of targets and recovery the trajectories. In [16], Danelljan et al. propose an adaptive low-dimensional variant of color attributes to improve tracking performance. Wu et al. compare current online tracking approaches in [17].

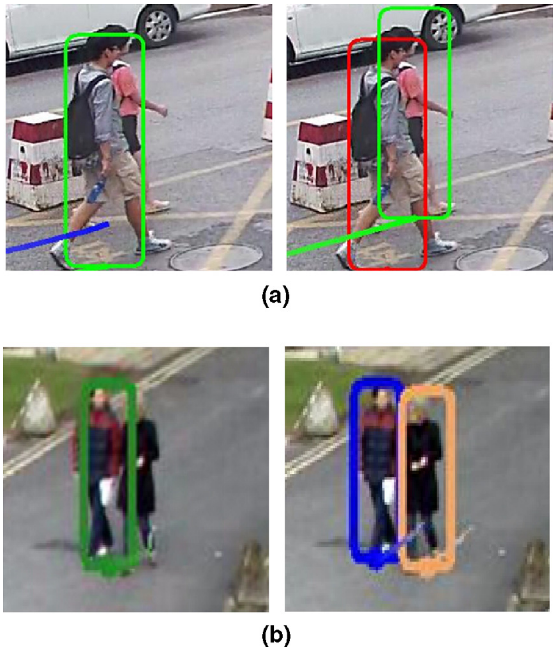

Fig. 1. Typical group errors. (a) shows tracking failure caused by mutual occlusion. (b) shows tracking failure caused by detection error.

In contrast with online approaches, offline approaches are used in accurate tracking scene. In recent years, offline approaches have become more and more popular. A certain latency and globally estimates all trajectories within a given time window are allowed in offline approaches. The offline approaches usually convert the tracking problem into a data association problem, and link image evidences into long trajectories. Min-cost network flow algorithms [18,19] are used to reformulate the task as a network flow problem, which can be solved in polynomial time. Berclaz et al. [20] capture the dependencies between video through network using Markovian dependencies. Pirsiavash et al. [21] propose 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. To avoid ID switches, an online CRF model is proposed to learn appearance features that discriminate among close targets for tracklet association [22]. Similarly, a more effective dynamic model leveraging nearby target positions is proposed in [23]. Label cost based tracker [2] addresses both data association and trajectory estimation by minimizing energy, which is composed of two categories of energy—discrete energy and continuous energy. The discrete energy is used to solve data association problem and the continuous energy to estimate trajectory. However, the correlation among targets which is important in visual tracking is not modeled into the framework. To overcome the drawback, Milan et al. [24] propose a mixed discrete-continuous CRF to fix the weakness. The above approaches still have some limitations. Interaction of targets, integrity of a trajectory and spatio-temporal constraints are overlooked.

Interaction of targets illustrates the mutual influence caused by close moving targets, so it is more important in pedestrian tracking. As shown in Fig. 1, when two pedestrians are close, detection errors (false alarms or multiple targets combined into one detection) occur frequently.

Integrity of trajectories refers to a trajectory with unique start and end points. While a trajectory with more than one pair of

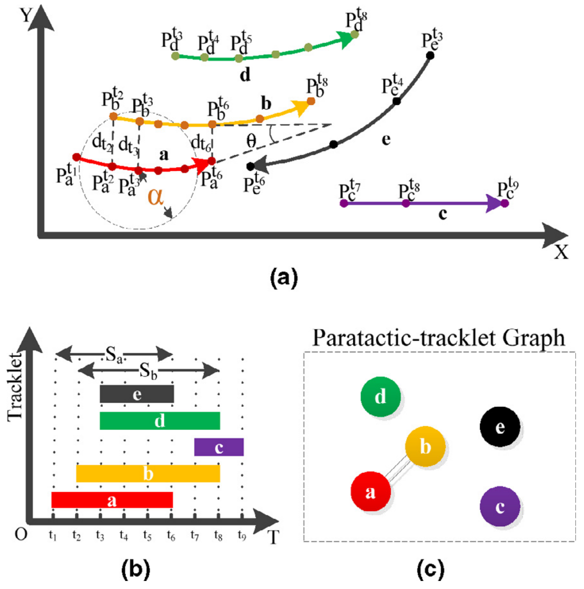
(a)
Y
X
d
b
a
e
c
α
θ
(b)
Tracklet
O
T
S_a
S_b
t_1 t_2 t_3 t_4 t_5 t_6 t_7 t_8 t_9
(c)
Paratactic-tracklet Graph

Fig. 2. Paratactic–tracklet graph (PTG). (a) illustrates the constraints in spatio domain and (b) shows the constraints in temporal domain. (c) is the PTG of tracklets including *a*, *b*, . . . , *e*.

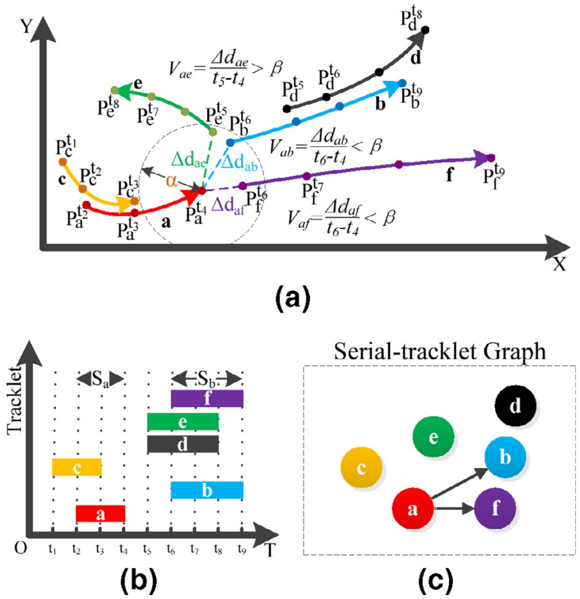
(a)
Y
X
$V_{ae}=\frac{\Delta d_{ae}}{t_5-t_4}>\beta$
$V_{ab}=\frac{\Delta d_{ab}}{t_6-t_4}<\beta$
$V_{af}=\frac{\Delta d_{af}}{t_6-t_4}<\beta$
Δd_{ae}
Δd_{ab}
Δd_{af}
α
(b)
Tracklet
O
T
S_a
S_b
t_1 t_2 t_3 t_4 t_5 t_6 t_7 t_8 t_9
(c)
Serial-tracklet Graph

Fig. 3. Serial–tracklet graph (STG). (a) illustrates the constraints in spatio domain and (b) shows the constraints in temporal domain. (c) is the STG of tracklets including *a*, *b*, . . . , *f*.

endpoints, the trajectory is not a good result regard to integrity. Without integrity, trajectories would outnumber targets and the problem of trajectory fragment occurs.

Spatio-temporal constraint embodies the tracklet association among different targets, *e.g.* paratactic or serial relation between two trajectories. The paratactic tracklets may have similar motion patterns and the serial tracklets may represent the same target. In labeling based tracker, the constraints are ignored, due to overlapping errors (one object has more than one trajectory at the same time) and combination errors (multiple objects have one trajectory).

3. Graph approach for inter-tracklet analysis

Interaction of close moving targets and the influences of scene to trajectories are always handled ineffectively or ignored, and ID switch, trajectory fragment caused by mutual occlusion, and the position estimation of false negatives have not been settled effectively. Therefore we introduce PSTG to describe the paratactic or serial relation among tracklets to solve interaction problem and integrity problem. Through constructing paratactic–tracklet graph(PTG) and serial–tracklet graph(STG), the potential relationship can be exploited among tracklets as well as between tracklets and outlier points (detections do not belong to any tracklets) around, resulting in a more accurate and integral tracking.

