Multi-camera tracking using a Multi-Goal Social Force Model

Riccardo Mazzon*, Andrea Cavallaro

Queen Mary University of London
Mile End Road, London E1 4NS, United Kingdom

Abstract

Tracking across non-overlapping cameras is a challenging open problem in video surveillance. In this paper, we propose a novel target re-identification method that models movements in non-observed areas with a modified Social Force Model (SFM) by exploiting the map of the site under surveillance. The SFM is developed with a goal-driven approach that models the desire of people to reach specific interest points (goals) of the site such as exits, shops, seats and meeting points. These interest points work as attractors for people movements and guide the path predictions in the non-observed areas. We also model key regions that are potential intersections of different paths where people can change the direction of motion. Finally, the predictions are linked to the trajectories observed in the next camera view where people reappear. We validate our multi-camera tracking method on the challenging i-LIDS dataset from the London Gatwick airport and show the benefits of the Multi-Goal Social Force Model.

Keywords: Multi-camera tracking, trajectory propagation, Social Force Model, London Gatwick airport dataset

1. Introduction

Wide indoor and outdoor sites are extensively monitored by networks of cameras whose Fields Of View (FOV) do not necessarily overlap, thus making the task of tracking a person across a network very challenging (Fig. 1). When dealing with multi-camera tracking, existing methods solve the trajectory association problem relying on a training phase to learn the relationships between camera pairs. Most algorithms are based on a minimization method in order to find the correspondences between trajectories from each camera in the network [1]. The minimization process usually aims of finding the best match between appearance and motion features of the target. Common strategies, that tackle this problem relying on appearance matching across cameras [2], can only be applied when people are well visible and recognizable. Other algorithms integrate appearance features with motion information and use traveling time and reappearance position within the next observed region as key features for the minimization process [3].

One of the first attempts to solve the multi-camera tracking problem is presented in [4], where Kettnaker and Zabih use a Bayesian formulation for path reconstruction in a non-overlapping camera network. Their main assumption is that one object can only be at one specific position at a certain time. Observation matching permits to obtain chains of observations between frames in order to create object trajectories across the different views. In a more recent work, Javed *et al.* [5] track across multiple cameras using pedestrian trajectories obtained from single-camera tracking in the observed regions and exploit the relationship between the FOV lines on the same common ground plane. The object motion across cameras is then estimated using a minimization of the Euclidean distance. Furthermore, Javed *et al.* [6] use inter-camera space-time and appearance probabilities to find an object in different cameras by maximizing the conditional probability of the corresponding observations. To match an object after it moved through non-observed regions, space-time and appearance models are learned and updated on-line. A further improvement of multi-camera tracking based on appearance and motion is presented in [1], where the Brightness Transfer Function (BTF) colors mapping between camera pairs is expected to be lying on a low–dimensional space. This lower dimensionality helps trajectory association that is performed by an optimization step on the available trajectories using the position and the appearance of the target. A similar problem is tackled in [7] where the appearance of the target is matched in the Consensus-Color Conversion of Munsell (CCCM) color space, the main paths are grouped by unsupervised clustering and the time needed for a target to go from one camera to the next is analyzed and learned by associating only potential targets. A different approach is presented in [3] where the appearance of people is matched across cameras using color, covariance matrix and Histogram of Oriented Gradients. The feature mapping across cameras is learnt on-line and the Hungarian algorithm is used to solve the association problem.

In the presence of non-observed areas, there are no direct measurements of a person that can be used to facilitate tracking across cameras. Predicting the exact position where a person exiting the FOV of a camera will appear in the FOV of the next camera is very challenging due to the presence of various barriers and potential interactions occurring in the non-

*Corresponding author
Email address: riccardo.mazzon@eecs.qmul.ac.uk (Riccardo Mazzon)
URL: http://www.eecs.qmul.ac.uk/~andrea/ (Andrea Cavallaro)

