DETECTING GROUP INTERACTIONS BY ONLINE ASSOCIATION OF TRAJECTORY DATA

Fan Chen[*]

School of Information Science
Japan Adv. Insti. of Sci. and Tech. (JAIST)
chen-fan@jaist.ac.jp

Andrea Cavallaro

Centre for Intelligent Sensing
Queen Mary University of London
andrea.cavallaro@elec.qmul.ac.uk

ABSTRACT

We propose a method for detecting group interactions for groups of varying number of objects. We model each object as a moving agent with a direction-aware interest map and group interactions as mutual interests between objects. After grouping objects into unit interactions individually in each frame, we solve the temporal association problem by tracking group interaction over consecutive frames. Optimal grouping is obtained by finding the maximum weight spanning tree of a directed graph formed by objects and their potential interactions. Experimental results show that our method obtained around 80% recalling rates on two publicly available datasets.

***Index Terms*—** Trajectory Analysis, Group Interaction, Hot-spot Detection

1. INTRODUCTION

In surveillance and team-sports analysis, it is usually difficult to specify for interacting objects the set of target behaviors or the number of involved objects. We aim to locate hot-spot events without estimating their types. Assuming that all objects intend to keep their individual moving status as long as possible [1], we use group interactions as clues to locate events both temporally and spatially from trajectory data. In this paper we propose a detection method of group interactions. One key issue in reliable detection arises from the subjectiveness in interaction definition. We handle this difficulty from both the definition and the evaluation methods.

Our target interaction is temporally-continuous spatial proximity of objects: we define them as containing at least two objects at any time; being temporally continuous; and they terminate when the configuration of dominant significance has been largely changed. An object only belongs to one group interaction at any moment, and might belong to different interactions in different periods. The termination condition helps to solve the ambiguity in the boundary of group actions, where the *dominance* of a member object is defined as the overall interests it received from all neighbouring objects.

Besides the above consideration, there are three main contributions in our method: we define a motion-direction-aware interest map for modelling the mutual influence between interacting objects; we model the isolation of unit interactions in each frame as an unsupervised clustering problem, and solve it as a rooted arborescence problem; we interpret the temporal association of group interactions as a tracking task of interactions.

After a brief review of prior works on automatic analysis of group interactions in Section 2, we describe our detection method of group interactions in Section 3. In Section 4, we evaluate the performance of the proposed approach on two standard datasets, and then conclude the paper in Section 5.

[*]This work is supported by the Japanese MEXT Grant-in-Aid for Young Scientists (B) No.23700110.

2. PRIOR WORKS

We roughly classify previous methods on group behaviors into two categories: methods *with or without a pre-specified set* of behaviors.

Methods dealing *with a pre-specified set* of target group behaviors and well-defined scenarios (e.g. detection of face-to-face conversations in a meeting [2] and detection of bag stealing or unattended luggage in video surveillance [3]) use prior knowledge about the scene and detection is performed based on multi-modal inputs, e.g. speech, facial expressions, pose and gaze [4]. With the prior knowledge of target behaviors, it is possible to define the intermediate status of interactions where HMM is applicable to find the optimal solution [5]. Some works also applied constraints to the number of involved objects, e.g. four individuals [5] or two people [3].

Methods *without a pre-specified set* of behaviors rely on detected interactions to understand behaviors of humans (e.g. users of mobile devices [6]) or animals (e.g. baboons [7]). Without enough prior knowledge on the target interactions, the grouping of interacting objects is usually performed in a much simpler way, for example by grouping objects that remain within a specified distance for at least a minimum duration of time [6][8]. In some constrained scenarios, including both spatial constraints (e.g., traffic roads) or temporal constraints (e.g., periodic pattern), clustering techniques could help to model and classify interactive patterns. In less constrained scenarios, more attention is given to several basic motion patterns (e.g. track, encountering, flock) [9].

There are also methods that do not explicitly detect interactions, but incorporate them for improving the performance of certain video analysis, e.g. video tracking [10][11].

Isotropic influence potential function (i.e., it has the same intensity regardless of its direction) tend to include irrelevant objects [10][12]. Visual focus of attention (VFOA) intends to indicate the direction a person is looking at [13]. The head direction is not only difficult to estimate, but could also be inconsistent with the real interested direction. A behavioral force model can describe social influences that an object received from other attractive or repulsive objects during pedestrian movements [1]. This social model was used, usually in a simplified form, for analyzing existed interaction data, e.g. the one with only the attractive force for modelling the desires of objects in mutual approaching [12], and the one with only the repulsive force for solving collisions during multi-target tracking [14]. However, most complex group interactions, such as many self-organized group behaviors, are only explainable by considering both attractive and repulsive forces [1]. We thus need to consider both forces for a better model of group interactions.

