Who Knows Who – Inverting the Social Force Model for Finding Groups

Jan Šochman
CMP, Dep. of Cyber., FEE, CTU
jan.sochman@cmp.felk.cvut.cz

David C. Hogg
School of Computing, University of Leeds, UK
d.c.hogg@leeds.ac.uk

Abstract

Social groups based on friendship or family relations are very common phenomena in human crowds and a valuable cue for a crowd activity recognition system. In this paper we present an algorithm for automatic on-line inference of social groups from observed trajectories of individual people. The method is based on the Social Force Model (SFM) – widely used in crowd simulation applications – which specifies several attractive and repulsive forces influencing each individual relative to the other pedestrians and their environment. The main contribution of the paper is an algorithm for inference of the social groups (parameters of the SFM) based on analysis of the observed trajectories through attractive or repulsive forces which could lead to such behaviour. The proposed SFM-based method shows its clear advantage especially in more crowded scenarios where other state-of-the-art methods fail. The applicability of the algorithm is illustrated on an abandoned bag scenario.

1. Introduction

Social groups are a very common phenomena in human crowds with empirical studies suggesting that about 74% of people come in a group to a social event [3] and about 50-70% (depending on the environment) are in a group during casual walking [18]. Despite this high percentage, the prevailing crowd behaviour models in today's simulation tools [6], computer graphics applications [16] and in particular in activity recognition and computer vision [2] are based on modelling each individual independently.

In this paper we propose an on-line algorithm for automatic detection of social groups in the crowd based on the analysis of the way the social relations influence the walking behaviour of the group members. The method is based on the Social Force Model (SFM) [13, 8] widely used in the crowd simulation community. In SFM, each individual's movement is influenced by several forces as illustrated in Fig 1, where the observed behaviour (*i.e.* the trajectories) differs depending on the presence or absence of the group
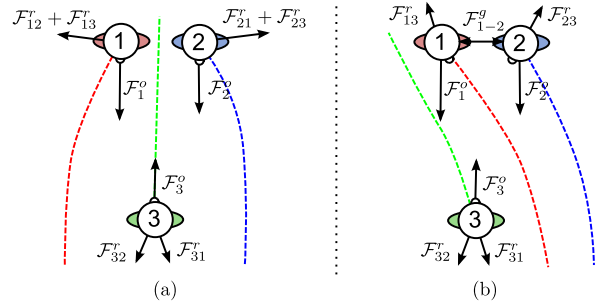

Figure 1. Depending on whether the individuals 1 and 2 (a) do not know each other or (b) know each other, Social Force Model produces different sets of trajectories combining together repulsive ($\mathcal{F}^r$), goal directed ($\mathcal{F}^o$), and group ($\mathcal{F}^g$) forces influencing the individuals.

relation (and thus the group or repulsive forces) between individuals. The proposed algorithm solves the inverted problem: knowing the trajectories, what are the social forces, and thus the relations, which caused that behaviour.

The algorithm is applicable in various fields. Think for instance of a scenario similar to that of PETS 2006 [2] where a family or a group of friends comes together and one of them leaves his/her bag with the others and separates from the group to buy a newspaper. Any threat detection system treating the individuals independently would inevitably report an abandoned bag as the criteria specified in PETS (the owner is further than a metres for more than b seconds) is fulfilled. An example of such system, but using social group inference, is given in the experiments section.

A system similar to the one above can also help to find parents of a lost child. A fully automatic group relations detection would also be valuable for calibration of crowd simulation models [17], for improved realism of computer generated crowd animations [11], for improving tracking [15] or to help robots to navigate better in a crowd.

Related Work

Although there exists a large body of work on group activity recognition [19], the focus is usually on well defined *collaborative* actions performed by more than one individual. In contrast, we are interested in *social* groups, or "who

knows who", rather than collaborative relations. We therefore limit the related work review to the approaches tackling social group modelling and inference only.

Modelling group behaviour. As noted recently [6], most of the commercial simulation tools lack ability to model social groups or their ability to do so is very limited. One approach for modelling walking people which has this ability is the Social Force Model (SFM) [8] and its variants. The original SFM does not actually contain a *de facto* group force. The model contains an attractive force, however it is mainly used for places and temporal relations (*e.g.* approaching an artist on the street) rather than a proper group formation principle. In [4] the model was extended to include groups and each individual is further assigned dependence and altruism levels which influence individual's behaviour. Several more forces were incorporated to the SFM in [10] to allow more realistic animation of agents engaged in a conversation.

However, all these attempts are based on human designed forces without proper evaluation. The model was only recently calibrated on real-world video sequences resulting in a model which realistically predicts avoidance behaviour of a walking group [12, 18] and later in a model with all its parameters, including group behaviour, estimated from real data [13].

Inferring social group relations. We are aware of only two approaches aiming explicitly at social group inference [9, 7] and one paper using social groups to improve tracking [15]. In [9], the groups are reported when two individuals keep close enough for a significant fraction of frames in a certain time window. The groups are further classified as voluntarily and involuntarily by computing the so called Perceived Personal Space, *i.e.* the free space in the visual field of the individual, using Voronoi diagrams. However, as we show in our experiments, such simple measures are not sufficient for reliable group inference in complex scenes. Moreover, the method suffers from separation of the inference into two sub-problems – grouping and voluntariness, which are naturally related and influence each other.