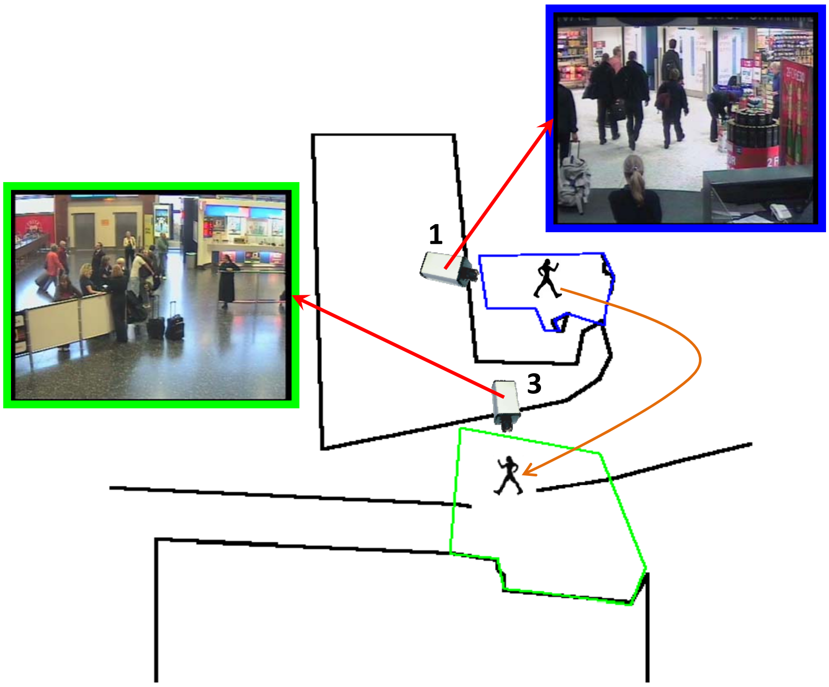


Figure 1: Example of person re-identification for multi-camera tracking. Top-view map of the London Gatwick airport [8]. The colored polygons indicate the FOV of Camera 1 (blue) and Camera 3 (green).

observed regions. Moreover, in the presence of a crowd, partial and complete occlusions will generate challenging situations for the above-described methods. Additional challenges are due changes in illumination conditions across cameras (e.g. the presence of a large window against an area with artificial illumination only), clutter (different people can look very similar) and different body poses.

In this paper, we tackle the multi-camera tracking problem by modeling the path of walking people without using appearance features. We predict where people move using a goal-driven model that creates hypotheses on where they are likely to reappear in order to facilitate the person re-identification process. To each person is assigned a set of possible goals [9], which are interest points in the site such as for example shops, doors, key points for the movement, exits, seats. In order to propagate people movements in non-observed regions, we use a motion model developed for crowd simulation [10]. Each person is modeled as an agent that can freely move onto the top-view map trying to reach the selected goals, avoiding barriers and walls while maintaining a desired speed. In order to tackle the multi-camera tracking problem, a matching process based on the spatio-temporal distances between predictions and single-camera tracking in the next observed region is performed. This process does not impose the assumption that points of view and illumination conditions are relatively consistent in the camera network. The main contributions of our work are: *a)* the use of a motion prediction model to estimate the positions of people in non-observed areas; *b)* the definition of multi-camera tracking as an on-line re-identification problem without using appearance features; *c)* the development of a simple parameter-based model for trajectory prediction that can be easily instantiated for a specific site. To the best of our knowledge, this is the first application and adaptation of a crowd simulation algorithm to a multi-camera tracking problem.

The paper is organized as follows: Section 2 discusses the related work on the field of motion modeling used for tracking. Section 3 presents the Social Force Model (SFM) and its modification for our goal-driven prediction. In Section 4 our model is validated using the i-LIDS dataset from the London Gatwick airport. Finally, Section 5 draws the conclusions and discusses future work.

2. Related work

We can identify three main strategies to categorize crowd simulation approaches based on how the relationships between pedestrians are modeled, namely macroscopic, microscopic and mesoscopic approaches [11]. *Macroscopic* approaches consider the crowd as an entity and movements are modeled as a flow that is followed by people. *Microscopic* approaches consider each person as an entity and the movement of each person is modeled by considering various factors such as interaction with other people and with the environment. Finally, *mesoscopic* approaches consider groups of people as entities and model their movements like a moving blob.

