Probabilistic Group-Level Motion Analysis and Scenario Recognition

Ming-Ching Chang Nils Krahnstoever Weina Ge
GE Global Research Center, One Research Circle, Niskayuna, NY, USA
{changm,gewe}@research.ge.com nils@krahnstoever.com

Abstract

This paper addresses the challenge of recognizing behavior of groups of individuals in unconstraint surveillance environments. As opposed to approaches that rely on agglomerative or decisive hierarchical clustering techniques, we propose to recognize group interactions without making hard decisions about the underlying group structure. Instead we use a probabilistic grouping strategy evaluated from the pairwise spatial-temporal tracking information. A path-based grouping scheme determines a soft segmentation of groups and produces a weighted connection graph where its edges express the probability of individuals belonging to a group. Without further segmenting this graph, we show how a large number of low- and high-level behavior recognition tasks can be performed. Our work builds on a mature multi-camera multi-target person tracking system that operates in real-time. We derive probabilistic models to analyze individual track motion as well as group interactions. We show that the soft grouping can combine with motion analysis elegantly to robustly detect and predict group-level activities. Experimental results demonstrate the efficacy of our approach.

1. Introduction

Environments such as schools, transportation hubs, sport venues, and public gatherings are typically characterized by a large number of people that exhibit frequent and complex social interactions [2, 27]. In order to identify activities and behaviors in such environment, it is necessary to understand the interactions taking place at a group level [19, 16, 14]. Understanding group-level interaction is particularly important in surveillance and security, where the gang related activities are the root cause of most criminal behaviors and disorderly conduct. A major goal of this work is to automatically detect and predict events of interest by understanding behaviors and activities at the group level.

There are at least three major issues in performing group-level behavior recognition. First, one needs to define the group structure from a varying number of individuals. Existing methods that track this group structure for behavior recognition [4, 9, 10, 14, 16, 22, 24] mostly rely on a hard decision over an explicit grouping. In our experience this leads to brittle reasoning systems that are sensitive to noise and tracking inaccuracies. Second, the spatial-temporal relationships among individuals can change rapidly, especially in busy environments, which poses the challenge of maintaining and tracking evolving group structures. Approaches based on hard grouping lack the capability to deal with ambiguity in grouping, which are often observed during a transition period when a person gradually joining or leaving a group. Third, given the temporally evolving group structures, efficient inference strategy needs to be investigated in order to perform event recognition.

To the best of our knowledge, our work is the first of the kind to perform and maintain a **soft** grouping structure throughout the entire event recognition stage. The main contribution of this paper is a probabilistic approach to both determine the group structure and perform robust reasoning on top of it. A key feature of our soft grouping strategy is that group-level activities must be represented on a per-individual basis. Our framework can probabilistically predict motion patterns and group-level activities of interest, such as "Are individuals i, j, and k forming a new group?" and "Are two groups going to meet in the future?". Our system operates based on two components:

- a *probabilistic group analysis* to reason about the soft group structure between individuals based on a connectivity graph defined using a track-to-track and a path-based connectivity measure.
- a *probabilistic motion analysis* to reason over the spatial-temporal pattern both at an individual and a group level to perform scenario recognition.

Our approach is capable of handling arbitrary number of individuals. The group representation is general and can be combined naturally with subsequent reasoning — analytic rules can be motivated directly from its probabilistic formulation in combining with other event inference modules. Our recognition framework is thus flexible in adapting new scenarios. Moreover, our model construction is intuitive (user-friendly for non-technical operators) and invariant to site-specific observations. This is because we con-

struct group activity models using scenario-specific predicates. We did not take a learning approach [11] for two reasons: (1) When data availability is limited, the learning of event models is difficult and requires considerable expertise. (2) The integration of domain knowledge in the model design phase is important. Our approach enables the end user to quickly design a new scenario recognition module based on combining existing modules.

This paper is organized as follows. §2 covers related works. §3 describes our probabilistic group analysis. §4 describes how we combine the soft group analysis and track motion analysis (which will be elaborated in §5) to recognize group-level activities. §6 describes the application domain with performance validation. §7 concludes the paper.

2. Related Works

Crowd behavior recognition has been an active research area [2, 27]. We briefly review works that perform group tracking in particular for event recognition. Saxena *et al.* [24] detect abnormal crowd events by tracking feature points using a KLT tracker and building a scenario recognition engine based on thresholding motion measurements. Hoogs *et al.* [14] detect crowd formation and dispersion via relational clustering solved using a spectral graph clique analysis. Activity models are manually specified a priori and limited to a single constant temporal state. Ge *et al.* [10] identify small groups of crowd to study pedestrian behavior. Groups are formed using bottom-up hierarchical clustering with hard decisions.

Ni *et al.* [19] recognize human group activities using three types of localized causalities. Their approach is based on learning concerning training samples and data labeling and is not automatic in requiring manual initialization of human tracks. Ryoo and Aggarwal [22] simultaneously estimate group structure and detect group activities using a stochastic grammar. Grouping is determined by reasoning over specific spatial-temporal relations. Lau *et al.* [16] estimate group structure without identifying individual blobs in tracking. Both the detection-track association and the grouping are hypothesized and posed as a recursive multi-hypothesis model selection problem. Although probabilistic clustering is performed, hard decisions are made in determining the group size.