Similar distance and direction-based measurements were used in [15] to improve tracking by jointly tracking and inferring the social groups. However, the inference is very slow and even incorrect groups could still be accepted if it improves tracking. In contrast, we are interested in the online inference of the true social relations among individuals based on their trajectories during casual walking.

Based on distance as well, but including also the velocity vector difference, the method proposed in [7] applies agglomerative clustering to the (whole) trajectories. The merging criterion takes into account the fraction of frames in which the individuals are seen close to each other and allows the addition of a person to the group only if they have

been close to at least half of its members. Although the paper mentions a sliding window variant of the approach, it offers no strategy for combining the sliding window results into a globally consistent explanation. Our experiments with the method also indicate that the method needs rather large temporal window to cluster the trajectories without too many false positives.

The paper is structured as follows. We start by giving the necessary details of the SFM in Sec 2. It is then used in Sec 3 to formulate the group inference problem as a prediction error minimisation. The proposed inference algorithm is described in Sec 3.1 and 3.2. The experimental evaluation is given in Sec 4 and we conclude in Sec 5.

2. Social Force Model

The Social Force Model (SFM) was originally proposed by Helbing and Molnar in [8] as a pedestrian simulation model based on analogies between gas particles or fluid movement and movement of a human crowd.

The SFM is a mathematical model describing movement of each individual considering the interactions with other individuals and the environment. Each individual i is driven by a (social) force $\mathcal{F}_i$. Although $\mathcal{F}_i$ is not a force in the physical sense (there is no action and reaction effect for instance, *i.e.* the force can be asymmetric), it is used as a real force to update the i-th individual's position $\mathbf{x}_i^t$ and velocity $\mathbf{v}_i^t$ as

$$\mathbf{v}_i^{t+s} = \mathbf{v}_i^t + sr^{-1}\mathcal{F}_i(X^t, \Omega^t) , \quad (1)$$
$$\mathbf{x}_i^{t+s} = \mathbf{x}_i^t + sr^{-1}\mathbf{v}_i^{t+s} , \quad (2)$$

where X^t comprises the positions of all individuals, s is a small frame difference and r is the video frame rate.

The parameters Ω^t of the social force $\mathcal{F}_i$ can be split into three groups, $\Omega^t = \{\Theta, \Phi^t, G^t\}$. The first group of parameters, Θ, are the parameters characteristic for human crowd dynamics (*e.g.* the way people avoid each other). These have been empirically estimated from observed behaviour in previous studies [13, 12] and are assumed to be fixed in the following (the values used are given in Appendix A). The second group of parameters, $\Phi^t = \{\phi_i^t\}_{i=1}^N$, represents parameters of each individual person like walking direction or current velocity. The third group of parameters is the group clustering, G^t. In the following, we will use $\mathcal{F}_i(X^t, G^t)$ instead of $F_i(X^t, \Omega^t)$ to indicate that the group clustering G^t is the unknown variable while the other parameters are either fixed or estimated from the positional data X^t.

The social force $\mathcal{F}_i$ itself is composed of four component forces

$$\mathcal{F}_i(X^t, G^t) = \mathcal{F}_i^o(X^t) + \mathcal{F}_i^w(X^t) + \\ \mathcal{F}_i^r(X^t, G^t) + \mathcal{F}_i^g(X^t, G^t) , \quad (3)$$
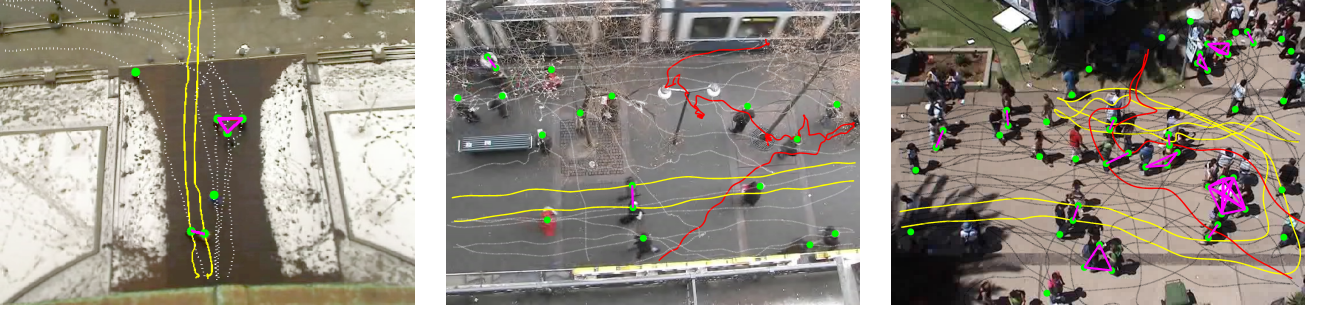
Figure 2. Snapshots from the ETH sequence (left), HOTEL sequence (middle) and the STUDENTS003 sequence (right). A number of trajectories are highlighted to demonstrate the complexity of the sequences. The straight thick lines connecting individuals indicate group links.