3. ONLINE TRACKING OF UNIT INTERACTIONS

3.1. Problem Formulation

Given L trajectories collected from N^F sampled frames, we denote the l^{th} trajectory at the i^{th} frame by $\mathbf{o}_{li} = \{a_{li}, \mathbf{x}_{li}, \mathbf{v}_{li}\}$. a_{li} is for

the availability of a trajectory, which takes 1 when it appears in the present frame and takes 0 otherwise. $\mathbf{x}_{li}$ is its spatial position. Its velocity $\mathbf{v}_{li}$ is computed from a local ΔL-frame long history,

$$\mathbf{v}_{li} = \frac{1}{\Delta L} \sum_{j=i-1}^{i-\Delta L} \frac{\mathbf{x}_{li} - \mathbf{x}_{lj}}{t_i - t_j} \tag{1}$$

where t_i is the time stamp of the i^{th} frame. $s_{li} \geq 0$ is the label of its interaction, where $s_{li} = 0$ if it joins no interaction. At the i^{th} frame, we group all moving objects as $\mathcal{G}_i = \{\mathbf{o}_{li} | a_{li} = 1\}$, and their interactive status as $\mathbf{s}_i = \{s_{li} | a_{li} = 1\}$. We infer the hidden interactive status from observation $\{\mathcal{G}_i\}$ by

$$\begin{aligned} \{\mathbf{s}_i^*\} &= \arg\max_{\{\mathbf{s}_i\}} \log P(\{\mathbf{s}_i\}|\{\mathcal{G}_i\}) \\ &= \arg\max_{\{\mathbf{s}_i\}} \log P(\{\mathcal{G}_i\}|\{\mathbf{s}_i\}) P(\{\mathbf{s}_i\}). \end{aligned} \tag{2}$$

We formulate it as a tracking problem under 1^{st} order dependency,

$$\mathbf{s}_i^* = \arg\max_{\mathbf{s}_i} \log P(\mathcal{G}_i|\mathbf{s}_i, \mathcal{G}_{i-1}) P(\mathbf{s}_i|\mathbf{s}_{i-1}^*) \tag{3}$$

and solve it in three steps: frame-wise grouping of objects into unit interactions;

$$\mathbf{s}_i^{\mathrm{U}} = \arg\max_{\mathbf{s}_i} \log P(\mathcal{G}_i|\mathbf{s}_i, \mathcal{G}_{i-1}) \tag{4}$$

temporal association of interactions between frames;

$$\mathbf{s}_i^* = \arg\max_{\mathbf{s}_i} \log P(\mathbf{s}_i|\mathbf{s}_{i-1}^*, \{\mathbf{s}_i^{\mathrm{U}}\}) \tag{5}$$

and a post-smoothing process.

3.2. Frame-wise Grouping of Unit Interactions

We analyse group interactions by evaluating the intention of each object interacting with other objects. The interaction is modelled as a behavior of multiple objects motivated by unidirectional/mutual interests. Assuming that the movement of each object is mainly driven by his desire in interacting with one other object, we derive

$$P(\mathcal{G}_i|\mathbf{s}_i, \mathcal{G}_{i-1}) \approx \prod_l P(\mathbf{o}_{li}|\mathbf{o}_{\mathcal{P}_{li}}, \mathbf{s}_i, \mathcal{G}_{i-1}) \propto \exp(\sum_l \mathcal{I}_{li,\mathcal{P}_{li}}) \tag{6}$$

$\mathcal{P}_{li}$ denotes its target object. For objects having no high interests on other objects, we simply let it focus on a virtual object $\mathbf{o}^{VR}$ with fixed interest $\mathcal{I}_{\mathrm{Thr}}$.

$$\mathcal{I}_{li,\mathcal{P}_{li}} = \begin{cases} \mathcal{I}_{li}(\mathbf{x}_{\mathcal{P}_{li}}) & \mathbf{o}_{\mathcal{P}_{li}} \in \mathcal{G}_i \\ \mathcal{I}_{\mathrm{Thr}} & \mathbf{o}_{\mathcal{P}_{li}} = \mathbf{o}^{VR} \end{cases} \tag{7}$$