For pedestrian tracking and motion prediction, most existing methods focus on modeling simple activities of a single person or the interactivity between two. In contrary, activities or interaction among groups, which occur more often in real scenarios, are much less devoted. Existing methods include using social force model [20, 18], floor field [1], correlated topic model [21], and use multi-camera tracking [8, 6]. Abnormal activities are detected using motion pattern [9], bag-of-social force features [18], mixture of dynamic textures [17], *etc*. The flow field tracking of Shah *et al.* [1, 18] handles very crowded scenes; however it does not operate at a group level.

3. Probabilistic Group Analysis

Defining a precise *grouping* of a crowd is challenging due to the complex social interactions and relations that are hard to measure. To handle the uncertainty in video tracking, we avoid explicit group segmentation and instead maintain a probabilistic measure.

3.1. Pairwise Grouping Measure

We first seek an instantaneous pairwise group affinity measure (that represents the probability of a pair of people belonging to a group), by checking if two individuals are physically close. Inspired by standard social norms from Hall's *proxemics* theory [12] for modeling inter-person spacial relations and the social force model [13] for modeling pedestrian dynamics, we define a pairwise grouping measure based on three main terms: the *distance* between two individuals, the *motion* (body pose and velocity) and the *track history*, as illustrated in Fig.1. How this direct connection probability can be extended to express group membership of non-direct neighbors will be described in §3.2.

The above pairwise grouping measure is defined straight from track observations, thus it favors people that are spatially close. Consider that affinity between people is not always isotropic, our individual-centric affinity is not so, either. Denote d_{ij} the Euclidean distance between two people i and j located at $\mathbf{x}_i$ and $\mathbf{x}_j$ on the ground plane (Throughout the paper, symbols i, j, k refer to a track of a person). We assume a person always faces one's motion direction and denote with ϕ_{ij} the angle between i's velocity vector and the relative position vector $\mathbf{p}_{ij} = \mathbf{x}_j - \mathbf{x}_i$ *w.r.t.* person j. In addition, the affinity varies with the velocity magnitudes of i and j. Overall the instantaneous affinity measure for some time t is hence a function

$$p_c^{\text{inst}}(i,j) = f(d_{ij}, \phi_{ij}, ||v_i||, ||v_j||), \tag{1}$$

where subscript c stands for connectivity, and the dependency of time t is made implicit for clarity. Fig.1(d) visualizes our concrete measure hidden behind the abstract def-

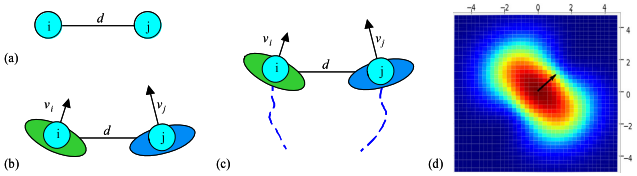


Figure 1. Pairwise group affinity measure: (a) Inter-person distance. (b) Distance and motion (velocity magnitude, direction, frontness/sidedness). (c) Distance, motion, and track history. (d) The instantaneous affinity measure of an individual at $(0, 0)$ with velocity vector $(1, 1)$ in an arrow. Color map depicts this probability kernel between 0 (blue) and 1 (red).

inition (Eq.1). Notice the probability is higher on the side of a person than in the front or back, which is a direct implementation of the aforementioned social norm that states people in a group are more likely to walk side-by-side.

To incorporate track history for robust estimation, we take account p_c^{inst} at time t over a window of T seconds (*e.g.* $T = 3$):

$$p_c^p(i,j;t) = \omega_1 \;\; p_c^{inst}(t) + \omega_2 \;\; \frac{\sum_{t_i \in T} p_c^{inst}(t_i)}{|t_i \in T|}, \quad (2)$$

where ω_1, ω_2 adjust the weights between the two terms of current status and the entire window history ($\omega_1 + \omega_2 = 1$). This improves overall robustness and avoid treating a sudden "passing by" event as an abrupt group change.

3.2. Path-based Group Connectivity

The pairwise affinity measure $p_c^p(i,j;t)$ is defined for two individuals i, j, independent of all other people in the crowd. Observe that two arbitrary individuals in a group do not necessarily have to be directly connected. Rather, it is sufficient that a connecting chain of bonds exists. Here we introduce a path-based group connectivity that estimates the pairwise grouping probability under the influence of others. We say that i and j are *connected*, if there exist pairwise connected intermediate individuals $i_0, ..., i_N$:

$$p_c^\pi(\{i \text{ and } j \text{ are connected via } i_0, ..., i_N\}) = p_c^p(i, i_0) \left[\prod_{k=0}^{N-1} p_c^p(i_k, i_{k+1}) \right] p_c^p(i_N, j). \quad (3)$$