where $\mathcal{F}_i^o$ is the force driving the individual towards their desired goal, $\mathcal{F}_i^w$ is the repulsive force keeping the individual from hitting walls, $\mathcal{F}_i^r$ is the repulsive force from other individuals not in the same group, and $\mathcal{F}_i^g$ is the force keeping the individual together with other people in the same group. The reader is kindly referred to [13, 12] for the exact definitions of the forces. For the purpose of clarity only the following more detailed definition is necessary here:

$$\mathcal{F}_i^r(X^t, G^t) = \sum_j \mathcal{F}_{ij}^r(X^t, G^t) , \qquad (4)$$

where $\mathcal{F}_{ij}^r$ are repulsive force components between individuals i and j. By definition, $\mathcal{F}_{ij}^r$ is zero between members of the same group. Also $\mathcal{F}_i^g$ is zero for all individuals not engaged in any group. These properties will be exploited later in the clustering algorithm in Sec 3.1.

In the following section we formulate the group inference task based on the SFM, propose an agglomerative clustering algorithm for solving the task and describe the on-line strategy for group inference.

3. SFM Group Parameter Estimation

Instead of using the SFM for generating simulations, we see it as a generative model and try to infer its hidden parameters given the tracking observations. In particular, the goal is to estimate the group clustering parameters, G^t.

The input to the SFM group parameters estimation at time t is the tracking positional data $X^t = \{\mathbf{z}_i^t\}_{i=1}^N$ for each of N individuals, where each $\mathbf{z}_i^t$ denotes a sequence of positions in several past frames and a few "future" (decision lag) frames $\mathbf{z}_i^t = (\mathbf{x}_i^{t-\delta_v}, \ldots, \mathbf{x}_i^t, \ldots, \mathbf{x}_i^{t+s})$. Note that the current position $\mathbf{x}_i^t$ only is not sufficient as the per individual parameters Φ^t are also unknown and need to be estimated.

Knowing the group clustering, the social force $\mathcal{F}_i$ computed either using equation (1) or (3) would be the same[1].

---

[1]Assuming for the moment that the values of $\mathbf{v}_i^t$ and $\mathbf{v}_i^{t+s}$ are known too. In practice they are estimated and thus causing another uncertainty.

Not knowing the group clustering, we define the individual *prediction error* for a given clustering G^t as

$$\epsilon_i(X^t, G^t) = \| \mathcal{F}_i^r(X^t, G^t) + \mathcal{F}_i^g(X^t, G^t) +$$
$$\mathcal{F}_i^o(X^t) + \mathcal{F}_i^w(X^t) - \frac{\mathbf{v}_i^{t+s} - \mathbf{v}_i^t}{sr^{-1}} \| . \qquad (5)$$

Note that the term derived from equation (1) is fixed given the tracking data X^t as well as the terms $\mathcal{F}_i^o$ and $\mathcal{F}_i^w$. The individual prediction error can thus be expressed as

$$\epsilon_i(X^t, G^t) = \| \mathcal{F}_i^r(X^t, G^t) + \mathcal{F}_i^g(X^t, G^t) + K_i \| , \qquad (6)$$

where $K_i = \mathcal{F}_i^o + \mathcal{F}_i^w - (\mathbf{v}_i^{t+s} - \mathbf{v}_i^t)/sr^{-1}$ is a constant vector independent of the unknown group clustering. However, even though constant, it still influences the optimisation due to the norm operator.

The task of finding the best group clustering G^{t*} is then formulated as a minimisation problem

$$G^{t*} = \arg\min_G \sum_i \epsilon_i(X^t, G) . \qquad (7)$$

In contrast to other models [15, 7], the group term does *not* decompose into pairwise interactions as the group force depends on *all* members of the group. This, on one hand, enforces stronger constraints on group members requiring them to be consistent with the whole group formation and not relying on group relation transitivity only. On the other hand it makes the problem more complex.

One possible optimisation strategy for the task (7) would be to use MCMC with simulated annealing to search for a maximum of $p(G) \propto exp(\sum_i \epsilon_i(X^t, G))$. Albeit quite general, MCMC requires many sampling steps to represent the probability distribution sufficiently which makes it slow by design. In our initial experiments it also did not show significant improvement compared to the clustering method proposed in Sec 3.1 and thus was abandoned.