We infer the target object that an object intends to interact with by

$$\{\mathcal{P}_{li}^{\mathrm{U}}\} = \arg\max_{\{\mathcal{P}_{li}\}} \sum_l \mathcal{I}_{li,\mathcal{P}_{li}} \tag{8}$$

We hence intend to find the spanning tree with the maximum interests for the directed graph built from all these objects. We build a directed graph, where each object is a node (Fig.1). For any two objects $\mathbf{o}_{li}, \mathbf{o}_{mi} \in \mathcal{G}_i$, we add one link from $\mathbf{o}_{li}$ to $\mathbf{o}_{mi}$ with weight $\mathcal{I}_{li}(\mathbf{x}_{mi})$, and one link from $\mathbf{o}_{mi}$ to $\mathbf{o}_{li}$ with weight $\mathcal{I}_{mi}(\mathbf{x}_{li})$. We add an extra root node, which receives upward links from all objects with $\mathcal{I}_{\mathrm{Thr}}$. This rooted maximum spanning tree problem is also called rooted arborescence, where Chu-Liu/Edmonds' algorithm (Complexity $O(E + VlogV)$) [15][16] is applied to find

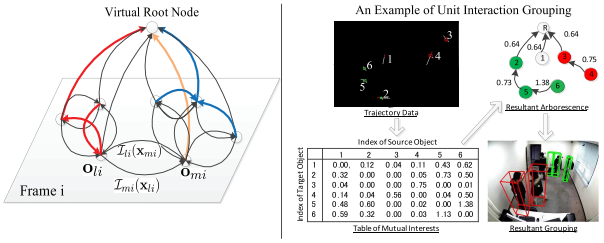

Fig. 1. Framewise detection of unit interactions by grouping objects.

the arborescence solution. Each unit interaction is then detected by computing the cluster of trajectories that are derived from the same path to the root node, among which clusters with only one object are discarded.

$$s_{li}^{\mathrm{U}} = \begin{cases} s_{\mathcal{P}_{li}}^{\mathrm{U}} & \mathbf{o}_{\mathcal{P}_{li}} \in \mathcal{G}_i \\ \text{New-Id} & \mathbf{o}_{\mathcal{P}_{li}} = \mathbf{o}^{VR} \end{cases} \tag{9}$$

An advantage of the rooted arborescence-based clustering is that it is fully unsupervised without the need to know the cluster number.

Observations on pedestrian behaviors include [1]: (i) pedestrians prefer to keep an individual desired speed in the sense of least energy-consuming; (ii) go straight ahead; (iii) have a private sphere (territorial effect) to avoid collisions, (iv) need more space in the walking direction than other directions; (v) show attractive effects, with a longer range of interactions than repulsive effects; and (vi) both attractive and repulsive forces are velocity-dependent.

We design the distribution map of the interests based on their positional relations and the current velocities (Fig.2(a)). In Fig.2(b), we show how typical two-object interactions are interpreted by the above interests. The interest of a moving object $\mathbf{o}_{li}$ in interacting with an object at a position $\mathbf{x}$ is defined as

$$\mathcal{I}_{li}(\mathbf{x}) = \mathcal{I}_{li}^{\mathrm{M}}(\mathbf{x}) + \mathcal{I}_{li}^{\mathrm{S}}(\mathbf{x}) \tag{10}$$

with the main interest $\mathcal{I}_{li}^{\mathrm{M}}(\mathbf{x})$ and the side interest $\mathcal{I}_{li}^{\mathrm{S}}(\mathbf{x})$. The interest received from the main interest field is defined as

$$\mathcal{I}_{li}^{\mathrm{M}}(\mathbf{x}) = \alpha_{li}(\mathbf{x}) \exp\left[-\frac{|\mathbf{x} - \mathbf{x}_{li}|^2}{2(\sigma_{li}^v)^2} - \frac{1 - |\mathcal{C}_{li}(\mathbf{x})|}{2\sigma_{li}^\theta}\right] \tag{11}$$

where $\mathcal{C}_{li}(\mathbf{x}) \equiv \cos(\mathbf{x} - \mathbf{x}_{li}, \mathbf{v}_{li})$, σ_{li}^v controls the spreading of the interests in the radius direction (Interest Effective Range), and σ_{li}^θ controls the distribution of interest in the angular direction,

$$\sigma_{li}^\theta = \log\left[1 + \frac{\widehat{v}}{|\mathbf{v}_{li}| + \epsilon}\right]. \tag{12}$$