We then set the connection probability between i and j to be the optimal path amongst all possible paths, which yields the highest probability:

$$p_c^\pi(\{i \text{ and } j \text{ are in same group}\}) = \max_{\text{all paths } P_k} p_c^\pi(\{i \text{ and } j \text{ are connected via path } P_k\}). \quad (4)$$

To find the optimal path, we first define the edge weight of the initial connection graph to be $\mathbf{G_0}(i,j) = -\log(p_c^p(i,j))$, whose values ranges from 0 to ∞. We then use Floyd's algorithm [25, Ch.32] to compute the all-pair shortest path in $O(n^3)$, where n is the number of tracked individuals. The resulting graph $\mathbf{G}_\mathbf{0}^\pi$ contains non-negative path weights. We then obtain the final probabilistic connection graph by $\mathbf{G} = p_c^\pi(i,j) = \exp(-G_0^\pi(i,j))$, where $p_c^\pi(i,j)$ is the path-based grouping probability.

The intuition behind this is that the grouping of (i,j) should directly depend on the path created by other individual k in between them. Our path-based metric could be viewed as a simplified solution of a more sophisticated flux-based model, where the connectivity between all pairs of individuals is formulated as a flow, and consider the accumulated flux as the grouping connectivity using the standard *maximum-flow, minimum-cut* algorithm [5, Ch.27]. However the computational cost for the flux-based metric is high (exponential). Our algorithm is also inspired by the spectral clustering [23] and path-based clustering [3] in the domain of pattern classification [7, Ch.10.9].

In case an explicit grouping is desired, we can adopt a proper graph cutting method on $\mathbf{G}$ such as using the hierarchical agglomerative clustering (*e.g. minimum spanning tree* (MST) [4]) or *modularity-cut* [26]. Fig.3 visualize $\mathbf{G}$ (in transparent edges) as well as some explicit grouping (in color polygons) in our test scenarios. Group segmentation is more robust if the hard decision is made only at the last stage of grouping. Our path-based grouping is less bias than the MST-grouping, since all pairs of paths are considered, whereas in [4] weaker connectivity are ignored in the clustering process.

We will show in the next section that we can perform many reasoning tasks using $\mathbf{G}$ without explicit grouping: counting the number of individuals in a group, determining if a group is forming or dispersing, modeling the movement of a group, and at a high level if separate groups are about to engage in aggressive activities such as a fight — all using similar probabilistic reasoning steps.

4. Probabilistic Group Structure Analysis and Scenario Recognition

We describe the key concept toward the flexibility and robustness of our approach, that is to *represent and reason group-level activity on an individual basis using the soft grouping graph* $\mathbf{G}$. This is a novel perspective because no decisive grouping is performed during the reasoning process. Since no explicit grouping is made, we must define the probability of a group-level scenario on an individual basis. For example, "the probability of the group that person i belongs to is chasing the group of j is 0.3". Inference using such probabilistic grouping over time leads to more robust reasoning, in particular on complex group scenarios. Table 1 provides an overview of group scenarios recognized by our system.

Group structure analysis: We analyze both the static group structures (size, compactness) and their dynamic changes (such as formation, dispersion) over time. The *size* of a group that person i belongs to is estimated as the expected value of the number of *healthy* tracks j that i is connected with:

$$G_s(i) = \sum_{\forall j} p_c^\pi(i,j) h(j), \quad (5)$$

where $h(j)$ is the track healthiness incorporated to deal with false and miss detections, by considering Kalman filter covariance and track lifetime.

We consider three status of a group structure: *(i)* the group is growing (formation), *(ii)* shrinking (dispersion), and *(iii)* remaining the same size (stable), with equations

Table 1. Probabilistic group-level scenario recognition.

Group scenario	**Probabilities for track i, or between tracks (i,j)**
Group formation	$p_g^f(i) = \text{sigmoid}(y_g^f, 1, 0.2),\ y_g^f = \sum_{\forall j \neq i} p_c^\pi(i,j;t) \cdot \left[1 - p_c^\pi(i,j;t_p)\right] \cdot \max(h(i), h(j))$
Group dispersion	$p_g^d(i) = \text{sigmoid}(y_g^d, 1, 0.2),\ y_g^d = \sum_{\forall j \neq i} p_c^\pi(i,j;t_p) \cdot \left[1 - p_c^\pi(i,j;t)\right] \cdot \max(h(i), h(j))$
Stable group	$p_g^s(i) = 1 - p_g^f(i) - p_g^d(i)$
Loitering group	$p_g^l(i) = 1 - \prod_{\forall j} \left\{1 - p_c^\pi(i,j) p^l(j)\right\}$
Stable loitering group	$p_g^{sl}(i) = p_g^s(i) p_g^l(i)$
Distinct groups	$p_g^\delta(i,j) = \prod_{\forall k} \{1 - \max(p_c^\pi(i,k)p_c^\pi(k,j), p_c^\pi(j,k)p_c^\pi(k,i))\}$
Close-by groups	$p_g^c(i,j;t) = 1 - \sum_{k \neq i,j} \left[1 - p^c(i,k;t)\right] \cdot \left[1 - p^c(k,j;t)\right]$
Group meeting	$p_g^{meet}(i,j) = 1 - \prod_{t=t_0 \text{ to } t_f} \left\{1 - p_g^c(i,j;t)\right\}$
Group following	$p_g^{flw}(i,j) = p_g^\delta(i,j) \cdot \left[1 - \prod_k \left\{p_g^\delta(i,k) + [1 - p_g^\delta(i,k)] \cdot [1 - p^{flw}(k,j)]\right\}\right]$
Group chasing	$p_g^{chs}(i,j) = p_g^\delta(i,j) \cdot \left[1 - \prod_k \left\{p_g^\delta(i,k) + [1 - p_g^\delta(i,k)] \cdot [1 - p^{chs}(k,j)]\right\}\right]$