Another possibility would be to formulate the task using CRMF in a similar manner to [15]. Unfortunately, the

Algorithm 1: SFM Prediction Error Minimisation

Input: $X^t = \{\mathbf{z}_i^t\}_{i=1}^N$ *// tracking data*
Init: $G = \{\{1\}, \ldots, \{N\}\}$ *// no groups*
     $L^t = zeros(N, N)$ *// link confidence*
     $S = \{1, \ldots, N\}$ *// individuals to consider*

Compute Φ^t from X^t
SFM predict: $X^t, G, \Phi^t \rightarrow \epsilon_i, \mathcal{F}_{ij}^r$ *// eq. (5)*

while S not empty
     $I = \arg\max_{i \in S} \epsilon_i$
     $J = \arg\max_{j \in \{1,\ldots,N\}} \mathcal{F}_{Ij}^r$

     $G' \leftarrow G$ with clusters containing I and J merged

     SFM predict: $X^t, G', \Phi^t \rightarrow \epsilon_i', \mathcal{F}_{ij}'^r$ *// eq. (5)*

     $\delta_{IJ} = \sum_i \epsilon_i - \sum_i \epsilon_i'$
     if $\delta_{IJ} > 0$ and $||\mathbf{x}_I^t - \mathbf{x}_J^t|| < \theta_{dist}$
         $G = G'$
         update L^t
         $\epsilon_i = \epsilon_i'$
         $\mathcal{F}_{ij}^r = \mathcal{F}_{ij}'^r$
     end

     $S = S \setminus \{I\}$
end

Output: link confidence matrix L^t

---

algorithm proposed in [15] was already too slow for real-time processing and we are dealing with a more complex problem due to the lack of pairwise decomposition here.

It should be also noted that we are not interested in solving the problem (7) for a single time step t only but would like to take into account the temporal evolution of the observations. This problem will be discussed later in Sec 3.2.

3.1. SFM Prediction Error Minimisation

We first focus on solving the problem (7) for a single time step t. The proposed algorithm (see Algorithm 1) is effectively an agglomerative clustering method with some domain specific modifications.

The main difference to a generic agglomerative clustering method is that the output of the algorithm is not the clustering G itself but the link confidence matrix L^t instead, i.e. the matrix of confidence value for each pair of individuals indicating how strong the evidence was for assigning the individuals into the same group. The rationale behind is that in many situations the individuals' actions are at the edge of group-like and individualistic and thus only slight variations decide the final clustering. Using the decision confidence thus help to distinguish these cases from really strong representatives of group-like behaviour.

The algorithm is given the tracking data X^t as an input and starts by assigning each individual into their own cluster in G, setting the link confidence matrix L^t elements to zero and by creating list of individuals S.

As the individual's parameters Φ^t do not change during the clustering procedure, they are precomputed from the tracking data X^t first. The individual parameters $\phi_i^t = \{\mathbf{v}_i^t, \mathbf{e}_i^0, \mathbf{v}_i^0 \, \mathbf{v}_i^{t+s}\}$ are estimated from the tracking data $\mathbf{z}_i$ as follows. The individual's current velocity $\mathbf{v}_i^t$ and the velocity $\mathbf{v}_i^{t+s}$ are computed as

$$\mathbf{v}_i^t = (\mathbf{x}_i^t - \mathbf{x}_i^{t-\delta_v})/(\delta_v r^{-1}), \tag{8}$$

$$\mathbf{v}_i^{t+s} = (\mathbf{x}_i^{t+s} - \mathbf{x}_i^{t+s-\delta_v})/(\delta_v r^{-1}). \tag{9}$$

Throughout the paper we use $\delta_v = 10$ to smooth the small variations in velocity direction due to noise in tracking data. The parameter s specifies how far in the future the SFM is asked to predict. If set too low (*e.g.* $s = 1$) the model will try to predict all the noise in the trajectory data, if set too high (*e.g.* $s = 75$), the model is no longer able to predict reliably the individual's behaviour in complex scenes and the inference causes larger decision delays. In the following we fix $s = 10$.

The goal direction $\mathbf{e}_i^0$ is set equal to $\mathbf{v}_i^t$ and the desired speed $\mathbf{v}_i^0$ is set to $\|\mathbf{v}_i^t\|$. As the algorithm tries to predict the group clusters only for small s, this simplification is acceptable. However, in case of a well structured scene with several typical exits, a goal inference algorithm could be used instead.

After setting Φ^t, the initial prediction error of all individuals (considered to be independent initially) is computed using equation (5). Not only the error itself is stored but also the individual components of the repulsive force $\mathcal{F}_{ij}^r$ are kept. Both are used to guide the clustering algorithm.

The main loop runs until all individuals have been examined for grouping. In each iteration a pair of individuals, (I, J), is selected as candidates for grouping. The first individual in the pair is the one with the largest prediction error. As the individual is not grouped yet (is still in S) the main reason for this high error could be only $\mathcal{F}_{Ij}^r$ for some j as $\mathcal{F}_I^g$ is zero (see equations (6) and (4)). Thus, the second individual selected is the one with maximum $\mathcal{F}_{Ij}^r$. Here, the search is over all individuals, not only those in S.

After the candidate pair is found, their corresponding clusters are merged resulting in a new clustering, G', which is used to compute new prediction errors ϵ_i' and new repulsive force components $\mathcal{F}_{ij}'^r$[2]. If the new clustering G' has lower prediction error than the original G, it is preserved, otherwise another attempt is made with a different pair.

Throughout the merging, the confidence associated with successfully linked pairs is kept in the matrix L^t. The con-

---

[2]Note that only the errors and forces for individuals from the two clusters containing the candidate pair need to be recomputed as the rest stays the same.

fidence is based on the prediction error decrement δ_{IJ}. For each pair it is computed as δ_{IJ}/m where m is the number of new "A knows B" links created by merging the two clusters. For instance, when the first individual is in a group of size three before merging and the second on in a group of size two, in total $2 \times 3 = 6$ new links are created and each is assigned confidence $\delta_{IJ}/6$ in the L^t matrix.