3.1. Paratactic–tracklet graph

Paratactic–tracklet. Two tracklets which have similar motion patterns in spatio-temporal neighbor are defined as paratactic tracklets. If tracklet *a* and tracklet *b* are paratactic tracklets, tracklet *a* and tracklet *b* cannot represent the same target. Paratactic–tracklet graph is used to illustrate this kind of relation between two paratactic tracklets.

A undirected graph $G_{\text{pa}} = (V_{\text{pa}}, E_{\text{pa}})$ is generated as shown in Fig. 2. Node set V_{pa} represents tracklet set $\mathcal{T}$. Edge set E_{pa} in G_{pa} is defined as

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

where $s(\mathcal{T})$, $p_t(\mathcal{T})$, α, $\cos(\mathcal{T}_1, \mathcal{T}_2)$ 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 $\mathcal{T}_2$.

Eq. (1) embodies that two paratactic tracklets should satisfy the following conditions: in temporal domain, two paratactic tracklets should have some overlaps, *e.g.* in Fig. 2b, tracklet *a* and tracklet *b* fulfil $S_a \cap S_b \neq \emptyset$, but tracklet *c* does not; in spatio domain, the distance between two paratactic tracklets should be less than the threshold α, *e.g.* in Fig. 2a, tracklet *a* and tracklet *b* fulfil $d_t < \alpha$, but tracklet *d* does not; in spatio-temporal domain, the velocity directions of two paratactic tracklets should be similar, *e.g.* in Fig. 2a, tracklet *a* and tracklet *b* fulfil $cos(\theta) > 0$, but tracklet *e* does not. Paratactic–tracklet graph of this case is constructed as Fig. 2c.

3.2. Serial–tracklet graph

Serial–tracklet. Two tracklets which may represent two trajectory fragments of a certain target in different periods are defined as serial tracklets. If tracklet *a* and tracklet *b* are serial tracklets, tracklet *a* and tracklet *b* can be either a trajectory fragments or two close tracklets. Serial–tracklet graph is used to show this kind of relation between two serial tracklets.

A directed graph $G_{\text{ser}} = (V_{\text{ser}}, E_{\text{ser}})$ is generated as shown in Fig. 3. Node set V_{ser} represents tracklet set $\mathcal{T}$. Edge set E_{ser} is defined as

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

where $s(\mathcal{T})$, $p_{\text{st}}(\mathcal{T})$, $p_{\text{en}}(\mathcal{T})$, α, v, β 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.

Eq. (2) embodies that two serial tracklets should satisfy the following conditions: in temporal domain, two serial tracklets should have no intersection, *e.g.* in Fig. 3b, tracklet *a* and tracklet *b* fulfil $S_a \cap S_b = \emptyset$, but tracklet *c* does not; in spatio domain, the distance between two serial tracklets' endpoints should be less than the threshold α, *e.g.* in Fig. 3a, tracklet *a* and tracklet *b* fulfil $\Delta d_{ab} < \alpha$, but tracklet *d* does not; in spatio-temporal domain, the estimated velocity over the gap between two serial tracklets should be reasonable,

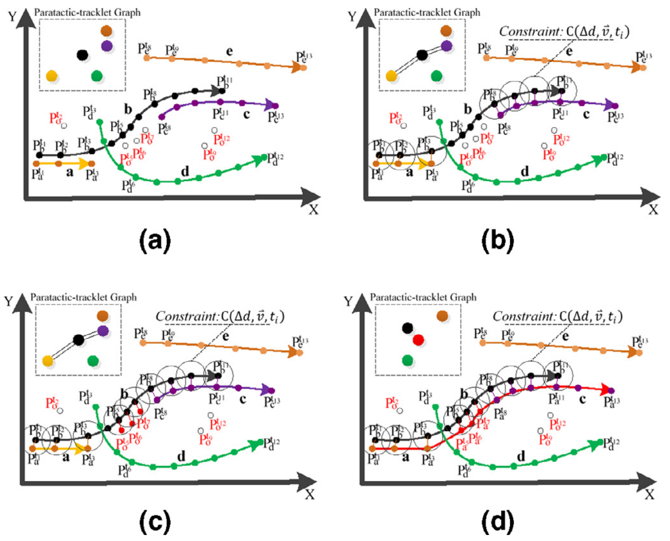
Paratactic-tracklet Graph
Constraint: $C(\Delta d, \vec{v}, t_i)$
a
b
c
d
e
X
Y
(a)
(b)
(c)
(d)

Fig. 4. Group tracklet reestimation. (a) is tracklet state before reestimation. (b) illustrates the propagation field of tracklet *b*. Outlier points around are considered in (c). (d) is the reestimation result.

e.g. in Fig. 3a, tracklet *a* and tracklet *b* fulfil $v_{ab} < \beta$, but tracklet *e* does not. It is obvious that the relation between tracklet *f* and tracklet *a* satisfies all these conditions. Therefore serial–tracklet graph is constructed as shown in Fig. 3c.

3.3. Tracklet group

Approaches [23,25,26] prove that integrity of correlation information among the tracklets has a significantly influence on the quality of multiple target tracking. In our PSTG-based tracking approach, tracklet group is introduced to improve correlation information and solve the problem of locating individual targets position in the case of missing image evidences. Tracklet group can help to solve ID switch and fragment in the situation of targets moving within groups effectively. Tracklet group analysis divides targets into paratactic groups and reestimate group tracklets in each group. As a result, it increase recall rate in multiple target tracking.

Group energy $\mathcal{E}_{\mathrm{gr}}$, indicating how well the labeling result $\mathcal{L}_{\mathrm{gr}}$ satisfies paratactic–tracklet graph, spatio-temporal proximity and motion similarity, is defined as

$$\mathcal{E}_{\mathrm{gr}} = \sum_{i}\sum_{i<j}(c_1\delta(l_{\mathrm{gr}}^{\mathcal{T}_i} = l_{\mathrm{gr}}^{\mathcal{T}_j}\&(\mathcal{T}_i,\mathcal{T}_j) \in E_{\mathrm{pa}}) + c_2(1-\delta(l_{\mathrm{gr}}^{\mathcal{T}_i} = l_{\mathrm{gr}}^{\mathcal{T}_j}\&(\mathcal{T}_i,\mathcal{T}_j) \in E_{\mathrm{pa}}))) \quad (3)$$

where c_1 is the weight of edges in paratactic–tracklet graph between two endpoints with the same label, otherwise c_2. δ is an indicator function.

Group energy $\mathcal{E}_{\mathrm{gr}}$ is minimized to obtain the best labeling result $\mathcal{L}_{\mathrm{gr}}$ based on paratactic–tracklet graph. Tracklets are grouped according to their labels, and tracklets with the same label *i* are grouped as $\mathcal{K}_i$. Asift appearance feature algorithm [27] is applied in each group $\mathcal{K}_i$ to calculate the number of targets in $\mathcal{K}_i$ marked as M_i, where $i \in N_{\mathrm{gr}}$.

In traditional data association based trackers [2,24], correlation information is formulated based on only tracklets with spatio-temporal proximity. To gain more correlation information to recover tracklets, unlike traditional approaches we extend tracklets' correlation through paratactic–tracklet graph. With tracklet graph, we enlarge propagation range of tracklet correlative information in time dimension. The workflow of group tracklet reestimation is shown in Fig. 4. Fig. 4a shows intermediate state of iteration. As shown in Fig. 4b, there is a propagation field around tracklet *b*. In this propagation field, tracklets with the similar direction can be propagated by tracklet *b*. Therefore, tracklet *a* satisfies the conditions. In other words, tracklet *b* can propagate correlation information through tracklet *a*, and vice versa (symmetry). Likewise, tracklet *c* can be propagated by tracklet *b*, so the information is propagated from tracklet *a* to tracklet *c* through tracklet *b*. That is an instance of enlarging propagation range in time dimension. Then tracklets *a*, *b* and *c* belong to the same group. Fig. 4c illustrates the tracklet recovery process with close outlier points using group information $\mathcal{K}$. As a result, tracklets *a* and *c* have been combined into a more integral tracklet as shown in Fig. 4d.