given in Table 1. The idea is to check all the neighbors of person i in $\mathbf{G}$ and see if there is a change in the connectivity. For example, if $\forall_{j \neq i}$, the group connectivity $p_c^\pi(i,j)$ is high at current time t and low at some previous time $t_p = t - T_w$, the probability of group formation of person i is high. We use a time window of $T_w = 30$ frames.

Group scenario recognition: In security, group loitering is of particular interest to municipalities, because it is likely related to (or often the prologue of) illegal activities *e.g.* gang activities and disorderly youth. Our analytical definition of a loitering person has three criteria: *(i)* is currently moving slowly, *(ii)* has been close to the current position at a point in time in the past that was at least T_{min} seconds ago and at most T_{max} seconds ago, and *(iii)* was also moving slowly at that previous point in time.

For each person i, the probability of the belonging group G_i is loitering is one minus the probability that all other individuals in the group are not loitering. This *inversion technique* will be used frequently in subsequent group scenario analysis. An attractive characteristic of our framework is its flexibility to recognize new scenarios by combining existing knowledge. As an example, we detect a *stable loitering group* by multiplying the probability of stable group and group loitering (Table 1). Throughout the paper we denote subscript g as group level probabilities. [1]

Pairwise group structure analysis: We consider two basic types of structure between groups: close-by groups and distinct groups. The former only considers pairwise group relationship at one time step while the latter considers track history. Two groups of a pair of people i and j are considered currently close-by if there exists no person k that both i and j are close to. Using the same inversion technique, we define the probability of close-by groups as: $p_g^c(i,j;t) = 1 - \forall_{k \neq i,j} p\big(k \text{ far from } i \text{ and } k \text{ far from } j \text{ at time } t\big)$, where the probability if two individuals being close is elaborated in §5.

The distinct groups p_g^δ between individuals i and j is modeled as $\{\forall_k$, there is no connectivity from (i,k) or $(k,j)\}$, using the same inversion technique. In addition, higher-level scenarios such as *stable distinct groups* and *stable loitering distinct groups* can respectively be defined as multiplications of component probabilities:

$$p_g^{s\delta}(i,j) = p_g^s(i) p_g^s(j) p_g^\delta(i,j), \tag{6}$$

$$p_g^{sl\delta}(i,j) = p_g^s(i) p_g^l(i) p_g^s(j) p_g^l(j) p_g^\delta(i,j). \tag{7}$$

Pairwise group scenario recognition: Building upon various per-track basis motion analysis for individuals in §5 such as meeting, following, and chasing, we can again recognize group-level scenarios using the soft group representation. The probability of the two groups of a pair of people i and j meeting at time t in the future is defined as:

$$p_g^{meet}(i,j) = 1 - \prod_{t=t_0 \text{ to } t_f} \left\{1 - p_g^c(i,j;t)\right\}, \tag{8}$$

where t_0 is the current time, t_f is the time extent in the future, and $p_g^c(i,j;t)$ is the probability of close-by groups.

We define the probability of a group G_i (where person i belongs to) follows an individual j as:

$$p_{gi}^{flw}(i,j) = p_g^\delta(i,j) \cdot (1 - p_{nf}(i,j)), \tag{9}$$

[1] Although intuitively a group satisfying loitering should be stable in a larger time scale, each individual of it could still be considered not in a stable group in a smaller time scale. We multiply factors by assuming these are independent.