3.2. On-line Group Detection

Applying Algorithm 1 at each time step results in a sequence of link confidence matrices L^t. However, the relation between these confidences and the final "A knows B" decision, is not straightforward. Think for instance of a meet-talk-split event or an obstacle avoidance manoeuvre. The interactions consist of time intervals when the individuals behave independently and intervals when they exhibit clear group-like behaviour. See Fig 5 for a few examples of L^t evolution. Thus, instead of applying a global measure on L^t, we classify two individuals as belonging to the same group once they exhibit a group like behaviour over a *short* temporal window.

Let us define a group activity confidence $A_{i,j}(t,w)$ for a pair of individuals i and j at time t over a temporal window of length w as

$$A_{i,j}(t,w) = 1/w \sum_{q=t-w+1}^{t} L^q(i,j) \ . \qquad (10)$$

A pair (i,j) is classified as belonging to the same group if $A_{i,j}(t,w) > \theta_A$ for some t. We further assume transitivity of the "knows" relation (if i knows j and j knows k, i knows k). The mean over the window could be replaced by a more robust measure like median if needed, however, in our experiments mean seemed to be sufficient. The effect of the window length w on the results is examined further in the experiments section.

4. Experiments

Datasets. The proposed method is tested on three publicly available video sequences of various complexity: the ETH and HOTEL sequences from the BIWI Walking Pedestrians dataset [14] and the STUDENTS003 [1] sequence. All sequences come with a partial annotation and ground plane homography which is necessary for measuring distances.

The ETH sequence is 8m 40s long and contains 386 individuals and 81 groups. The sequence is rather sparse with only few interactions between individuals and rather straight trajectories. An example frame from the sequence with a few highlighted typical trajectories is shown in Fig 2, left.

The HOTEL sequences contains slightly more complex interactions among individuals. It is 7m 23s long, contains
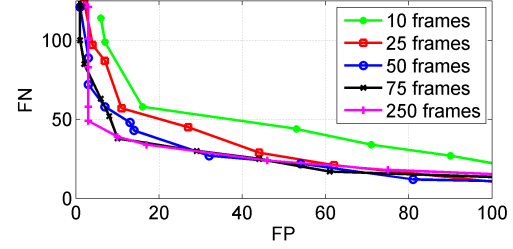

Figure 3. Varying the window size w on the STUDENTS003 sequence.

287 individuals and 55 groups. A typical frame from the sequence is shown in Fig 2, middle.

The STUDENTS003 sequence is 3m 40s long and contains 417 individuals and 112 groups (Fig 2, right). The crowd density is much higher compared to the ETH sequence and the behaviour exhibited is often much more complex as indicated by the highlighted trajectories.

All three sequences come with partial annotation only. We completed the trajectory annotation by marking the position of each individual every few frames and interpolating the rest of the frames using cubic splines.

Compared Methods. In the following experiments we compare the proposed algorithm with our implementation of [7] and also test against our adaptation of a part of the model proposed in [15] as detailed below. We do not show a comparison with [9] as they use a simpler version of the distance measure than the one proposed in [15] which, as will be shown below, is not sufficient for group inference in complex scenes. The method [7] is tested in an "off-line mode" as the strategy for combining sliding window decisions is not given and our experiments with the method have shown that too many false positives are generated if the decision are simply merged.

The method proposed in [15] optimises jointly over both the trajectories and group relations. The probability term for all trajectories H, their group assignment G, original image data I and model parameters Θ being optimised is

$$\begin{aligned} \log P(H,G,I|\Theta) &= E^{mo}(H|\Theta) + E^{app}(H|I,\Theta) \\ &+ E^{prior}(G|\Theta) + E^{pos}(H,G|\Theta) \qquad (11) \\ &+ E^{ang}(H,G|\Theta) + E^{tr}(G|\Theta) - \log Z(I,\Theta) \ , \end{aligned}$$

where the energy terms express respectively the expected motion and appearance, the prior on the number of groups in the scene, the expected position and velocity angle difference given the group/non-group membership and the group membership transitivity constraints. The final term normalises the probability to sum to one.