$\widehat{v}$ is the average speed computed from all the trajectories, and ϵ is a very small value to prevent the division by zero. We assume that moving objects focus more in their positive moving directions, which is controlled by an extra weight $\alpha_{li}(\mathbf{x})$, i.e.,

$$\alpha_{li}(\mathbf{x}) = \begin{cases} 1, & \mathcal{C}_{li}(\mathbf{x}) \geq 0 \\ \beta + (1-\beta) * \tanh(\frac{\widehat{v}}{|\mathbf{v}_{li}| + \epsilon}), & \mathcal{C}_{li}(\mathbf{x}) < 0 \end{cases} \tag{13}$$

$\alpha_{li}(\mathbf{x})$ approaches 1 when the speed $\mathbf{v}_{li}$ is approaching zero, and decays to β when the speed $\mathbf{v}_{li}$ goes to infinity. As for the side interest, we model it as only interested in very close objects, but less sensitive in the angular direction,

$$\mathcal{I}_{li}^{\mathrm{S}}(\mathbf{x}) = \exp\left[-\frac{|\mathbf{x} - \mathbf{x}_{li}|^2}{2(w^v \sigma_{li}^v)^2} - \frac{|\mathcal{C}_{li}(\mathbf{x})|}{2w^\theta \sigma_{li}^\theta}\right] \tag{14}$$

where w^v and w^θ are two additional weights.

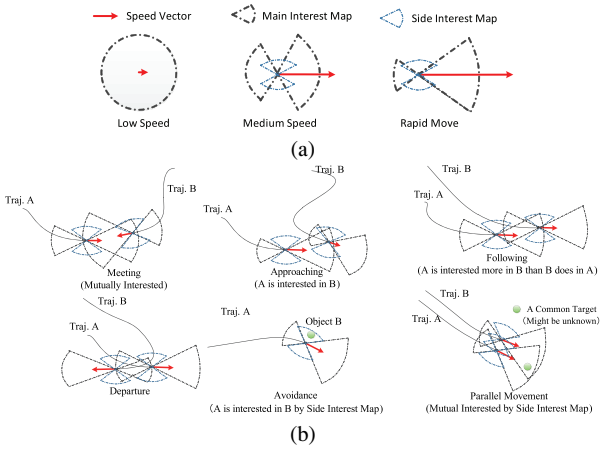

Fig. 2. An interaction is modelled as a mutual interest between object trajectories, where each object has main interests and side interests. (a) Interest map under different speed; (b) Unit interactions in terms of interest map.

3.3. Inter-frame Association of Unit Interactions

Defining $\mathbf{I}_{ik} = \{l | s_{li}^{\mathrm{U}} = k, k > 0\}$, we build two vectors with equal length of $|\mathbf{I}_{ik} \cup \mathbf{I}_{jk'}|$ from two unit interactions $\mathbf{I}_{ik}$ and $\mathbf{I}_{jk'}$, i.e., $\overrightarrow{\mathcal{A}}_{ik} = \{\mathcal{A}_{ik}(\mathbf{o}_{li}) | l \in \mathbf{I}_{ik} \cup \mathbf{I}_{jk'}\}$ and $\overrightarrow{\mathcal{A}}_{jk'} = \{\mathcal{A}_{jk'}(\mathbf{o}_{lj}) | l \in \mathbf{I}_{ik} \cup \mathbf{I}_{jk'}\}$.$\mathcal{A}_{ik}(\mathbf{o}_{li})$ is the accumulated interests that object $\mathbf{o}_{li}$ received from other objects in the interaction $\mathbf{I}_{ik}$,

$$\mathcal{A}_{ik}(\mathbf{o}_{li}) = \begin{cases} \sum_{m \in \mathbf{I}_{ik} \setminus \{l\}} \mathcal{I}_{mj}(\mathbf{x}_{li}) & l \in \mathbf{I}_{ik} \\ 0 & \text{otherwise} \end{cases} \tag{15}$$