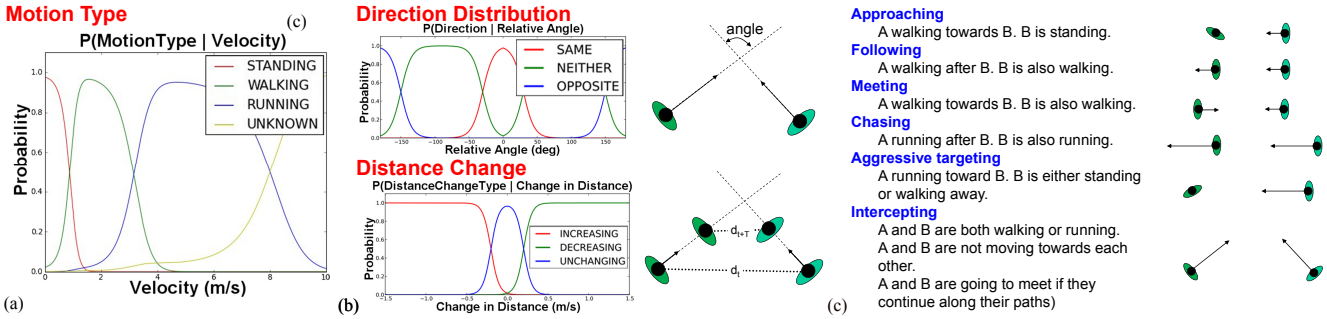


Figure 2. (a) **motion type:** The probability of a person being in different motion types is modeled by a set of sigmoid functions using velocity. (b, top) **relative motion direction:** The probability of different relative moving directions between two people is modeled by a set of sigmoid functions using relative angles. (b, bottom) **relative distance change:** The probability of different relative distance changes between two people is modeled by a set of sigmoid functions using the change in inter-person distance. (c) **pairwise interaction:** Illustration of different types of interaction between two people.

where p_{nf}, the probability of G_i not following individual j is defined as

$$p_{nf}(i,j) = \prod_k \left(p_g^{\delta}(i,k) + (1 - p_g^{\delta}(i,k)) \cdot (1 - p^{flw}(k,j))\right).$$

The intuition is that we consider each individual independently, taking account of two cases: either individual k and follower i are in different groups, or the k and i are in one group but k is not following j.

Next we use Eq. 9 to model the case where a group of individuals is following another group:

$$p_g^{flw}(i,j) = p_g^{\delta}(i,j) \cdot (1 - p'_{nf}(i,j)), \qquad (10)$$

where the probability of G_i not following G_j is defined similarly as before, by taking account of two cases: either individual k and j are in different groups, or k and j are in the same group but k is not followed by i:

$$p'_{nf}(i,j) = \prod_k \left(p_g^{\delta}(j,k) + (1 - p_g^{\delta}(j,k)) \cdot (1 - p_{gi}^{flw}(i,k))\right).$$

The group-level chasing scenario can be defined similarly, in that the probability p^{chs} of chasing individual should be used. We further define a family of related group-level scenarios such as *group approaching*, *group aggressive targeting*, and *group intercepting*, respective to the pairwise interaction outlined in Fig.2(c) in a similar fashion.

5. Probabilistic Individual Motion and Interaction Analysis

This section describes our per-track basis motion analysis, where the equations are summarized in Table 2. We organize the motion scenarios into three main categories (§5.1), which serves as the building block for group motion analysis: *(i)* individual motion types, *(ii)* relative motion direction between pairs, and *(iii)* relative distance change between pairs. These components can be combined with the motion prediction modules (Fig.2) described in §5.2 to predict several scenarios such as *approaching*, *following*, and *chasing* in §5.3.

5.1. Person Motion Analysis

We consider the following motion types M^T of a person: *standing*, *walking*, *running*, and an additional *unknown* state. Denote v the velocity of a person, we use the sigmoid function to model the posterior distribution $p(M^T|v) \simeq p(v|M^T)p(M^T)$, visualized in Fig.2(a). We also use velocity measurements to estimate the relative motion direction between pairs of people. We allow three relative motion directions $M^D = \{same, opposite, neither\}$, which are conditioned on the angle between their velocity vectors. A graphical visualization of the posterior is shown in Fig.2(b, upper).

5.2. Motion Prediction

We utilize the time varying prediction of location probability distributions produced by the Kalman tracker to reason about the chance if two people will spatiotemporally get close. This probability is used in group-level recognition such as close-by and the meeting of groups.

Denote the location prediction of individuals i and j as $x^i(t) \simeq N(z_t^i, S_t^i)$ and $x^j(t) \simeq N(z_t^j, S_t^j)$ respectively, where N denotes normal distribution, z_t is the estimated ground plane location and S_t is the associated uncertainty. The probability that two people are close to each other at time t can be estimated by assuming that the true locations of i and j are indeed $x^i(t)$ and $x^j(t)$: $p^c(i,j;t) = \theta\ (||x^i(t) - x^j(t)|| - \sigma_c)$, where θ denotes thresholding. However, we do not know the exact location of i and j at time t. We hence perform a numerical integration over all possible locations with a set of sample points and weights that represent the distributions $N(x^i; z_t^i, S_t^i)$ and $N(x^j; z_t^j, S_t^j)$. The predicted probability of two people to be close at time t is:

$$p^c(i,j;t|z_t^i, S_t^i, z_t^j, S_t^j) = \sum_{m,n} (\theta(||x_m^i - x_m^j|| - \sigma_c) w_m^i w_m^j). \qquad (11)$$

Table 2. Probabilistic scenario recognition for/between individuals.