Having the trajectories fixed, only the terms dependent on G become relevant. Thus we define a distance for each pair h_i and h_j of trajectories in the current frame as a log
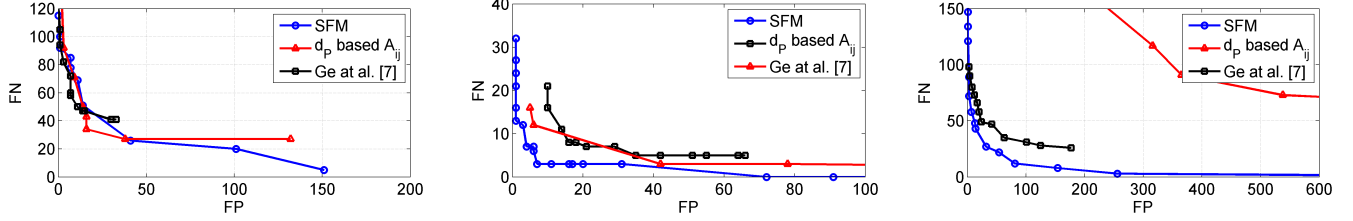
Figure 4. Pairwise links DET curve for the ETH (left), HOTEL (middle) and STUDENTS003 (right) sequences.

likelihood ratio

$$
\begin{aligned}
d_P(i,j) =& E^{pos}(h_i, h_j | g_{ij} = 1) + E^{ang}(h_i, h_j | g_{ij} = 1) - \\
& E^{pos}(h_i, h_j | g_{ij} = 0) - E^{ang}(h_i, h_j | g_{ij} = 0) \ ,
\end{aligned}
$$

where g_{ij} indicates whether the individuals i and j are in the same group or not.

The energy terms are computed over a temporal window as in the original paper and the group activity confidence A_{ij} is computed as in equation (10) with $d_P(i,j)$ substituted instead of $L(i,j)$. For both, the proposed method and the above adaptation of [15], we use $w = 50$ in equation (10). This value is justified by the experiment shown in Fig 3 where the results of the proposed method for STU-DENTS003 sequence are shown for varying window length. When set too low, many false positives are detected for a comparable false negative rate. On the other hand, even though the results do improve with longer window, the decision lag becomes significant for an on-line system. Choosing $w = 50$ is thus a reasonable compromise between precision and responsiveness.

Comparison. We evaluate the methods by calculating DET curves for pairwise links precision (Fig 4). Each method is run over the whole sequence and found groups are compared with the ground truth. Each pair of individuals is reported as a false positive if the individuals were assigned to the same group by the method but are not in the same group in the ground truth annotation and as a false negative when the method fails to assign them to the same group. Each method is tested over a range of values of its free parameter (θ_A threshold for the SFM and d_P-based group activity confidence, threshold on the length of interaction for [7]) to generate the DET curve.

As expected, for the relatively sparse and simple ETH sequence all three methods produce similarly good results as most of the groups are isolated in space and time. A larger difference in results can be seen for the HOTEL and STUDENTS003 sequences which are much more crowded and complex. The grouping based on d_P fails here as the distance and velocity based measure is not strong enough to distinguish between group and non-group relations and produces many false links between groups, especially in the STUDENTS003 sequence. The method [7] still performs quite well, but using the whole trajectories – when run on-
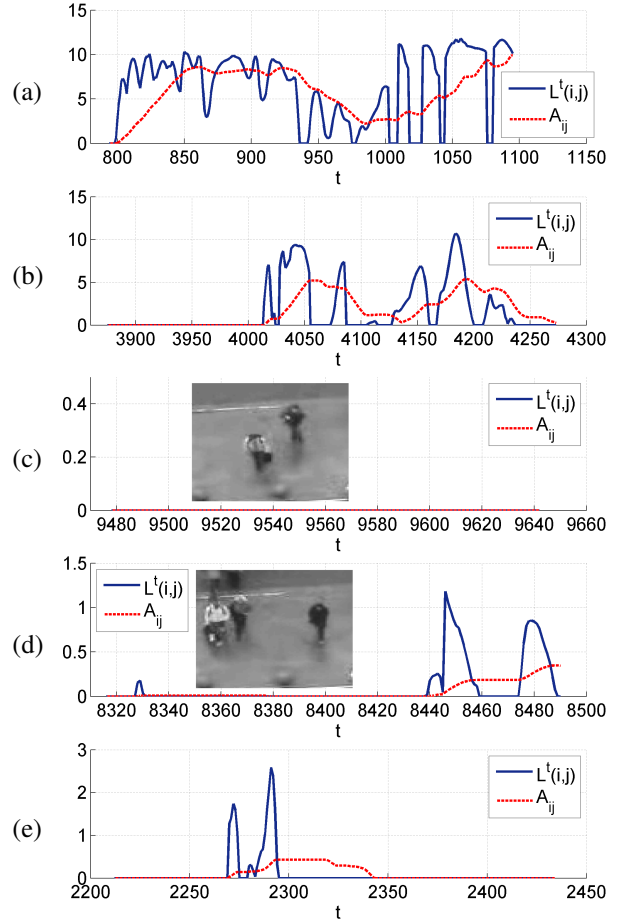

Figure 5. Link confidence $L^t(i,j)$ and group activity confidence A_{ij} for various pair interactions in the ETH sequence: (a) walking together (the highlighted pair from Fig 2, left), (b) meeting and leaving together, (c) walking close but slightly behind, (d) walking aside but a bit too far, (e) overtaking.

line the number of false positives increases considerably. The best results in this complex scenarios are achieved by the proposed SFM-based method. The success of the proposed method could be attributed to the fact that it not only measures proximity of the individuals but it also takes into account the context in which the proximity occurred. The method runs real-time at about 100fps.