Macroscopic approaches are used for person tracking in high-density crowds, where individuals are difficult to be isolated. In this case the holistic crowd movements can be modeled as a flow. Hughes [12] defines crowds as *thinking* fluids and crowd movements are modeled by fluid attributes. This method was applied for dense crowd simulations at the exit of sport events [13]. Furthermore, Rodriguez *et al.* [14] use a macroscopic approach for tracking in crowds in unstructured environments where people can have heterogeneous movements. The crowd movements are modeled by Correlated Topic Model (CTM), where the topic is the high-level crowd movement and the word combination describes different motion patterns.

Ali and Shah [15] employ a mesoscopic approach in structured environments where dense crowds have homogeneous flows. Their method is based on *floor fields* and people appearance patches, applied to scenarios with cameras placed far from the observed scene.

Microscopic approaches are more suitable for modeling and predicting the movements of a single person. In [16] crowd simulation is performed by learning people movements from real sequences. Their model uses single-camera tracking results in order to obtain realistic crowd behaviors. One of the first applications of a microscopic model to a computer vision problem is reported in [17] where a Discrete Choice Model (DCM) is the basis of a low-complexity tracking algorithm aimed at following people in crowded scenarios. Single pedestrian movements are predicted in the next frame using a discrete grid and the prediction is performed by DCM tuned by a learning phase. Another microscopic approach for crowd modeling is the Social Force Model (SFM) firstly presented by Helbing and Molnar in [18] and subsequently refined in [19]. The SFM models the forces that guide a person toward a certain goal while avoiding barriers, walls, and other people. In [20] two escape scenarios are simulated using the SFM. The authors study the crowd simulation in order to understand how people behave in different situations and note that the average crowd density abruptly increases in the case of the closed exit, compared for example to the case of a person collapsing. SFM has also been used for abnormality detection. In [21] the SFM guides the movement of a set of particles spread in the scene where the interaction forces between the agents (in this case particles) are computed using optical flow. Abnormalities are detected by finding uncommon patterns on social interaction forces over time. Furthermore, SFM has been applied to single-camera tracking. In [10] the parameters for the SFM are learned from a set of tracking results and the model is applied in simple scenarios where the detection task is already solved. Another tracking method that make use of the SFM is reported in [22], where it is demonstrated how single-camera tracking can perform better if the motion model follows a minimization process of the social forces involved in the SFM. It is important to note that although [21] and [22] modify the SFM, they do not consider forces due to the environment since in their scenarios no obstacles or walls that constrain people movements are present.

A different microscopic approach for single-camera tracking is presented in [23] where Pellegrini *et al.* define the Linear Trajectory Avoidance (LTA) method. This method differs from SFM because, instead of defining people movements using energy potentials, an *expected point* where people are likely to move to is used and a global optimal solution is found in order to assign the next step to each target. An improvement of this method is reported in [24] where the stochastic LTA (sLTA) is introduced. Compared to the original LTA the final decision is based on a mixture of Gaussians that describes where people are likely to move. Vasquez *et al.* [25] present an approach based on Growing Hidden Markov Model (GHMM) to predict the goal of a moving target by studying its movements. The proposed algorithm considers the site map divided by a Voronoi diagram. The learning and prediction steps of the GHMM are calculated on-line using information from the available observations. This work provides one of the first attempts of long-term (but not instantaneous as in tracking) people movements prediction toward a goal. However the algorithm is only developed for single-camera scenarios.

3. The multi-camera multi-goal SFM

3.1. Overview of the proposed approach

In this paper we develop a modified Social Force Model for multi-camera tracking. The multi-camera tracking problem is formulated as an on-line target re-identification problem where one person exiting from one camera view is identified in the next camera view (where observations are available again), after having crossed non-observed areas. We assume to be known an approximate map of the environment and we integrate it with a modified Social Force Model [19] to model the behavior of walking people toward different goals within the map. We initialize the SFM with information from one observed region and then we let the model propagate the path within the non-observed areas, based on a set of interest points (goals) and barriers. The best predicted path is then selected based on the available information in the next observed region where a target reappears.