Track analysis	Probabilities for track i, or between tracks (i, j)
Track healthiness	$h(i)$, obtained from Kalman filter tracking confidence and lifetime
Person loitering	$p^l(i)$ (omitted in this paper due to space limit)
Person motion type	$p^{mt}(M^T\|, i)$, $M^T = \{$ *standing, walking, running, unknown* $\}$, Fig.2(a)
Relative motion direction	$p^{md}(M^D\|i, j)$, $M^D = \{$ *same, opposite, neither* $\}$, Fig.2(b, upper)
Relative distance change	$p^{dc}(D^C\|i, j)$, $D^C = \{$ *increasing, decreasing, unchanging* $\}$, Fig.2(b, lower)
Track to track pairwise metric	$p_c^p(i, j)$ incorporating front/sided-ness, velocity, and motion track history in Eq.2
Track to track path connectivity	$p_c^\pi(i, j)$, obtained from $p_c^p(i, j)$ after all-pair shortest path computation
Person meeting	$p^{meet}(i, j) = 1 - \prod_{t=t_0 \text{ to } t_f} \{1 - p^c(i, j; t)\}$, where p^c is defined in Eq.11
Person following	$p^{flw}(i, j) = p(M_i^T = \textit{walking})p(M_j^T = \textit{walking})\left[1 - \text{sigmoid}(d_{itc}(i, j), \mu_{d_{itc}}, \sigma_{d_{itc}})\right]$
Person chasing	$p^{chs}(i, j) = p(M_i^T = \textit{running})p(M_j^T = \textit{running})\left[1 - \text{sigmoid}(d_{itc}(i, j), \mu_{d_{itc}}, \sigma_{d_{itc}})\right]$

5.3. Recognizing Pairwise Interaction Scenarios

In estimating probabilistic *meeting*, *i.e.* to determine if two individuals will meet in a future time interval $t \in [0, T]$, it is not trivial to remove the time dependency by marginalizing over the time t, since locations $x^i(t)$ and $x^j(t)$ are not necessarily independent between time steps. We hence chose to select discrete time slices t_a and infer how probable it is that two targets are going to be close at least at one time t_s from a set $\{t_s | s = 0, \ldots, N-1\}$. This is again the inversion technique that $p^{meet}(i, j)$ equals 1 minus the probability that the targets *never* got close (Table 2).

Probabilistic *following/chasing* is estimated based on the person motion types and the prediction of future locations. In a following event, we first compute the interception distance $d_{itc}(i, j)$ between a follower i and a target j by computing the shortest distance between the current group location of j and the predicted location of i (extrapolated using the current velocity estimation). If both people are walking and their interception distance is small, we consider it a following event. Similarly, for both are running, it is then a chasing event (Table 2).

Other scenarios such as probabilistic *approaching*, *aggressive targeting*, and *intercepting*, summarized in Fig.2(c) can be modeled similarly by incorporating motion pattern and motion prediction modules. These can be combined to detect high-level scenarios. For example, the probability of two targets quickly running toward, and meeting at the same location is:

$$\left[p^{md}(\textit{opposite}|i, j) + p^{md}(\textit{neither}|i, j)\right] \cdot p^{dc}(\textit{decreasing}|i, j) \cdot p^{mt}(\textit{running}|i) \cdot p^{mt}(\textit{running}|j) \cdot p^{meet}(i, j).$$

6. Implementation, Results and Evaluation

Video tracking system: Our system is equipped with four standard CCTV cameras for data collection and testing. The tracking system [15] comprises of multiple calibrated and synchronized cameras performing tracking cooperatively. Person detections in each view are projected onto the ground plane in 3D and are then fed into a centralized Kalman tracker operating on the ground plane. Our system runs automatically in real-time (15 to 30 fps) in recognizing pre-defined events for about 5-20 people or more.

To ensure real-life performance, was first deployed our system in a courtyard to test on pedestrian surveillance. We then participate an official field test on the MPR dataset (http://mockprisonriot.org), where several correction officers volunteered to enact security relevant behaviors such as agitated arguments, fights, contraband exchange in an abandoned prison yard in West Virginia, USA. As many activities of interest for correctional settings are related to gang activities, the enacted scenarios often has the presence of multiple groups.

Fig.3 illustrates snapshots of several scenarios detected by our system. Fig.3(a) shows our soft grouping connectivity in the prison yard scenario when 6 enacted inmates are about to form two groups to challenge each other and fight. Our path-based probabilistic grouping scheme successfully identifies the two main groups. Fig.3(b-h) depicts samples of detected group activities in several scenarios selected from both datasets.

We next discuss two specific scenarios in related security applications, where the recognition involves complex spatial-temporal reasoning over low-level and high-level interpretations. We show that our probability framework is flexible and adaptable to solve them.