The distance between $\mathbf{I}_{ik}$ and $\mathbf{I}_{jk'}$ is defined as

$$\mathcal{D}(\mathbf{I}_{ik}, \mathbf{I}_{jk'}) = 1 - \frac{\cos(\overrightarrow{\mathcal{A}}_{ik}, \overrightarrow{\mathcal{A}}_{jk'}) \min(|\overrightarrow{\mathcal{A}}_{ik}|, |\overrightarrow{\mathcal{A}}_{jk'}|)}{\max(|\overrightarrow{\mathcal{A}}_{ik}|, |\overrightarrow{\mathcal{A}}_{jk'}|)} \tag{16}$$

to reflect the interest shift between these two unit-interactions. For example, when a configuration $[1, 1, 0]$ shifts to $[1, 0, 1]$ or $[1, 0, 0]$, both cases result at a distance of 0.5, which means that half of the elements have changed in the interaction.

The current interaction is regarded to continue unless most objects in this interaction have shifted their major interests. We thus associate unit interactions in consecutive frames by using Eq.5 with

$$\begin{aligned} &P(\mathbf{s}_i | \mathbf{s}_{i-1}^*, \{\mathbf{s}_i^{\mathrm{U}}\}) \\ &\propto \exp\left[-\sum_{lm} \frac{\delta_{s_{li}^{\mathrm{U}}, k} \delta_{s_{mi-1}^{\mathrm{U}}, k'}}{|\mathbf{I}_{ik}||\mathbf{I}_{i-1,k'}|} \mathcal{D}(\mathbf{I}_{ik}, \mathbf{I}_{i-1,k'}) \delta_{s_{li}, s_{mi-1}^*} \right] \end{aligned} \tag{17}$$

subject to $\forall l \neq m, s_{li} = s_{mi} \iff s_{li}^{\mathrm{U}} = s_{mi}^{\mathrm{U}}$. $\delta_{a,b}$ is the Kronecker delta. This is solvable by bipartite graph mapping. A post-thresholding is then performed on two associated unit interactions $\forall l, m, l \in \mathbf{I}_{ik}, m \in \mathbf{I}_{i-1,k'}, s_{li} = s_{mi-1}^*$,

$$s_{li}^* = \begin{cases} s_{mi-1}^* & \mathcal{D}(\mathbf{I}_{ik}, \mathbf{I}_{i-1,k'}) < \mathrm{Thr}^{\mathrm{Conn}} \\ \text{New-Id} & \text{otherwise} \end{cases} \tag{18}$$

3.4. Post-processing of Noisy Objects

When a time lag is allowed, we can include a post-smoothing step that removes those noisy objects before the above post-thresholding:

1: Filter out objects that last less than T_1^{Thr} frames;

$\exists i_1, i_2, \prod_{i=i_1}^{i_2-1} \delta_{s_{li}, s_{l,i+1}} = 1, \delta_{s_{l,i_1-1}, s_{li_1}} = \delta_{s_{li_2}, s_{l,i_2+1}} = 0$
$0 \leq i_2 - i_1 < T_1^{\mathrm{Thr}} \rightarrow s_{li} = 0, i_1 \leq i \leq i_2$.

2: Define the dominance of object $\mathbf{o}_{li}$ in interaction $\mathbf{I}$ as

$$W_{\mathbf{I}}(\mathbf{o}_{li}) = \frac{\sum_{\mathbf{I}_{ik} \in \mathbf{I}} \mathcal{A}_{ik}(\mathbf{o}_{li})}{\sum_l \sum_{\mathbf{I}_{ik} \in \mathbf{I}} \mathcal{A}_{ik}(\mathbf{o}_{li})} \tag{19}$$

and remove the object if it satisfies $W_{\mathbf{I}}(\mathbf{o}_{li}) < 10\%$;

3: Discard short interaction $\mathbf{I}$, if $\sum_{\mathbf{I}_{ik} \in \mathbf{I}} 1 < 20$ frames;

We then perform the post-thresholding with $\mathrm{Thr}^{\mathrm{Conn}} = 0.5$ for a finer division over the resultant sequence.

4. EXPERIMENTAL RESULTS

We use two datasets of trajectories to validate the proposed interaction detection method: the JAIST dataset [17] and the APIDIS basketball dataset [18]. We use ground-truth trajectories and interactions, so as to focus on the accuracy evaluation of interaction detection. In order to suppress the subjectiveness in manual annotations, we only label interactions with very clear semantic meanings. The annotated events in JAIST DB are listed in Table 1. Events in the APIDIS DB are related to clear activities, including offence-defence, tip-off, and player exchange. We only label as ground-truth events with unambiguous semantic meaning and, for each event, we only label the minimum set of unarguable member objects (e.g. gawkers were not labeled in a fighting event). Finally, in our evaluation criteria, we define the overload rate of objects and interactions. A correctly detected group interaction must include all key objects in the minimum set. Inclusion of borderline objects and over-segmentation of events are evaluated by the overload rate of objects and interactions, respectively. We provide limited results here, and invite readers to visit the supplemental material [19] for more results.