Let $\mathcal{M}$ be a top-view map of the site under surveillance that includes areas observed as well as non-observed by the FOVs of M cameras $C_1, C_2, \ldots, C_M$ used to monitor the area. Observed areas are mapped in $\mathcal{M}$ by homography projection [26]. Let $(x, y) \in \mathcal{M}$ be a point in the top-view. Let N people $P_1, P_2, \ldots, P_N$ walk onto $\mathcal{M}$ and let $\mathbf{p}_i(t) = (x_i(t), y_i(t)) \in \mathcal{M}$ be the position of person P_i at time t. Finally, let $B \in \mathcal{M}$ be the set of points $\mathbf{p}_B = (x_B, y_B)$ corresponding to barriers and walls that people can not cross.

We indicate with $\mathbf{p}_i^c(t) = (x_i^c(t), y_i^c(t)) \in \mathcal{M}$ the position of person P_i within the FOV of camera C_c, where $t \in [T_{start_i^c}, T_{end_i^c}]$ is the time interval during which P_i is visible in C_c. Without loss of generality, we consider camera C_1 to be the first camera when the person appears in the scene (i.e. we know $\mathbf{p}_i^1(t)$ with $t \in [T_{start_i^1}, T_{end_i^1}]$). When $t > T_{end_i^1}$ we assume P_i is not in the FOV of any camera and we start estimating the movement of P_i. In particular, we define $\Psi_i^* = \{\mathbf{p}_i^{*j}(t)\}$ where $\mathbf{p}_i^{*j}(t) \in \mathcal{M}$, $j = 1, 2, \ldots, N_{\Psi_i^*}(t)$, and $N_{\Psi_i^*}(t)$ is the number of position hypotheses at time t where person P_i is likely to walk.

Since in a complex site people have different goals to reach and hence different behaviors, a unique fixed goal for all the people is not a good model for the estimation of people behavior [9]. We tackle this problem by introducing a Multi-Goal Social Force Model (MG-SFM). We spread in the scene $|G|$ different goals that correspond to interest points in the site, such as shops, cafeterias, exits, seats, etc.

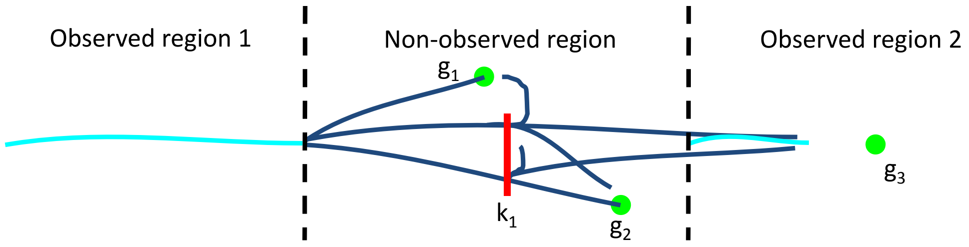


Figure 2: Schematic representation of the evolution of a path using the proposed approach. Cyan: trajectories in the observed regions. Green: goals. Red: key regions where new predictions are generated toward the goals. Blue: predicted trajectories toward the goals.

For each $\mathbf{p}_i^{*j}(t) \in \Psi_i^*$ a goal $\mathbf{g}_i^j$ is assigned where $\mathbf{g}_i^j \in G$ and G is the set of possible goal positions (interest points) onto $\mathcal{M}$. $\mathbf{p}_i^{*j}(t)$ and $\mathbf{g}_i^j$ will be considered as a pair in the rest of the paper. As it is difficult to exactly define the desired goal of each person over time, we generate hypotheses of people movements by introducing a set of new predictions toward G when the already existing trajectories in Ψ_i^* reach key regions in the environment (i.e. a crossing of possible paths selected using the map of the environment). We define these key regions as $\mathbf{k}_1, \mathbf{k}_2, \ldots \mathbf{k}_K$, where K is the number of key regions in $\mathcal{M}$.