Case study I: The scenario of *flanking maneuver* is a spatiotemporal configurations exhibited by groups, where one or more aggressive and dominating groups spread out to surround a victim group prior to an attack. Refer to Fig.3(g) for an example occurred in a prison yard. We consider the probabilistic flanking condition $p^{flk}(i, j : k)$, where a victim k is flanked by two others i and j if: (1) i, j are in a group, i, k and j, k are in different groups, the distance $d(i, j)$ is large enough than $d(i, k)$ and $d(j, k)$, and the angle θ_f between $\overrightarrow{ki}$ and $\overrightarrow{kj}$ is large enough. The event probability is then:

Figure 3. Snapshots of several group-level activities captured by our system, where in each case the top depicts one (out of many) camera views and the bottom depicts a top-down 2D planar view, except (h). (a) Cyan lines depict our probabilistic path-based grouping connectivity p_c^{π} as a complete graph of all tracks; lower transparency indicates higher probability. (b) Detected stable loitering groups, where the transparency indicates probability p_g^{sl}. (c) Three detected pairs of stable distinct groups, where each yellow edge depicts $p_g^{s\delta}$. Observe that only people of different groups trigger strong signals. (d) Detected group formation, where the 4 people gather together such that their overall group size is increasing. We denote 'F' in red for formation, 'S' in green for stable, and their transparency indicates probabilities p_g^f and p_g^s. (e) Detected group dispersion in the same sequence after (d), where the 4 individuals disperse ('D' in yellow for p_g^d). (f) Detected group following p_g^{flw} in red edges connecting two major groups, where one is following another. (g) Detected flanking maneuver, where 4 enacted inmates are surrounding 2 victims for an attack. Red circle indicates high probability of being flanked, while green circle indicates low probability. Observe that the grouping connectivity between members in the aggressive group is very strong. (h) Detected contraband exchange with a certain probability.

$$p^{flk}(i,j:k) = p_c^{\pi}(i,j)\,[1-p_c^{\pi}(i,k)]\,[1-p_c^{\pi}(j,k)]\cdot \quad (12)$$
$$\text{sigmoid}(d_{ratio},\mu_{d_r},\sigma_{d_r})\cdot\text{sigmoid}(\theta,\mu_{\theta},\sigma_{\theta}),$$

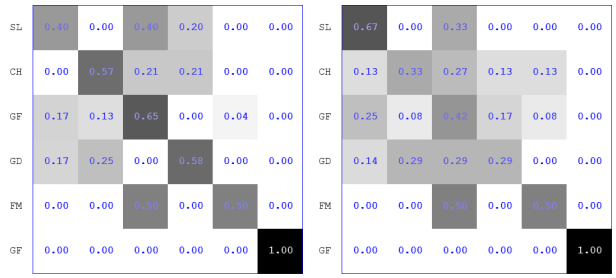
where μ_θ and σ_θ control how wide should the attackers i and j spread in order to flank the victim k; the distance radio $d_{ratio} = \frac{2d(i,j)}{d(i,k)+d(j,k)}$ controls the proper distance between the attackers and the victim. We consider all pairs of i and j for every individual k in accumulating all probabilities. Thus, the probability of individual k is flanked is $1 - p(k \text{ is not flanked by any others})$:

$$p^{flk}(k) = 1 - \prod_{\forall \text{ pairs of } i,j} \big[1 - p^{flk}(i,j:k)\big]. \quad (13)$$

Case study II: Another important application in prison security is to monitor if there is any *contraband handoff* between inmates, where improvised knives, drugs, messages and others are exchanged. This scenario requires two individuals to physically get close to each other during the handoff, Fig.3(h) and also that a short while of T seconds ago, the individuals are not close yet but are approaching each other. There is at least one person walking during the approach, while the second person might be standing or walking. The scenario is modeled as $p(\{\text{i and j are exchanging contraband at time t}\}) =$

$$[p^{mt}(\textit{walking}|i;t_p = t-T) + p^{mt}(\textit{walking}|j;t_p) - p^{mt}(\textit{walking}|i;t_p)p^{mt}(\textit{walking}|j;t_p)]\cdot$$
$$p^{meet}(i,j;t_p)p^{md}(\textit{opposite}|i,j,t_p)\big[1-p^{c}(i,j;t_p)\big]p^{c}(i,j;t).$$

Validation: We have performed evaluation on various group-level events against manually labelled ground truth as shown in Fig.4. The confusion matrix labels are in the order of stable loitering group (SL), contraband handoff (CH), group formation (GF), group dispersion (GD), flanking maneuver (FM), and group following (GF). Note that our evaluation uses videos with complex group interactions where multiple events of interest can occur simultane-



(a) Ground truth (b) Test results

Figure 4. Confusion matrix comparing the accuracy of the group-level event classification (see text).

ously. For example, prior to a gang fighting event while a group of people is loitering, another group can be forming. The contraband hand-off will inevitably include a meeting as a part. Therefore, the diagonal values are not always 1 in the ideal cases as a direct result of composite events (as compared to a conventional confusion matrix) shown in Fig.4(a). Fig.4(b) shows our test result, which fairly resembles the table generated from the ground truth.

7. Conclusion

We have described a probabilistic framework to recognize group-level activity in many scenarios using a novel soft grouping metric and track-based motion analysis. We use a graph between individuals within a scene to determine group membership and interactions, where sound probabilistic estimates can be combined to handle new scenarios. The approach is per-track basis, fully automatic and efficient. In the future, for crowded scenes we can augment our method by first generating a large scale clustering and for each cluster perform detailed group analysis. Another future direction is to recover minor tracking errors by exploiting the probabilistic reasoning about the groups as a feedback.