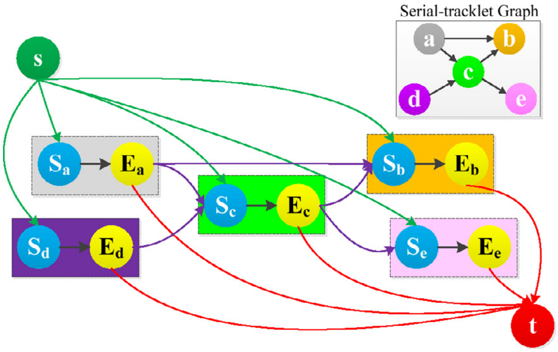

Fig. 5. Tracklet integrity graph. (For interpretation of the references to colour in the text, the reader is referred to the web version of this article.)

3.4. Trajectory integrity

The actual tracklet of a single object in continuous tracking should have only one pair of endpoints (one start point and one end point), but in actual tracking result, a certain target usually has multiple tracklets because of detection errors or tracklet fragments. In our graph approach, we introduce integrity to measure continuity of trajectory.

We propose trajectory energy $\mathcal{E}_{\mathrm{tr}}$ to describe the integrity of trajectories. It is composed of three parts of energy—start energy, end energy and connection energy.

A graph $G_{\mathrm{tr}} = (V_{\mathrm{tr}}, E_{\mathrm{tr}})$is generated as shown in Fig. 5, after serial–tracklet graph of $\mathcal{T}$ is available. Node set 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

$$E_{\mathrm{tr}} = \{(i,j)|i \in V, j \in follow(V_i)\} \quad (4)$$

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

$$follow(v_i) = \begin{cases} \{v \in V|(v_i, v) \in E_{\mathrm{ser}}\} \cup \{t\} & v_i \in V - \{s,t\} \\ V - \{s,t\} & v_i = s \\ \emptyset & v_i = t \end{cases} \quad (5)$$

Fig. 5 is defined as tracklet integrity graph. 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). The weight of each edge is defined as

$$w((v_i, v_j)) = \begin{cases} w_{\mathrm{st}}^{j} & (v_i, v_i) \in E_s \\ w_{\mathrm{en}}^{i} & (v_i, v_j) \in E_t \\ w_{\mathrm{co}}^{(i,j)} & \text{otherwise} \end{cases} \quad (6)$$

where $w_{\mathrm{st}}, w_{\mathrm{en}}, w_{\mathrm{co}}$ are respectively the weights of edges in E_s, E_t and E_c.

From all the above, tracklet energy is defined as Eq. (7).

$$\mathcal{E}_{\mathrm{tr}} = \sum_{(m,n)\in E_{\mathrm{tr}}} (w(m,n)) \cdot \mathrm{active}((m,n)) \quad (7)$$

where $w(m,n)$ is the weight of edge (m,n) and *active*(e) is the active indicator function of edge *e*. When *e* is active(*cf.* Section 4.5), its value

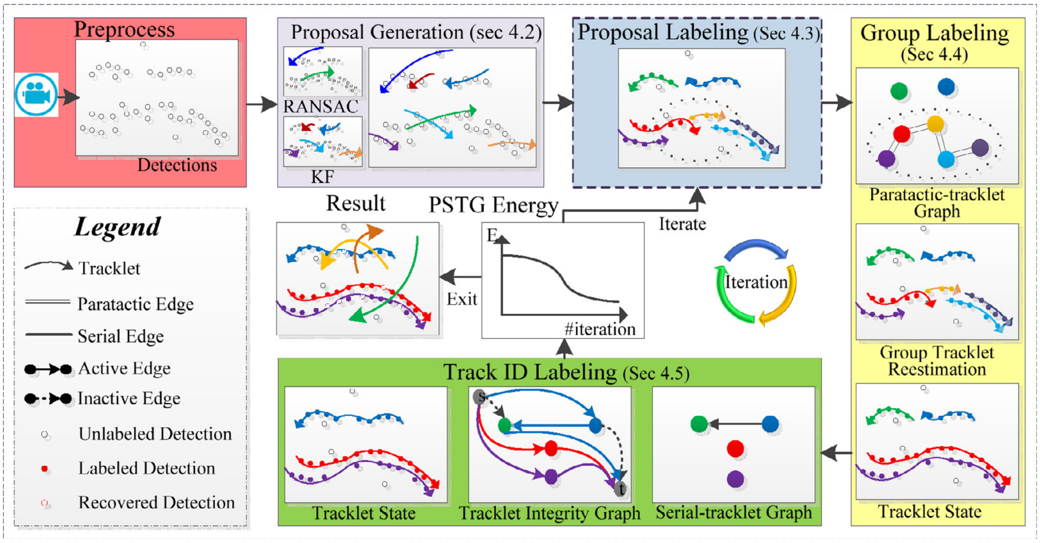

Fig. 6. PSTG-based tracking framework. Each iteration consists mainly of three steps: proposal labeling, group labeling and track ID labeling. Proposal labeling can be any algorithm processing tracklet proposals.

is 1, and otherwise is 0. When one edge is active, it means the two tracklets connected by the edge represent the same target.

4. Multi-label optimization for multi-target tracking

4.1. Architecture of multi-label optimization

We propose a novel framework called PSTG-based tracking, shown in Fig. 6 to solve the above problems – interaction of targets, integrity of trajectories and spatio-temporal constraint. In our 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.

In preprocessing, detections of a video sequence are given by a target detector, and tracklet proposals are generated in proposal generation. Then there is an iteration optimization which consist of proposal labeling, group labeling and track ID labeling. Proposal labeling produces simple tracklets. Group labeling and track ID labeling are respectively used to handle interaction of targets and integrity of trajectories.

The process of proposal labeling in this framework can be replaced by any approach which can achieve tracklet proposals. In the paper, we apply a label cost approach [2] to generate tracklet proposals and compare our approach with original ones [2,24] to show the effect of PSTG analysis.

We propose a tracking framework of multi-label optimization to show the effectiveness of graph approach. As we pursue tracking by detection, it is assumed that detection set D has been given. We denote the ith detection at frame t as d_t^i.

We settle tracking problem by solving multi-label optimization iteratively. As shown in Fig. 6, in each iteration, proposal labeling, group labeling and track ID labeling are applied respectively. Proposal labeling is used to produce low confidence tracklets. Group labeling reestimates the trajectory fragments in each group and track ID labeling enhances the integrity of trajectories generated by group labeling.