Without loss of generality, we assume C_2 to be the next camera where person P_r is visible ($T_{start_r^2}$ is the time step when P_r reappears). We consider all the predictions $\mathbf{p}_i^{*j}(t)$ at time $t \in [T_{start_r^2} - \Delta_t, T_{start_r^2} + \Delta_t]$, where Δ_t is a time interval, and we set their next goal to $\mathbf{p}_r^2(T_{start_r^2})$. Then we let the predictions evolve over time along with the observed trajectory (the new goal) $\mathbf{p}_r^2$ for T_{proj} frames. Finally, from all $\mathbf{p}_i^{*j}$ we select the closest prediction in space to $\mathbf{p}_r^2$ in order to re-identify P_r (ideally P_r is re-identified with P_i when they represent the same person).

Figure 2 shows examples of predictions obtainable with the proposed approach: the algorithm finds the next position of a pedestrian starting from the observations in the first camera and uses this information to estimate the path a person is expected to follow when observations are available again in the next camera.

3.2. Multi-Goal Social Force Model

In order to estimate how person P_i moves in the non-observed areas we modify the Social Force Model [19]: each person is modeled as an autonomous agent that walks within the environment toward a specific goal, avoids barriers and walls, and maintains a desired speed. We assume that people crossing non-observed areas will maintain the speed they had in the previous camera view and that there are no interactions between people. Let each person P_i have mass m_i and be guided by the forces that describe the desired movements according to the surrounding constrains. We model an attractive force $\mathbf{f}_{iD}^{*j}(t)$ toward a specific goal and a repulsive force $\mathbf{f}_{iB}^{*j}(t)$ from walls and barriers. Finally, the displacement of P_i over time is defined by $d\mathbf{v}_i^{*j}(t)/dt$. The dynamics of the SFM is therefore formulated as:

$$m_i \frac{d\mathbf{v}_i^{*j}(t)}{dt} = \mathbf{f}_{iD}^{*j}(t) + \sum_B \mathbf{f}_{iB}^{*j}(t). \tag{1}$$

As abrupt movements of walking people are less likely to happen, we define a temporal smoothing process similar to the one reported in [22] in order to estimate the next step by considering the velocity[1] in the previous steps and actual forces. Compared to [22] we use a weighted average of the two components and we use more than only one previous step for smoothness:

$$\mathbf{p}_i^{*j}(t+1) = \mathbf{p}_i^{*j}(t) + \left(w \frac{d\mathbf{v}_i^{*j}(t)}{dt}\tau + (1-w)\overline{\mathbf{v}}_i^{*j}(t) \right), \tag{2}$$

where $\overline{\mathbf{v}}_i^{*j}(t) = \frac{\mathbf{p}_i^{*j}(t) - \mathbf{p}_i^{*j}(t-T_p)}{T_p}$ is the actual velocity calculated as the average velocity of the previous T_p frames, τ is the interval during which the variation of velocity is calculated. The magnitude of the displacement is directly proportional to τ. We fix $\tau = 1$ as we calculate τ at each time step. $w \in [0, 1]$ is the weight given to the actual velocity and $1 - w$ the one given to the previous velocity. The movement smoothness is inversely proportional to w and high values of w can lead to abrupt displacement of the target over time. Figure 3 shows different trajectory behaviors at varying w.