Analysis. To gain intuition into robustness of Algorithm 1, Fig 5 depicts the link confidence $L^t(i,j)$ and the

group activity confidence A_{ij} for several pairs from the ETH sequence. First, the plot in Fig 5a shows a typical confidence history for a couple walking together through the scene without strong interactions with other individuals. The drops in $L^t(i, j)$ indicate moments when the individuals move apart slightly due to gesticulation or too strong waddling gait. In particular, during the period of low confidence between frames 934 and 1000 one individual attempts to enter the building but stops and returns as the other one does not follow. They then both stand close to each other in front of the building for another 4s before entering the building together.

A more complex group behaviour is shown in Fig 5b. Here two people meet in the middle of the scene and after a short greeting walk out of the scene together with slightly higher level of independence.

The next two plots in Fig 5 demonstrate the model's sensitivity to relative position of the group members (embedded is a snapshot of the situation). In the plot (c) one individual follows another very closely but a bit behind and to the side. Even though their distance and velocity difference stays small for most of the time, the relative position indicates they do not know each other. Similarly in the plot (d) the individuals keep similar distance and velocity for most of the observed frames and even walk next to each other but the distance is big enough to distinguish the fact that they do not belong to the same group. Only when entering the building they get closer so the confidence increases slightly.

Finally, a dynamic interaction between two individuals is depicted in Fig 5e. Here one individual overtakes the other and due to the lack of space needs to get very close. However, even though they are temporarily adjacent, the interaction is again assigned low group activity confidence.

An important aspect of the algorithm, which helps in these situations, is the exploitation of the link confidence L^t. For instance in the overtaking or walking aside examples, if *any* positive value of L^t would mean a positive response, the accumulation of this weak evidence could easily trigger a false grouping of the individuals together.

Abandoned bag detection. Fig 6 depicts series of our initial tests on the abandoned bag scenarios. The sequences were tracked using our implementation of [5]. We augmented the abandoned bag rule from PETS [2] to "a bag is abandoned if the owner *or someone the owner knows* is not within a meters from the bag for more than b seconds". The scenarios are staged so that the original rule gives false alarms in sequences (b) and (d) as the owner leaves his bag unattended for more than 30 seconds. The proposed method was able to identify the group relations, indicating the proposed approach can help to reduce the number of false positive alarms in this surveillance task.

5. Conclusions

We presented an on-line algorithm for social group inference from trajectories of multiple individuals. The social group inference is formulated as a SFM prediction error minimisation. The proposed agglomerative clustering algorithm compares the effects of applying the group or repulsive forces and groups only individuals for which the group forces approximate the tracking data better.

The algorithm was tested on three publicly available sequences with increasing complexity showing its potential especially in more crowded scenes with complex interactions between individuals. Its applicability was also demonstrated using the abandoned bag scenarios involving social interactions.

Appendix A

The goal, wall and group forces were implemented according to the formulae given in [13] and the repulsive force according to equation (4.5) in [12]. The values of the parameters Θ used in our implementation of the Social Force Model are listed in Table 1 together with references to the source of those particular values. For further details and the meaning of each parameter see the referenced papers. In most cases the empirically estimated values used by the referred papers were adopted except for the repulsive force parameters where the distribution means are used.

A few modifications of the SFM for some limiting cases were also necessary for inference purposes. The main reason for these modifications is that for simulation the group relations are known and thus the proper forces are applied, whereas in the inference problem both group and repulsive forces are tested.

First, a modification of the repulsive force is needed as the repulsive force from [12] works only for people walking towards each other. Since the repulsive force influences avoidance behaviour which is desirable even in cases like overtaking, we make the force symmetric by adding $\pi/2$ to the angle between $\mathbf{v}_i$ and $\mathbf{v}_j$ when it is smaller than $\pi/2$.

It is also necessary to test if people walk in the opposite direction when computing f^{vis} part of the group force (see [13] for the definition). Otherwise, every two people facing each other are considered as a strong candidates (f^{vis} being zero) for grouping. To avoid this problem, we simply skip all the candidates for grouping with velocity angle difference above $\pi/2$.

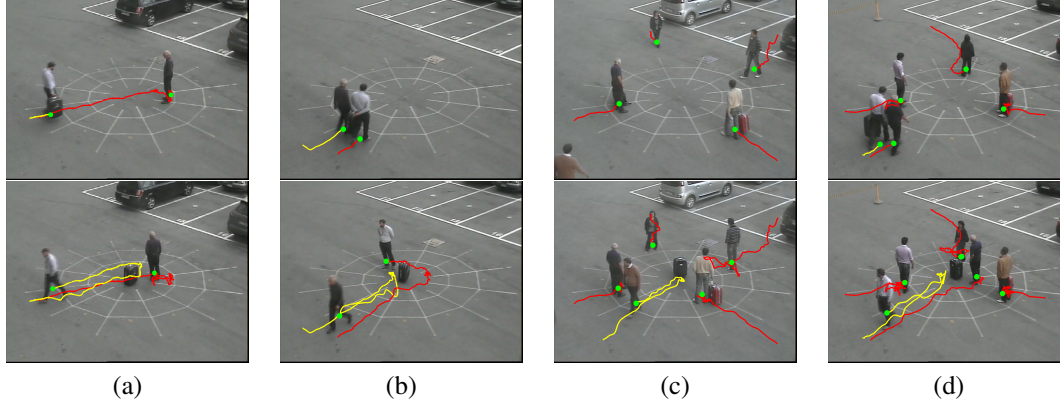<center>(a)   (b)   (c)   (d)</center>

Figure 6. Snapshots from four sequences of abandoned bag scenarios. In each sequence an individual drops a bag in the middle and leaves the scene. In sequences (a) and (c) the individual comes alone and in (b) and (d) with a companion. In all cases the proposed method was able to correctly find the social groups and thus allowed correct abandoned bag detection.