Proposal labeling. Given proposal set

$$\mathcal{H} = \{\mathcal{H}^1, \mathcal{H}^2, \ldots, \mathcal{H}^{N_h}, \mathcal{H}^0\}, \tag{8}$$

where $\mathcal{H}^0$ is for outlier detections, *e.g.* false alarms. Proposal is a short track for further process which is generated in Section 4.2. The goal of proposal labeling is to label each d_t^i by a label

$$l_{\text{pr}}^{d_t^i} \in \{1, 2, \ldots, N_h, 0\}, \tag{9}$$

which indicates d_t^i corresponds to detection in $\mathcal{H}^{l_{\text{pr}}^{d_t^i}}$. Detection d_t^i with $l_{\text{pr}}^{d_t^i} = 0$ is an outlier detection. Proposal energy $\mathcal{E}_{\text{pr}}$ is minimized to obtain a suitable proposal labeling result $\mathcal{L}_{\text{pr}}$. Detections with the same label are put into one set, and B-splines are used to fit the detections in one group to tracklets. B-splines are piecewise polynomial functions which can represent complex tracklets. The spline set with N_{pr} elements is denoted as $\mathcal{T}_{\text{pr}} = \{\mathcal{T}_{\text{pr}}^1, \mathcal{T}_{\text{pr}}^2, \ldots, \mathcal{T}_{\text{pr}}^{N_{\text{pr}}}\}$.

Group labeling. The goal of group labeling is to label each $\mathcal{T}_{\text{pr}}^i$ by a label

$$l_{\text{gr}}^i \in \{1, 2, \ldots, N_{\text{gr}}\}, \tag{10}$$

which indicates $\mathcal{T}_{\text{pr}}^i$ belongs to group l_{gr}^i. Group is a target set in which every element has a similar motion pattern. Group energy $\mathcal{E}_{\text{gr}}$ is minimized to obtain a suitable group labeling result $\mathcal{L}_{\text{gr}}$. Trajectories with the same label is put into one group. $\mathcal{T}_{\text{gr}} = \{\mathcal{T}_{\text{gr}}^1, \mathcal{T}_{\text{gr}}^2, \ldots, \mathcal{T}_{\text{gr}}^{N_{\text{gr}}}\}$ is generated by reestimating trajectories in one group.

Track ID labeling. The goal of track ID labeling is to label each $\mathcal{T}_{\text{gr}}^i$ by a label

$$l_{\text{in}}^i \in \{1, 2, \ldots, N_{\text{in}}\}, \tag{11}$$

which indicates tracklet $\mathcal{T}_{\text{gr}}^i$ is a trajectory fragment of target l_{in}^i. Tracklet energy $\mathcal{E}_{\text{in}}$ is minimized to obtain a suitable tracklet labeling result $\mathcal{L}_{\text{in}}$. The integrity of tracking improves through trajectory optimization with $\mathcal{L}_{\text{in}}$.

Then, energy of multi-label framework $\mathcal{E}$ is defined as

$$\mathcal{E} = \mathcal{E}_{\text{pr}} + \mathcal{E}_{\text{gr}} + \mathcal{E}_{\text{in}} \tag{12}$$

where $\mathcal{E}_{\text{pr}}, \mathcal{E}_{\text{gr}}, \mathcal{E}_{\text{in}}$ are introduced in Sections 4.3, 4.4, 4.5 respectively.

The algorithm of PSTG tracking is shown in Algorithm 1. Line 1 generates proposal set according to Section 4.2. The implements of line 3, line 6 and line 8 are respectively given in Algorithms 2, 3 and 4. More details about line 4 are introduced in Section 3.1. New proposals produced by line 3, line 6 and line 8 are added to $\mathcal{H}$ in line 9. Meanwhile, proposals which are unused for several iterations (three iterations in our experiments) are removed from $\mathcal{H}$.

4.2. Proposal generation

As shown in Fig. 6, the first step of the approach is proposal generation. In our experiment, the proposals fall into two categories—proposals generated through RANSAC (RANdom SAmple Consensus)

Algorithm 1 PSTG tracking.

```
Require: Detection Set $\mathcal{D}$
Ensure: Track set $\mathcal{T}$
  generate proposal set $\mathcal{H}$ using RANSAC and KF
  while energy decrease do
      $\mathcal{T}_{\text{pr}} = ProposalLabeling(\mathcal{D}, \mathcal{H})$
      $\mathcal{A}_{\text{PTG}} = PTGAnalysis(\mathcal{T}_{pr})$
      if There is any groups in $\mathcal{A}_{\text{PTG}}$ then
          $\mathcal{T}_{\text{gr}} = GroupTrackletReestimation(\mathcal{A}_{\text{PTG}})$
      end if
      $\mathcal{T}_{\text{in}} = TrackIDLabeling(\mathcal{T}_{\text{gr}})$
      $\mathcal{H} = Update(\mathcal{H}, \mathcal{T}_{\text{in}})$
      calculate energy with Eq. (12)
  end while
  $\mathcal{T} = \mathcal{T}_{\text{in}}$
```

Algorithm 2 Proposal labeling.

```
Require: Proposal set $\mathcal{H}$, Detection Set $D$
Ensure: Track set $\mathcal{T}_{\text{pr}}$
  for $i \in \mathcal{H}$ do
      for $j \in \mathcal{D}$ do
          calculate detection-proposal weight $w_{\text{det}}^{i,j}$
      end for
      calculate proposal weigh $w_{\text{tra}}^{i}$
  end for
  set the trajectory number punishment $\gamma$
  label cost optimization with $w_{\text{det}}, w_{\text{tra}}, \gamma$
  form $\mathcal{T}_{\text{pr}}$ with optimization result
```

Algorithm 3 Group tracklet reestimation.

```
Require: Target Number $M_i$, Detection Set $D_i$, Start Time $T_s$, End Time
    $T_e$
Ensure: Track Set $\mathcal{T}$
for $i = 1 \rightarrow M_i$ do
    $p_i \leftarrow initial\ position$
    $list_i \leftarrow \emptyset$
end for
$t \leftarrow T_s$
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
$\mathcal{T} \leftarrow BSpline(list^i)$
```

Algorithm 4 Track ID Labeling.

```
Require: Tracklets Set $\mathcal{T}_{\text{gr}}$
Ensure: Track Set $\mathcal{T}_{\text{in}}$
construct STG with $\mathcal{T}_{\text{gr}}$
for $i \in \mathcal{T}_{\text{gr}}$ do
    for $j \in \mathcal{T}_{\text{gr}}$ do
        calculate weight $c(i, j)$
    end for
end for
binary programming Eq. (28)
form $\mathcal{T}_{\text{in}}$ with optimization result
```

and proposals generated through the method described below. The former are short tracklets considering spatio-temporal proximity, and the latter are tracklets considering motion model $m = (p, v)$ which describes motion status of a certain target with position p and velocity v.

Given detection set D, we associate each detection d_t^i with a certain motion model m_t^i. The motion model set $\mathcal{M}$ in frame t is described as $\mathcal{M}_t = \{m_t^1, m_t^2, \ldots, m_t^{N_t}\}$, where N_t is the number of detections in frame t.

A Kalman filter based binary optimization approach is proposed for producing proposals. We convert the model-detection association problem between adjacent frames into a bipartite graph mapping problem. Let

$$V_1 = \{m_t^1, m_t^2, \ldots, m_t^{N_t}\} \tag{13}$$

be the set of motion models in frame t and

$$V_2 = \{d_{t+1}^1, d_{t+1}^2, \ldots, d_{t+1}^{N_{t+1}}, d^{\text{out}}\} \tag{14}$$

be the set of detections in frame $t + 1$, where d^{out} is a virtual node used to link with motion models which have no suitable detections to link with, *e.g.* false alarms. The edge set E_{bi} is defined as

$$\begin{aligned} E_{bi} = &\{(i, j) | i \in V_1, j \in V_2, w(i, j) < \tau\} \\ &\bigcup \{(i, d^{\text{out}}) | i \in V_1\}, \end{aligned} \tag{15}$$

where $w(i, j)$ and τ are respectively the weight of edge between two nodes and the threshold of linking weight.