Table 1. Interactions for the JAIST Database. FI:Fight, FD:Fall-down, SS:Steal-Suitcase, DS: Drag-Suitcase, EB: Exchange-Bag, SC: Stop-and-Chat, PW:Pickup-Wallet. (The recalling rate is computed at the optimal affective range 1075mm obtained in Fig.3.)

Event Type	FI	FD	SS	DS	EB	SC	PW	Total
Times	4	3	9	17	9	6	3	51
Recalled	4	3	6	16	5	5	1	40

We considered three criteria for performance evaluation. For interaction $\mathbf{I}$, we define $\mathcal{O}(\mathbf{I}|\mathcal{F}) = \{l | \exists \mathbf{I}_{ik} \in \mathbf{I}, i \in \mathcal{F}, l \in \mathbf{I}_{ik}\}$ as the set of its objects within the set of frames $\mathcal{F}$. Given N^{GT} ground-truth and N^{DET} detected interactions, we define the match function between the p-th ground-truth interaction to the q-th detected one,

$$\mathcal{M}_{pq}^{\mathrm{GT}} = \begin{cases} 1, & \mathcal{O}(\mathbf{I}_p^{\mathrm{GT}}|\mathcal{F}_{pq}) \subset \mathcal{O}(\mathbf{I}_q|\mathcal{F}_{pq}) \text{ and} \\ & |\mathcal{O}(\mathbf{I}_q|\mathcal{F}_{pq})| - |\mathcal{O}(\mathbf{I}_p^{\mathrm{GT}}|\mathcal{F}_{pq})| <= E_{obj} \\ 0, & \text{otherwise} \end{cases}$$

In their overlapped period, the interaction should include all objects in the ground-truth interaction and no more than E_{obj} other objects. The positive recall rate is defined as

$$\mathcal{R}^{\mathrm{PR}} = \frac{\sum_p sgn(\sum_q \mathcal{M}_{pq}^{\mathrm{GT}})}{N^{\mathrm{GT}}}, \tag{20}$$

where $sgn(Q) = 1$ if $Q > 0$ and 0, otherwise. The spatial completeness of detected group interactions is evaluated by the overload rate of objects,

$$\mathcal{R}^{\mathrm{OLOBJ}} = \frac{\sum_{pq} |\mathcal{O}(\mathbf{I}_q|\mathcal{F}_{pq})| \mathcal{M}_{pq}^{\mathrm{GT}}}{\sum_{pq} |\mathcal{O}(\mathbf{I}_p^{\mathrm{GT}}|\mathcal{F}_{pq})| \mathcal{M}_{pq}^{\mathrm{GT}}}, \tag{21}$$

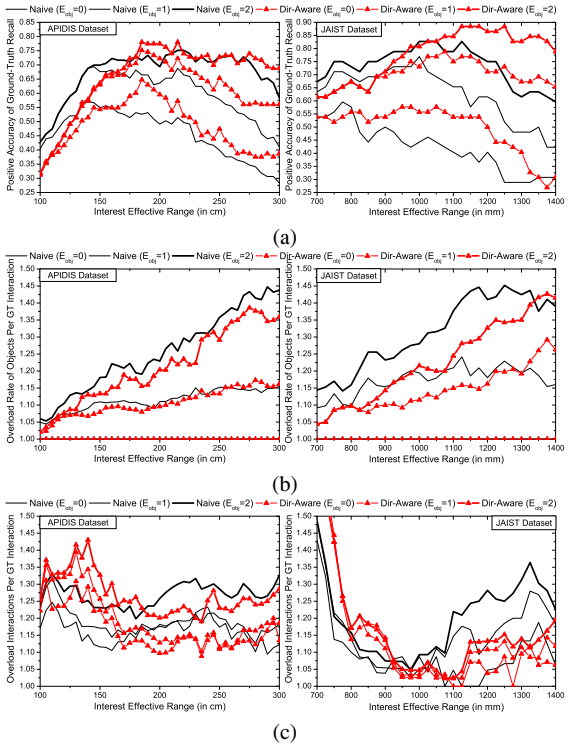

Fig. 3. Performance evaluation of the proposed interaction detection method. (a) Positive Recall Rate of Ground-truth Interactions;(b) Overload Rate of Objects per GT Interaction;(c) Overload Rate of Interactions per GT Interaction.