General parameters	$d_0 = 0.7$
Wall repulsive force [12]	$a = 10, b = 0.1$
Group force [13]	f^{vis}: $\beta_1 = 4, \phi = \pi/2$ f^{att}: $\beta_2 = 3$ f^{rep}: $\beta_3 = 1$
Repulsive force [12]	$A = 4.5, \gamma = 0.35, \lambda = 2$ $n = 2, n' = 3$

Table 1. Θ parameters values used in our implementation of SFM.

References

[1] http://graphics.cs.ucy.ac.cy/downloads/crowds-software-and-data. 5

[2] *Ninth IEEE International Workshop on Performance Evaluation of Tracking and Surveillance (PETS)*, New York, USA, June 18 2006. 1, 7

[3] A. F. Aveni. The not-so-lonely crowd: Friendship groups in collective behavior. *Sociometry*, 40(1):96–99, 1977. 1

[4] A. Braun, S. R. Musse, L. P. L. de Oliveira, and B. E. J. Bodmann. Modeling individual behaviors in crowd simulation. *International Conference on Computer Animation and Social Agents*, 0:143, 2003. 2

[5] M. D. Breitenstein, F. Reichlin, B. Leibe, E. Koller-Meier, and L. V. Gool. Online multi-person tracking-by-detection from a single, uncalibrated camera. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 99, 2010. 7

[6] R. Challenger, C. W. Clegg, M. A. Robinson, and M. Leigh. Understanding crowd behaviours: Simulation tools. Technical report, University of Leeds, 2009. 1, 2

[7] W. Ge, R. Collins, and B. Ruback. Automatically detecting the small group structure of a crowd. In *Workshop on Applications of Computer Vision*, pages 1–8, 2009. 2, 3, 5, 6

[8] D. Helbing and P. Molnar. Social force model for pedestrian dynamics. *Physical Review E*, 51:4282, 1995. 1, 2

[9] J. Jacques, A. Braun, J. Soldera, S. Musse, and C. Jung. Understanding people motion in video sequences using Voronoi diagrams. *Pattern Analysis & Applications*, 10:321–332, 2007. 2, 5

[10] D. Jan and D. R. Traum. Dynamic movement and positioning of embodied agents in multiparty conversations. In *Proceedings of the 6th international joint conference on autonomous agents and multiagent systems*, pages 1–3, 2007. 2

[11] A. Lerner, E. Fitusi, Y. Chrysanthou, and D. Cohen-Or. Fitting behaviors to pedestrian simulations. In *Proceedings of the ACM SIGGRAPH/Eurographics Symposium on Computer Animation*, August 2009. 1

[12] M. Moussaïd, D. Helbing, S. Garnier, A. Johansson, M. Combe, and G. Theraulaz. Experimental study of the behavioural mechanisms underlying self-organization in human crowds. *Proceedings of the Royal Society B: Biological Sciences*, 276(1668):2755–2762, 2009. 2, 3, 7, 8

[13] M. Moussaïd, N. Perozo, S. Garnier, D. Helbing, and G. Theraulaz. The walking behaviour of pedestrian social groups and its impact on crowd dynamics. *PLoS ONE*, 5(4):e10047, 04 2010. 1, 2, 3, 7, 8

[14] 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. 5

[15] S. Pellegrini, A. Ess, and L. Van Gool. Improving data association by joint modeling of pedestrian trajectories and groupings. In *ECCV 2010*, volume 6311, pages 452–465. Springer Berlin / Heidelberg, 2010. 1, 2, 3, 4, 5, 6

[16] C. W. Reynolds. Flocks, herds and schools: A distributed behavioral model. In *Proceedings of the 14th annual conference on computer graphics and interactive techniques*, pages 25–34, 1987. 1

[17] C. Rudloff, T.Matyus, and D. S. See and. Can walking behavior be predicted? an analysis of the calibration and fit of pedestrian models. 90th Annual Meeting of the Transportation Research Board, January 2011. 1

[18] H. Singh, R. Arter, L. Dodd, P. Langston, E. Lester, and J. Drury. Modelling subgroup behaviour in crowd dynamics dem simulation. *Applied Mathematical Modelling*, 33(12):4408–4423, 2009. 1, 2

[19] P. Turaga, R. Chellappa, V. Subrahmanian, and O. Udrea. Machine recognition of human activities: A survey. *Circuits and Systems for Video Technology, IEEE Transactions on*, 18(11):1473 –1488, 2008. 1