We generate a bipartite graph $G_{bi} = (V_1, V_2, E_{bi})$ to solve the matching problem. A binary variable $l_{i,j}$ is used to indicate whether motion model $i \in V_1$ links with detection $j \in V_2$ in current frame. Our matching problem can be written as the following binary optimization

$$\min_{(i,j) \in E_{bi}} \quad y = \sum w(i, j) l_{i,j} \tag{16}$$

$$\text{s.t.} \qquad l_{i,j} \in \{0, 1\} \qquad (i, j) \in E_{bi} \tag{17}$$

$$\sum_i l_{i,j} = 1 \qquad i \in V_1 \tag{18}$$

$$\sum_j l_{i,j} \leq 1 \quad j \in V_2 - \{d_{\text{out}}\} \tag{19}$$

The objective function Eq. (16) is minimized with respect to three sets of constraints. Eq. (17) converts the optimization problem into a binary optimization problem, and $l_{i,j}$ can be either 0 or 1. Eq. (18) shows each motion model must match a certain detection, including virtual detection d^{out}. Eq. (19) reflects that each detection in the current frame should be associated with one motion model at most.

With the optimization problem solved, proposal set is generated as

$$\mathcal{H} = \{\mathcal{H}^1, \ldots, \mathcal{H}^{N_h}\} \tag{20}$$

where N_h is the number of proposals.

4.3. Proposal labeling

In proposal labeling, tracker maps elements in detection set D with elements in proposal set $\mathcal{H}$. Proposal energy $\mathcal{E}_{\text{pr}}$ which is used to measure the performance of the proposal labeling $\mathcal{L}_{\text{pr}}$, is defined as

$$\mathcal{E}_{\text{pr}} = \mathcal{E}_{\text{det}} + \mathcal{E}_{\text{tra}} \tag{21}$$

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

$$\mathcal{E}_{\text{det}}(h_l, d_t^i) = \begin{cases} s_{d_t^i} \cdot \|h_l(t) - p_t^i\|^2 & l \in N_h \\ s_{d_t^i} \cdot O & l = 0 \end{cases} \tag{22}$$

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

$$\mathcal{E}_{\text{tra}}(h_l) = \gamma + \text{score}(h_l) \tag{23}$$

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

Proposal energy $\mathcal{E}_{\text{pr}}$ is minimized to obtain a suitable proposal labeling result $\mathcal{L}_{\text{pr}}$. To apply fitting method to these detections in different groups, we can gain tracklet set $\mathcal{T}_{\text{pr}}$. Algorithm 2 shows the process of proposal labeling. Label cost optimization(lines 6) is based on [28].

4.4. Group labeling

In group labeling, tracker maps elements in $\mathcal{T}_{\text{pr}}$ with elements in group set $\mathcal{K}$. A trajectory reestimation method (see below) is applied in each group in order to improve the quality of tracklets. Paratactic–tracklet graph is generated with the information of $\mathcal{T}_{\text{pr}}$. Then group set $\mathcal{K} = \{\mathcal{K}_1, \mathcal{K}_2, \dots, \mathcal{K}_{N_k}\}$ and number set $M = \{m_1, m_2, \dots, m_{N_k}\}$ are calculated by the method described in Section 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 labeling are taken into. The trajectory reestimation in a group is shown as Algorithm 3, then tracklet set $\mathcal{T}_{\text{gr}}$ is gained through the algorithm. Due to the same time interval of video sampling frequency, speed in Algorithm 3 can be represented by the distance between target positions in two adjacent frames.

4.5. Track ID labeling

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

$$w_{\text{st}}^{(i)} = \min_{p \in P_{\text{ex}}} \|p_{\text{st}}^{(i)} - p\| \tag{24}$$

$$w_{\text{en}}^{(i)} = \min_{p \in P_{\text{ex}}} \|p_{\text{en}}^{(i)} - p\| \tag{25}$$

where P_{ex} is the set of exit positions in tracking area. $p_{\text{st}}^{(i)}$ and $p_{\text{en}}^{(i)}$ are respectively the positions of start point and end point of tracklet $\mathcal{T}^i$.

$w_{\text{co}}^{(i,j)}$ in serial–tracklet graph is defined as

$$w_{\text{co}}^{(i,j)} = \begin{cases} \|p_{\text{st}}^{(j)} - p_{\text{en}}^{(i)}\| & (i,j) \in E_{\text{ser}} \\ \lambda & (i,j) \notin E_{\text{ser}} \end{cases} \tag{26}$$

where λ is the constant cost between two unlinked nodes.

As shown in Section 3.4, track ID energy $\mathcal{E}_{\text{in}}$ is defined as

$$\mathcal{E}_{\text{in}} = \sum_{L \in \mathcal{L}} (w_{\text{st}}^{(L^{(1)})} + w_{\text{en}}^{(L^{(\text{end})})} + \sum_{(i,j) \in (L^{(t)}, L^{(t+1)})} w_{\text{con}}^{(i,j)}) \tag{27}$$

where $L = \{L^{(1)}, L^{(2)}, \dots, L^{(\text{end})}\}$ is an active trail (as defined after two paragraphs).

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

$$\min \quad y = \sum_{(i,j) \in E_{\text{in}}} c(v_i, v_j) l_{\text{in}}^{(i,j)} \tag{28}$$

$$\text{s.t.} \quad l_{\text{in}}^{(i,j)} \in \{0, 1\} \qquad (i,j) \in E_{\text{in}} \tag{29}$$

$$\sum_v l_{\text{in}}^{(v,w)} = 1 \qquad w \in V - \{s\} \tag{30}$$

$$\sum_v l_{\text{in}}^{(w,v)} = 1 \qquad w \in V - \{t\} \tag{31}$$

Binary linear programming is applied to solve the optimization problem. As a result, we can gain $\mathcal{L}_{\text{in}} = \{l_{\text{in}}(i,j), i \in V_{\text{in}}, j \in V_{\text{in}}\}$. Active edge set E_{active} is defined as

$$E_{\text{active}} = \{(v_i, v_j) \in E_{\text{in}} | l_{\text{in}}^{(i,j)} = 1\} \tag{32}$$

The active trail $\text{Trail}_i = \{s, v_i^{(1)}, v_i^{(2)}, \dots, v_i^{(n)}, t\}$ from start node s to target node t can be gained through E_{active}.

The tracklets in an active trail $\text{Trail}^{(i)}$ are considered the trajectory fragments of target i. Trajectory set $\mathcal{T}_{\text{in}}$ is gained by applying B-splines method to these tracklets. Algorithm 4 shows the process of track ID labeling.

5. Experiments

Dataset. We have tested our approach on various datasets (both public datasets and actual surveillance sequences) and achieved very competitive results. Here we demonstrate the evaluation results on two public datasets and three surveillance sequences of kindergarten, crossing and school gate. Two public datasets are used to compare our approach with [2] and [24]. The two public sequences are the View_001 (795 frames) of PETS2009-S2L1 and Stadtmitte (179 frames) of TUD dataset. For a fair comparison, we use the same detection set as [2] and [24]. So all tracking approaches are based on the same image evidence input.

To further verify the effectiveness, we evaluate our approach on the several surveillance sequence—kindergarten, crossing and school gate. The scene of kindergarten has two main characteristics: 1) It is very common that children and adults walk in groups. 2) Children are easily blocked by adults due to their height gaps. In addition, the targets in the scene are sometimes blocked by the tree and motorbikes. Then experiments on sequences of crossing and school gate are given. In the scene of crossing, with the camera installed in a low position, target far from the camera is easily blocked by the target near the camera and even the light near the camera. There is also mutual occlusion situation in the scene of school gate.