while the temporal completeness is evaluated by the over-segmentation rate of the ground-truth interactions,

$$\mathcal{R}^{\mathrm{OL}} = \frac{\sum_p \sum_q \mathcal{M}_{pq}^{\mathrm{GT}}}{\sum_p sgn(\sum_q \mathcal{M}_{pq}^{\mathrm{GT}})}. \tag{22}$$

Figure 3 shows the value of the three metrics with respect to various interest ranges σ_{li}^{v} under three different error tolerance thresholds E_{obj}(i.e.,0,1,2), where $\beta = 0.7, w^{v} = 0.4$ and $w^{\theta} = 2$. Comparing our method to an isotropic interest map used in [10], we observed that it not only obtains a higher recalling accuracy of annotated interactions, but also has a lower overload rate of both objects and interactions. The optimal values of σ_{li}^{v} are around $1.1m$ for the indoor dataset, and $1.8m$ for the basketball dataset, respectively. These values are physically meaningful and close to our expectation. As an overall evaluation, if we tolerate one extra redundant object, we obtain around 80% recalling accuracy in both datasets.

We also make experiments to investigate the robustness of the proposed method against tracking errors. There are mainly four types of tracking errors: inaccurate localization, missing detections, false alarms and id-switches. We simulate the first two kinds of tracking errors on our ground-truth trajectories, i.e., inaccurate localization by introducing Gaussian noises and missing detections by randomly removing objects at each frame. The overload rate of objects remains quite stable in both cases, as shown in Fig. 4(a), where inaccurate localization and missing detections mainly decrease the recall rate and over-segment interactions. The proposed method outputs robust detection results with a low over-segmentation rate of

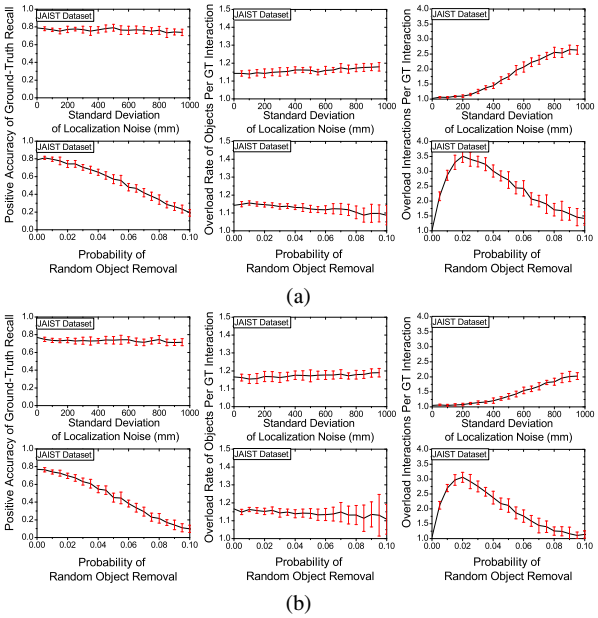

Fig. 4. Evaluation on the robustness of the proposed interaction detection method against simulated tracking noises, under two different pre-smoothing strengths. (a) Results with pre-smoothing (21-frame temporal window); (b) Results with pre-smoothing (41-frame temporal window). Both take the affective range as 1075mm.

events (as below 1.5 here), when the standard deviation of Gaussian noise is lower than 400mm, which is achievable in most conventional tracking methods (e.g.[20]). The recall precision remain high when the probability of random object removal is smaller than 0.02, while the over-segmentation rate of events reaches 3.5. They can be partially corrected by pre-smoothing of trajectories and post-removal of boundary objects. Our manual data are also smoothed from temporally down-sampled labeling. In Fig. 4(b), we enlarged the temporal window for pre-smoothing the trajectories, which clearly alleviates over-segmentation of events. Note that interpolation of missing objects was not performed in our pre-smoothing process, because it is less practical in the real application. False alarms and id-switching increase overload rate of objects by including non-existing or duplicated objects, which are more difficult to simulate. False-alarms usually result at short-term trajectories, most of which could be easily pre-filtered out. Id-switching is more complicated, whose solution requires extra observation data in general.