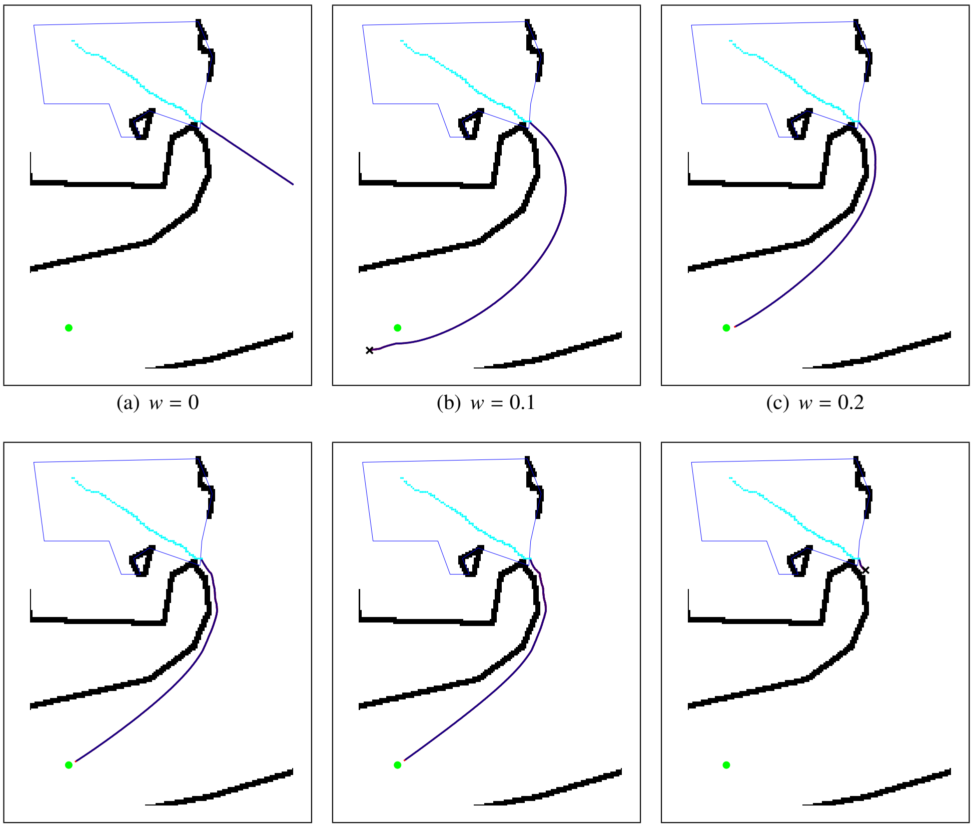
A goal is a point or an area of interest that would be reached at a desired speed following the minimum path, if there would not be any constrains such as walls and barriers. These desires are taken into account as:

$$\mathbf{f}_{iD}^{*j}(t) = m_i \frac{v_i^0 \mathbf{e}_i^{0*j}(t) - \overline{\mathbf{v}}_i^{*j}(t)}{\tau_i}, \tag{3}$$

where v_i^0 is the desired speed toward the direction $\mathbf{e}_i^{0*j}(t)$ of the goal to reach, and τ_i is the time relaxation parameter. $\mathbf{f}_{iD}^{*j}(t)$ is the force that pushes the target to reach the desired velocity by calculating the difference between desired and actual velocities. Note that v_i^0 does not depend on the specific prediction j but only on the desired speed of person P_i.

The desired speed v_i^0 is a key feature for our model. We have tested three different strategies for desired speed calculation using observations from the first observed region: the aver-

[1] Velocity is the 2D displacement of a point, while speed is the magnitude of the velocity.



(a) $w = 0$

(b) $w = 0.1$

(c) $w = 0.2$

(d) $w = 0.3$

(e) $w = 0.4$

(f) $w = 0.5$

Figure 3: Trajectory propagation examples generated using different values of w (see Eq. 2 for details) on the top-view map. Black: barriers. Cyan: trajectories from the observed region (FOV of the camera is the blue region). Purple: propagated trajectory. Green dot: goal to reach. Black cross: example of stopped prediction because its speed is too small.

age speed using the complete trajectory here referred to as MG-SFM-AVG; the maximum speed registered within a time interval of $2 * T_p$ (MG-SFM-MAX50); the maximum speed registered within a time interval of T_p (MG-SFM-MAX25). Results for the three strategies are reported in Fig. 8 and discussed in Sec. 4.