Implementation details. To make this comparison as fair as possible, we use public ground truths, and the detector based on SVM classification of histograms of oriented gradients (HOG) [29] evidence. And we have marked the ground truth of kindergarten, crossing and school gate by hand.

Parameters. The weights α and β are respectively set to the average width and the average height of detections in PTG and STG. Proposal labeling works with parameter $\gamma = 0.1$ and group labeling does with $C_1 = 0.35$ and $C_2 = 1.0$. Finally, we set λ in track ID labeling to the length of frame diagonal. Parameters are set empirically in our experiment.

Evaluation metrics. We use CLEAR MOT to measure tracking result. The multiple object tracking precision (MOTP ↑), multiple object tracking accuracy (MOTA ↑), recall(Rcll ↑), precision(Prcn ↑), the number of false alarms per frame(FAR ↑), the number of mostly($\geq$ 80%) tracked trajectories (MT ↑), the number of partially (20–80%)

Table 1
Our tracking result on PETS2009-S2L1.

Method	MOTA	MOTP	Rcll	Prcn	FAR	MT	PT	ML	FP	FN	IDs	FM
Proposal	88.1	77.7	91.3	97.4	0.14	17	2	0	112	405	37	30
Group	91.1	79.3	94.2	97.5	0.14	18	1	0	113	268	34	27
Track ID	91.9	79.1	94.9	97.6	0.14	18	1	0	110	239	26	20

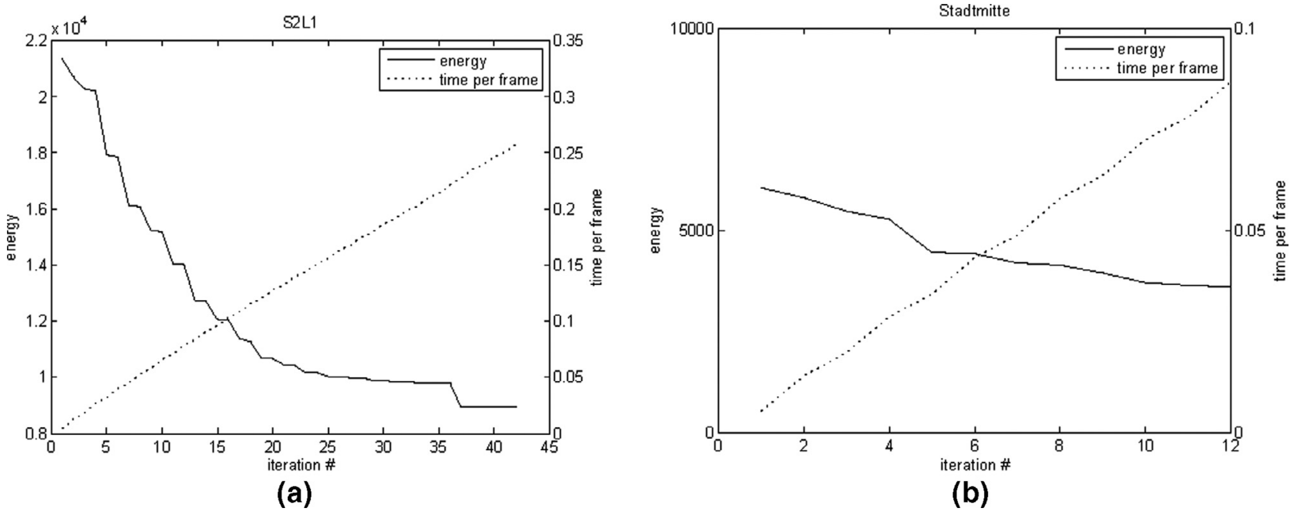

Fig. 7. Energy minimization and time consuming. The solid line shows energy minimization and the dash line shows time consuming.

tracked trajectories(PT), the number of mostly lost trajectories(ML ↓), the number of false positives (FP ↓), the number of false negatives (FN ↓), identity switching (IDs ↓) and the number of trajectory fragments (FM ↓) are used. ↑ is a positive indicator meaning the higher the value, the better, while ↓ means the lower the value, the better.

5.1. Framework verification

Two intermediate results(only using proposal labeling and adding only group labeling) and final result on PETS2009-S2L1 are shown in Table 1. 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. For the same reason, Rcll and Prcn both grow.

The results of energy optimization processing in S2L1 and Stadtmitte are shown in Fig. 7, and the solid lines in Fig. 7 represent the relationship between iteration loop processing and the energy optimization. In the S2L1 dataset, our approach only uses 42 iterations to find the optimal result while [24] needs near 400 iterations. In the Stadtmitte dataset, only 12 iterations are needed in our approach while near 40 iterations are needed in [24]. Moreover, our approach outperforms [24] in both two datasets and it shows that group labeling (PTG analysis) and track ID (STG analysis) make the approach more effective.

5.2. Computational time

We implemented our approach in Matlab without code optimization or parallelization and tested it on a PC with 3.0 GHz CPU and 8 GB memory. The computation efficiency of our approach is shown in Fig. 7. As a result, our approach (0.08 – 0.26 s per frame) works faster than that in [2,24] (1–2 s per frame). Therefore, the framework of PSTG improves performance of energy function optimization.

5.3. Quantitative evaluation on public datasets

PETS2009-S2L1. Table 2 illustrates the results compared with some state-of-the-art approaches on PETS2009-S2L1. Andriyenko et al. [2] convert data association into labeling problem as we do. And Milan et al. [24] are an improvement of Andriyenko et al. [2]. 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. At the same time, precision as well as FP is improved significatively. However, compared with [24], FN increases, because short tracklets have not formed for some missing detections behind the billboard and tracklet graphs cannot embody isolated outlier points.

TUD-Stadtmitte. Comparison with [24] on TUD-Stadtmitte is shown in Table 2. The results are similar with PETS2009-S2L1. Fig. 8 illustrates our tracking results on PETS2009-S2L1 and TUD-Stadtmitte.

5.4. Quantitative evaluation on surveillance sequences

Kindergarten. Tracking results on kindergarten are shown in Table 2. There are eight trajectories in the ground truth, and with our approach, six of them are mostly ($\geq 80\%$) tracked. The rest two trajectories are those of a woman in blue and a man appearing only in the

Table 2
Quantitative results (The best performances are marked in bold).

Sequence	Method	MOTA	MOTP	Rcll	Prcn	MT	PT	ML	FP	FN	IDs	FM
PETS2009-S2L1 (up to eight targets)	[2]	89.3	56.4	–	–	–	–	–	–	–	–	–
	[24]	90.3	74.3	**96.8**	94.1	**18**	**1**	**0**	282	**148**	**22**	**15**
	Ours	**91.9**	**79.1**	94.9	**97.6**	**18**	**1**	**0**	**110**	239	26	20
TUD-Stadtmitte (up to five targets)	[24]	56.2	61.6	**69.1**	85.6	**4**	**6**	**0**	134	**357**	15	**13**
	Ours	**58.7**	**62.6**	68.5	**88.8**	**4**	**6**	**0**	**100**	364	**14**	14
Kindergarten (up to six targets)	[2]	83.1	80.5	85.4	98.8	5	3	**0**	4	57	5	5
	Ours	**91.6**	**80.8**	**92.3**	**99.2**	**6**	**2**	**0**	**3**	**30**	**0**	**3**
Crossing (up to five targets)	[2]	81.4	81.3	83.4	**100**	4	2	**0**	**0**	48	6	6
	Ours	**97.9**	**82.4**	**98.3**	99.7	**6**	**0**	**0**	1	**5**	**0**	**1**
School gate (up to four targets)	[2]	74.2	**75.6**	76.3	**98.3**	5	2	**0**	**4**	70	2	**5**
	Ours	**91.5**	74.7	**94.6**	96.9	**6**	**1**	**0**	9	**16**	**0**	6