Acknowledgement. This work was supported by grant #2009-SQ-B9-K013 awarded by the National Institute of Justice. The opinions, findings, and conclusions or recommendations expressed in this publication are those of the authors and do not necessarily reflect the views of the Department of Justice.

References

[1] S. Ali and M. Shah. Floor fields for tracking in high density crowd scenes. In *ECCV*, pages II: 1–14, 2008. 2

[2] J. Candamo, M. Shreve, D. B. Goldgof, D. B. Sapper, and R. Kasturi. Understanding transit scenes: a survey on human behavior-recognition algorithms. *ITSS*, 11(1):206–224, 2010. 1, 2

[3] H. Chang and D.-Y. Yeung. Robust path-based spectral clustering. *Pattern Recogn.*, 41(1):191–203, 2008. 3

[4] M.-C. Chang, N. Krahnstoever, S. Lim, and T. Yu. Group level activity recognition in crowded environments across multiple cameras. In *AVSS W. AMMCSS*, 2010. 1, 3

[5] T. H. Cormen, C. E. Leiserson, R. L. Rivest, and C. Stein. *Introduction to Algorithms*. The MIT Press, 2nd ed., 2001. 3

[6] F. Cupillard, F. Bremond, and M. Thonnat. Group behavior recognition with multiple cameras. In *WACV*, pages 177–183, 2002. 2

[7] R. O. Duda, P. E. Hart, and D. G. Stork. *Pattern Classification*. John Wiley and Sons, Inc., second edition, 2001. 3

[8] R. Eshel and Y. Moses. Tracking in a dense crowd using multiple cameras. *IJCV*, 88(1):129–143, 2010. 2

[9] C. Garate, P. Bilinski, and F. Bremond. Crowd event recognition using HOG tracker. *PETS*, pages 1–6, 2009. 1, 2

[10] W. Ge, R. T. Collins, and B. Ruback. Automatically detecting the small group structure of a crowd. In *WACV*, pages 1–8, 2009. 1, 2

[11] S. Gong and T. Xiang. Recognition of group activities using dynamic probabilistic networks. In *ICCV*, pages 742–749, 2003. 2

[12] E. T. Hall. *The Hidden Dimension*. Anchor, 1966. 2

[13] D. Helbing and P. Molnar. Social force model for pedestrian dynamics. *Physical Review E*, 51:4282–4286, 1995. 2

[14] A. Hoogs, S. Bush, G. Brooksby, A. Perera, M. Dausch, and N. Krahnstoever. Detecting semantic group activities using relational clustering. In *WMVC*, pages 1–8, 2008. 1, 2

[15] N. Krahnstoever, P. Tu, T. Sebastian, A. Perera, and R. Collins. Multi-view detection and tracking of travelers and luggage in mass transit environments. In *PETS*, 2006. 6

[16] B. Lau, K. O. Arras, and W. Burgard. Multi-model hypothesis group tracking and group size estimation. *IJSR*, 2(1):19–30, 2009. 1, 2

[17] V. Mahadevan, W. Li, V. Bhalodia, and N. Vasconcelos. Anomaly detection in crowded scenes. In *CVPR*, 2010. 2

[18] R. Mehran, A. Oyama, and M. Shah. Abnormal crowd behavior detection using social force model. In *CVPR*, pages p35–942, 2009. 2

[19] B. Ni, S. Yan, and A. Kassim. Recognizing human group activities with localized causalities. In *CVPR*, pages 1063–6919, 2009. 1, 2

[20] S. Pellegrini, A. Ess, K. Schindler, and L. van Gool. You'll never walk alone: Modeling social behavior for multi-target tracking. In *ICCV*, pages 261–268, 2009. 2

[21] M. Rodriguez, S. Ali, and T. Kanade. Tracking in unstructured crowded scenes. In *ICCV*, pages 1389–1396, 2009. 2

[22] M. Ryoo and J. Aggarwal. Stochastic representation and recognition of high-level group activities. *IJCV*, pages 1–18, 2010. 1, 2

[23] J. M. Santos, J. Marques de Sa, and L. A. Alexandre. Legclust—a clustering algorithm based on layered entropic subgraphs. *PAMI*, 30(1):62–75, 2008. 3

[24] S. Saxena, F. Brémond, M. Thonnat, and R. Ma. Crowd behavior recognition for video surveillance. In *ACIVS*, pages 970–981, Berlin, Heidelberg, 2008. Springer-Verlag. 1, 2

[25] R. Sedgewick. *Algorithms in C++*. Addison-Wesley Professional, second edition, 1992. 3

[26] T. Yu, S. Lim, K. Patwardhan, and N. Krahnstoever. Monitoring, recognizing and discovering social networks. In *CVPR*, 2009. 3

[27] B. Zhan, D. N. Monekosso, P. Remagnino, S. A. Velastin, and L.-Q. Xu. Crowd analysis: a survey. *Mach. Vis. Appl.*, 19(5/6):345–357, 2008. 1, 2