5. CONCLUSIONS

We described a method to detect group interactions from trajectory data without pre-specified sets of group behaviors, based on associating unit interactions. Experimental results show that the proposed method can extract group interactions with a variable number of group members. The method was validated on two publicly available datasets.

As future work, we will consider to solve the detection as a unified optimization task. We are also interested in semantic understanding of detected group interactions, such as recognizing certain events or detecting abnormalities.

6. REFERENCES

[1] D. Helbing, P. Moln'ar, I. Farkas, and K. Bolay, "Self-organizing pedestrian movement," *Environment and Planning B-planning and Design*, vol. 28, pp. 361–383, 2001.

[2] O. Brdiczka, M. Langet, J. Maisonnasse, and J.L. Crowley, "Detecting human behavior models from multimodal observation in a smart home," *IEEE Trans. Automation Science and Engineering*, vol. 6, no. 4, pp. 588–597, 2009.

[3] M. Taj and A. Cavallaro, "Recognizing interactions in video," in *Intelligent Multimedia Analysis for Security Applications*, pp. 29–57. 2010.

[4] D. Gatica-Perez, "Automatic nonverbal analysis of social interaction in small groups: A review," *Image Vision Comput.*, vol. 27, no. 12, pp. 1775–1787, Nov. 2009.

[5] O. Brdiczka, J. Maisonnasse, and P. Reignier, "Automatic detection of interaction groups," in *ICMI '05*, 2005, pp. 32–36.

[6] Y. Wang, E.P. Lim, and S.Y. Hwang, "Efficient mining of group patterns from user movement data," *Data Knowl. Eng.*, vol. 57, no. 3, pp. 240–282, 2006.

[7] A.J. King, C. Sueur, E. Huchard, and G. Cowlishaw, "A rule-of-thumb based on social affiliation explains collective movements in desert baboons," *Animal Behaviour*, vol. 82, no. 6, pp. 1337 – 1345, 2011.

[8] T. Yu, S.N. Lim, K. Patwardhan, and N. Krahnstoever, "Monitoring, recognizing and discovering social networks," in *CVPR'09*, June, pp. 1462–1469.

[9] H. Jeung, M.L. Yiu, and C.S. Jensen, "Trajectory pattern mining," in *Computing with Spatial Trajectories*, Yu Zheng and Xiaofang Zhou, Eds., pp. 143–177. Springer New York, 2011.

[10] Z. Khan, T. Balch, and F. Dellaert, "Mcmc-based particle filtering for tracking a variable number of interacting targets," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 11, pp. 1805–1918, Nov. 2005.

[11] K. Smith, D. Gatica-Perez, and J.M. Odobez, "Using particles to track varying numbers of interacting people," in *CVPR'05*, june 2005, vol. 1, pp. 962–969.

[12] X. Cui, Q.S. Liu, M.C. Gao, and D.N. Metaxas, "Abnormal detection using interaction energy potentials," in *CVPR'11*, 2011, pp. 3161–3167.

[13] L. Bazzani, D. Tosato, M. Cristani, M. Farenzena, G. Pagetti, G. Menegaz, and V. Murino, "Social interactions by visual focus of attention in a three-dimensional environment," *Expert Systems*, 2012, in print.

[14] S. Pellegrini, A. Ess, K. Schindler, and L.J. Van Gool, "You'll never walk alone: Modeling social behavior for multi-target tracking," in *ICCV'09*, 2009, pp. 261–268.

[15] Y.J. Chu and T.H. Liu, "On the shortest arborescence of a directed graph," *Science Sinica*, vol. 14, pp. 4, 1965.

[16] J. Edmonds, "Optimum branchings," *J. Research of the National Bureau of Standards*, vol. 71, no. B, pp. 7, 1967.

[17] Jaist-MVS-Database, "http://www.jaist.ac.jp/%7echen-fan/multivision/jaistmvsdb.html," 2012.

[18] APIDIS-Basketball-Database, "http://www.apidis.org/dataset/," 2012.

[19] Additional-results, "http://www.jaist.ac.jp/project/prime-proj/results-hotspotdetection.htm," 2012.

[20] Our Tracking Results, "http://www.jaist.ac.jp/project/prime-proj/results-tracking-tad.htm," 2012.