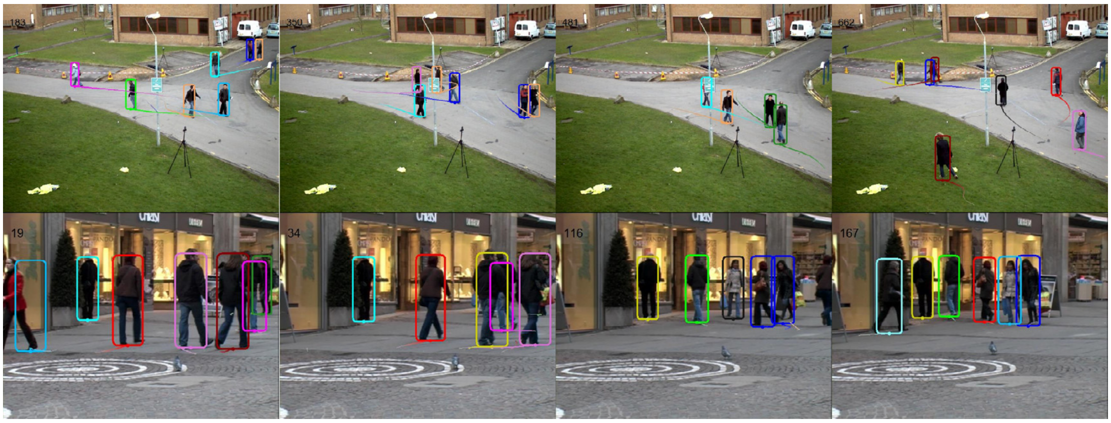

Fig. 8. Our results on PES2009-S2L1 and TUD-Stadtmitte.

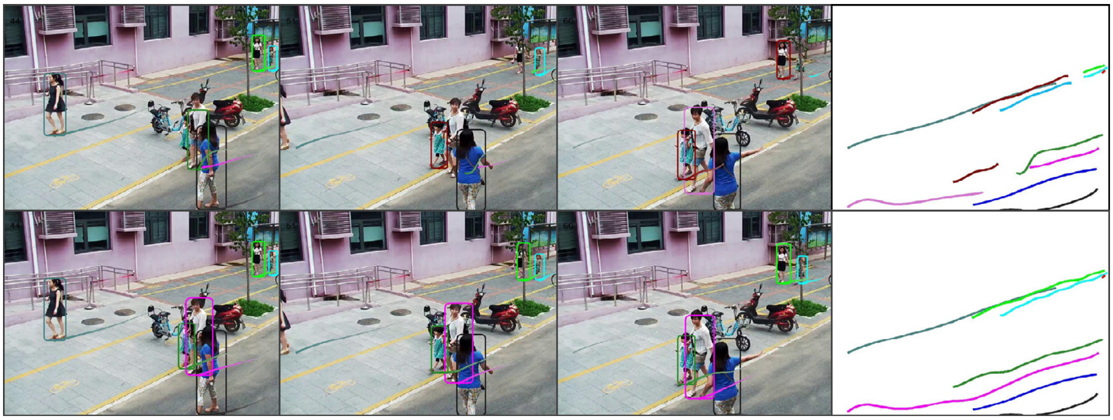

Fig. 9. Results on kindergarten. The results of [2] (top). Our results (bottom). (For interpretation of the references to colour in the text, the reader is referred to the web version of this article.)

last four frames. The woman in blue is not complete after 70 frames in the video, so that detector cannot detect her successfully. The trajectory of the man is recovered in three of four frames (75%), so it is not mostly (≥ 80%) tracked trajectory. All these indicators compared with [2] are improved significantly. In the video, there is target–block occlusion (as shown in Fig. 9, the woman in light green rectangle and the boy in cyan rectangle are both blocked by the tree) and target–target occlusion (as shown in Fig. 9, the boy in dark green rectangle is blocked by the woman in the magenta rectangle, the woman magenta rectangle is blocked by the woman in black rectangle). Our approach works well in all these situations, and it proves that our approach is effective. Fig. 9 illustrates the results of [2] and our tracking results on kindergarten.

Crossing. Table 2 shows the results on crossing. There are six trajectories in the ground truth. All of them are mostly (≥ 80%) tracked with our approach, while only four are mostly (≥ 80%) tracked with [2]. Compared with [2], most of indicators (MOTA, Rcll, MT, PT, FN, IDs and FM) are improved significantly, and MOTP increases at the same time. In the video, there is target–block occlusion (*e.g.*, as shown in Fig. 10, the two man in right side are both blocked by the light) and target–target occlusion (as shown in Fig. 10, the woman in blue rectangle is blocked by the woman in the light green rectangle). Our approach works well in all these situations, and it proves that our approach is effective. Fig. 10 illustrates the results of [2] and our tracking results on crossing.

School gate. Tracking results on school gate are shown in Table 2. There are seven trajectories in the ground truth. Six of them are mostly (≥ 80%) tracked with our approach, but five of them are mostly (≥ 80%) tracked with [2]. MOTA, Rcll, MT, PT, FN and IDs are improved significantly compared with [2]. In the video, there is target–target

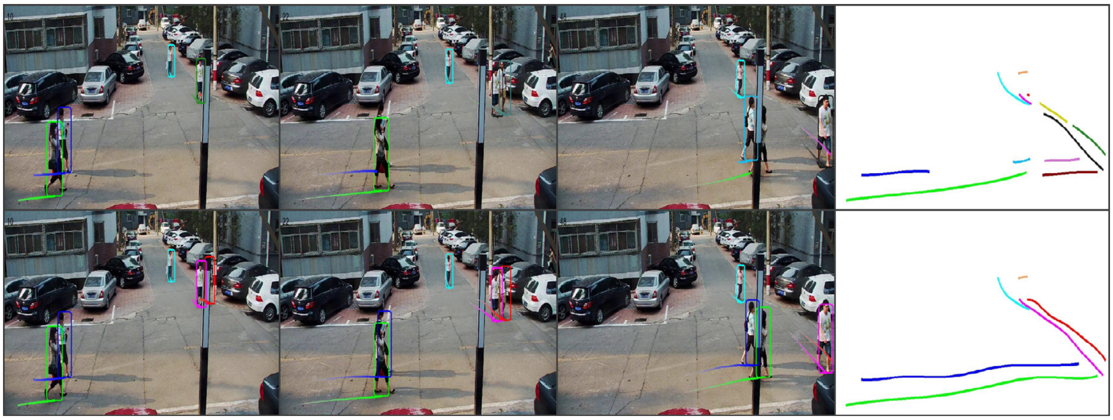

Fig. 10. Results on crossing. The results of [2] (top). Our results (bottom). (For interpretation of the references to colour in the text, the reader is referred to the web version of this article.)