A monotonically decreasing force $\mathbf{f}^{*j}_{iB}(t)$ is also considered that acts from barriers and walls to each person [10]. As suggested in [10] we model this force with an inverted exponential proportional to the Euclidean distance $d^{*j}_{iB}(t)$ between person P_i predictions and barriers B. In addition to this, as walking people are influenced only by what happens in front of them [10], we restrict $\mathbf{f}^{*j}_{iB}(t)$ to the barriers in the range $[-90°, 90°]$ of the direction of motion of P_i and to the "visible" barriers from the actual position of the pedestrian. Figure 4(a) shows the range of influence of the barriers on a person and Fig. 4(b) reports a schematic representation of the influence of barriers force on people movements, formalized as:

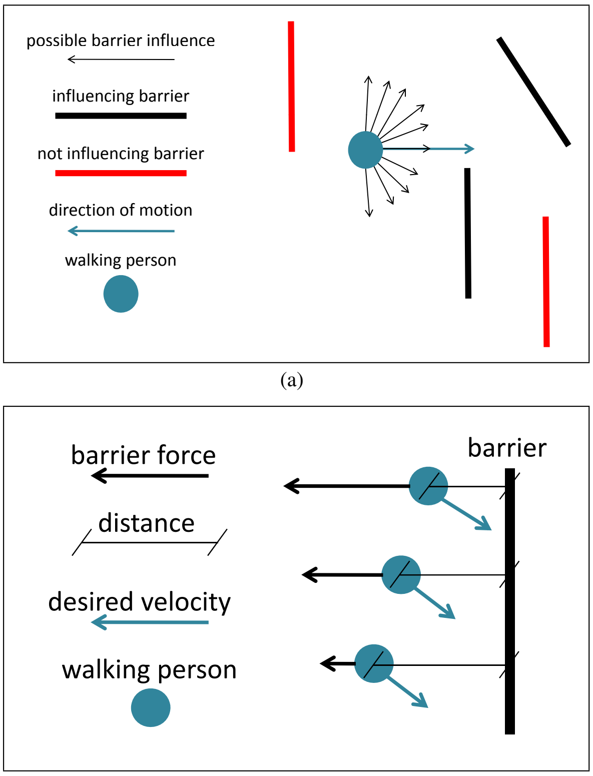
$$\mathbf{f}^{*j}_{iB}(t) = A_B e^{-\frac{d^{*j}_{iB}(t)}{B_B}}, \qquad (4)$$

where A_B is the weight associated to the barriers force (high values correspond to high repulsion force from the barriers), B_B is the interaction range that enlarges or reduces the area of influence of the barriers on people movements.

We predict how each person moves toward each goal using Eq. 2. At time step $T_{end^1_i} + 1$ (when person P_i is no more visible from camera C_1), we generate $|G|$ predictions toward each goal in G and we let them propagate onto $\mathcal{M}$. Since walking people change their view of the environment, it is likely that the direction of motion and their goal change over time. To model this behavior, multiple new predictions are further generated when an existing prediction reaches a key region $\mathbf{k}$. For instance, if prediction $\mathbf{p}^{*1}_1(\bar{t})$ toward goal $\mathbf{g}^1_1$ has reached the key region $\mathbf{k}_1$ at time $\bar{t}$, we generate $|G| - 1$ new predictions toward $G/\left\{\mathbf{g}^1_1\right\}$ (we exclude the goal already followed by $\mathbf{p}^{*1}_1(\bar{t})$), and we include[2] them in Ψ^*_1. The next step of MG-SFM removes from Ψ^*_i the predictions that do not appropriately model realistic scenarios. In particular, we remove each $\mathbf{p}^{*j}_i(t)$ with distance from its goal $\mathbf{g}^j_i$ less than $\epsilon_g > 0$, and we remove each $\mathbf{p}^{*j}_i(t)$ that corresponds to a prediction with speed $v^{*j}_i(t) < \epsilon_v * v^0_i$, where $v^{*j}_i(t) = |\bar{\mathbf{v}}^{*j}_i(t)|$ and $0 < \epsilon_v < 1$.