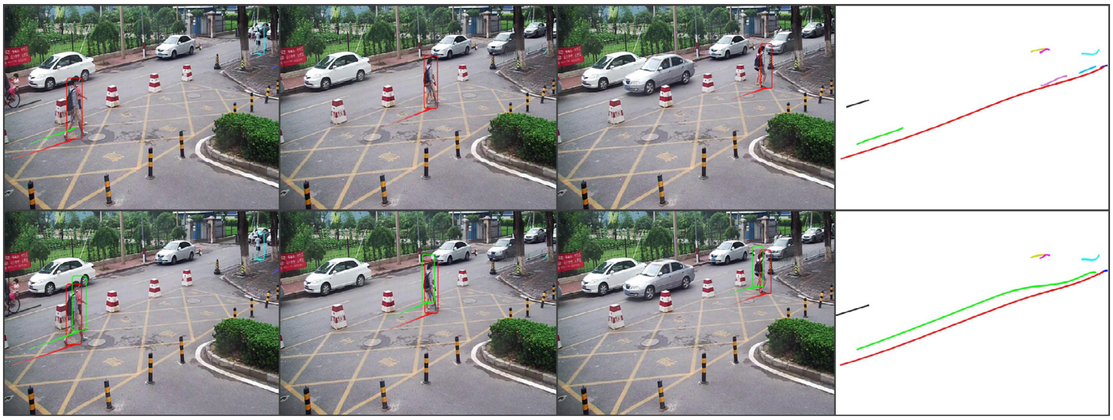

Fig. 11. Results on school gate. The results of [2] (top). Our results (bottom). (For interpretation of the references to colour in the text, the reader is referred to the web version of this article.)

occlusion (as shown in Fig. 11, the woman in light green rectangle is blocked by the man in the red rectangle). Our approach works well in the situation, and it proves that our approach is effective. Fig. 11 illustrates the results of [2] and our tracking results on school gate.

6. Conclusion

In this paper, a PSTG-based framework for multi-target tracking is proposed to handle long-term occlusion within group and tracking failure caused by interaction of targets. The thesis achieves breakthrough on the paratactic–serial tracklet graph approach describing the correlation among tracklets and a multi-label optimization method for trajectory estimation. Contrary to label cost tracking methods, we avoid the tracking failure caused by interaction of targets through modeling 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. In addition, this paper puts forward a multi-label optimization method which embody proposal labeling, group labeling and track ID labeling. And the PSTG energy which can be minimized by multi-label optimization is presented to make the trajectory estimation more accurate. Our experiments demonstrate the validity of our approach on various public datasets and actual surveillance sequences, achieving very competitive results, both visually and in terms of quantitative evaluation with respect to ground truth are also given. Besides improving the tracking accuracy especially with long-term occlusion caused by interaction of targets, the PSTG-based model presents the multi-target association in a social group which is concerned in various applications.

In future work, we plan to integrate more outlier points into our framework to achieve a higher recall rate, thereby further raising the tracking performance. Moreover, the graph-based algorithm we proposed enhance the correlation among tracklets in time domain and it can also be designed to improve the relation in space domain, therefore multi-target tracking in multi-view scene even cross camera tracking can be implemented using our tracklet graph-based approach.

Acknowledgments

This study was partially supported by the National Natural Science Foundation of China (no. 61272350), the National High Technology Research and Development Program of China (no. 2013AA01A603) and the National Aerospace Science Foundation of China (no. 2013ZC51). Supported by the Programme of Introducing Talents of Discipline to Universities and the Fundamental Research Funds for the Central Universities.

Supplementary material

Supplementary material associated with this article can be found, in the online version, at 10.1016/j.cviu.2015.06.002 .

References

[1] M. Andriluka, S. Roth, B. Schiele, Monocular 3D pose estimation and tracking by detection, in: 2010 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, 2010, pp. 623–630.

[2] A. Andriyenko, K. Schindler, S. Roth, Discrete-continuous optimization for multi-target tracking, in: 2012 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, 2012, pp. 1926–1933.

[3] C. Huang, B. Wu, R. Nevatia, Robust object tracking by hierarchical association of detection responses, in: Computer Vision—ECCV 2008, Springer, 2008, pp. 788–801.

[4] Z. Kalal, J. Matas, K. Mikolajczyk, Pn learning: bootstrapping binary classifiers by structural constraints, in: 2010 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, 2010, pp. 49–56.

[5] X. Shi, H. Ling, J. Xing, W. Hu, Multi-target tracking by rank-1 tensor approximation, in: 2013 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, 2013, pp. 2387–2394.

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

[7] D. Wang, H. Lu, M.-H. Yang, Online object tracking with sparse prototypes, IEEE Trans. Image Process. 22 (1) (2013) 314–325.

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

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

[10] Y. Lu, T. Wu, S.-C. Zhu, Online Object Tracking, Learning, and Parsing with and-or Graphs, 2014, pp. 3462–3469.

[11] S.-H. Bae, K.-J. Yoon, Robust Online Multi-object Tracking based on Tracklet Confidence and Online Discriminative Appearance Learning, 2014, pp. 1218–1225.

[12] J. Supancic, D. Ramanan, Self-paced learning for long-term tracking, in: 2013 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2013, pp. 2379–2386.

[13] K. Schindler, Continuous energy minimization for multi-target tracking, IEEE Trans. Pattern Anal. Mach. Intell. 36 (1) (2014) 58–72.

[14] K. Yamaguchi, A.C. Berg, L.E. Ortiz, T.L. Berg, Who are you with and where are you going, in: 2011 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, 2011, pp. 1345–1352.

[15] M.D. Breitenstein, F. Reichlin, B. Leibe, E. Koller-Meier, L. Van Gool, Online multi-person tracking-by-detection from a single, uncalibrated camera, IEEE Trans. Pattern Anal. Mach. Intell. 33 (9) (2011) 1820–1833.

[16] M. Danelljan, F. Khan, M. Felsberg, J. van de Weijer, Adaptive color attributes for real-time visual tracking, in: 2014 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2014, pp. 1090–1097.

[17] Y. Wu, J. Lim, M.-H. Yang, Online object tracking: a benchmark, in: 2013 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, 2013, pp. 2411–2418.

[18] A.A. Butt, R.T. Collins, Multi-target tracking by Lagrangian relaxation to min-cost network flow, in: 2013 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, 2013, pp. 1846–1853.

[19] Z. Wu, A. Thangali, S. Sclaroff, M. Betke, Coupling detection and data association for multiple object tracking, in: 2012 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, 2012, pp. 1948–1955.

[20] J. Berclaz, F. Fleuret, E. Turetken, P. Fua, Multiple object tracking using k-shortest paths optimization, IEEE Trans. Pattern Anal. Mach. Intell. 33 (9) (2011) 1806–1819.

[21] H. Pirsiavash, D. Ramanan, C.C. Fowlkes, Globally-optimal greedy algorithms for tracking a variable number of objects, in: 2011 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, 2011, pp. 1201–1208.

[22] B. Yang, R. Nevatia, An online learned CRF model for multi-target tracking, in: 2012 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, 2012, pp. 2034–2041.

[23] S. Pellegrini, A. Ess, K. Schindler, L. Van Gool, You'll never walk alone: modeling social behavior for multi-target tracking, in: 2009 IEEE 12th International Conference on Computer Vision, IEEE, 2009, pp. 261–268.

[24] A. Milan, K. Schindler, S. Roth, Detection-and trajectory-level exclusion in multiple object tracking, in: 2013 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, 2013, pp. 3682–3689.

[25] Z. Qin, C.R. Shelton, Improving multi-target tracking via social grouping, in: 2012 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, 2012, pp. 1972–1978.

[26] M. Feldmann, D. Franken, W. Koch, Tracking of extended objects and group targets using random matrices, IEEE Trans. Signal Process. 59 (4) (2011) 1409–1420.

[27] J.-M. Morel, G. Yu, Asift: a new framework for fully affine invariant image comparison, SIAM J. Imaging Sci. 2 (2) (2009) 438–469.

[28] A. Delong, A. Osokin, H.N. Isack, Y. Boykov, Fast approximate energy minimization with label costs, Int. J. Comput. Vis. 96 (1) (2012) 1–27.

[29] N. Dalal, B. Triggs, Histograms of oriented gradients for human detection, in: IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2005. CVPR 2005, vol. 1, IEEE, 2005, pp. 886–893.