Figure 5 shows four examples of trajectory prediction in non-observed regions using the parameter setting explained in

[2] For the new predictions we include in Ψ^*_1 the same positions of $\mathbf{p}^{*1}_1(t)$ for $t = [T_{end^1_1} + 1, \bar{t}]$, and from $\bar{t} + 1$ onward we make the predictions toward the new assigned goals.



(a)

(b)

Figure 4: Influence of the presence of barriers on the movement of a person: (a) the influence is limited to the visible range $[-90°, 90°]$ in the direction of motion; (b) the force generated by a barrier is inversely proportional to the exponential of the distance from the barrier itself [10].

Sec. 4. Using the same parameter setting, we test our model in order to calculate the distance (in time and space) of our predictions with respect to frame step and position of person reappearance. Table 1 shows the results for MG-SFM-AVG, MG-SFM-MAX50, and MG-SFM-MAX25 calculated on 42 people going from one observed region to the next. For each person we consider the 60 closest predictions in time to the reappearance time step and we calculate the average distance to the reappearance position. The results are shown in columns 2-4 of Table 1. We see that for MG-SFM-MAX25, 81% of the predictions are within 20 units (as the radius of the green circle in Fig. 5). Furthermore, we analyze how synchronized our predictions are to the reappearance time step. We take the 60 closest predictions in space to the position of reappearance and we calculate the average difference with the time step of reappearance. Columns 5-7 of Table 1 show the complete results and we can see that over 50% of our predictions are within 25 frames (1 second on the used dataset) when applying MG-SFM-MAX25.

	Radius (units)			Time (frames)		
	5	**10**	**20**	**25**	**75**	**125**
MG-SFM-AVG	45%	57%	67%	31%	62%	71%
MG-SFM-MAX50	48%	59%	71%	33%	79%	86%
MG-SFM-MAX25	45%	62%	81%	52%	79%	90%

Table 1: Testing for MG-SFM predictions on 42 trajectories from the London Gatwick airport [8]. See text for the complete explanation of MG-SFM-AVG, MG-SFM-MAX50, and MG-SFM-MAX25. Columns 2-4: Average percentage of predictions within the indicated radius centered on the person reappearance position of the 60 closest predictions (in time) to the reappearance time step. Columns 5-7: Average percentage of time synchronization within the indicated frames between predictions and time step of person reappearance of the 60 closest predictions to the position of reappearance.

As predicting the exact position and the exact time instant when a person reappears is very challenging, when a generic person P_r, where $r = 1, 2, \dots, N$, appears in C_2 we consider good hypotheses for P_r all the predictions $\mathbf{p}_i^{*j}(t) \in \Psi_i^*$, where $i = 1, 2, \dots, N$, $j = 1, 2, \dots, N_{\Psi_i^*}(t)$, $t \in [T_{start_r^2} - \Delta_t, T_{start_r^2} + \Delta_t]$, and $\Delta_t = 2 * T_p$. Δ_t is chosen to be proportional to T_p (Eq. 2) in order to obtain a large enough time window for the final matching between predictions $\mathbf{p}_i^{*j}(t)$ and observations $\mathbf{p}_r^2(t)$. Note that P_i varies among all the available people trajectories within the specific time interval since we are now tackling the person re-identification problem. In particular, we propagate the predictions $\mathbf{p}_i^{*j}(t+t_r)$ toward $\mathbf{p}_r^2(T_{start_r^2}+t_r)$ with $t_r = 0, 1, \dots, T_{proj}-1$ using Eq. 2, where

$$T_{proj} = \min(T_{end_r^2} - T_{start_r^2} + 1, T_p). \tag{5}$$

We define $d_{ir}^{*j}(t)$ to be the Euclidean distance between $\mathbf{p}_i^{*j}(t + T_{proj})$ and $\mathbf{p}_r^2(T_{start_r^2} + T_{proj})$, and for each P_i we calculate

$$\chi_{ir} = \min_j \min_t \left(d_{ir}^{*j}(t)\right), \tag{6}$$

where j and t vary as defined above. By sorting χ_{ir} a ranking for the association of P_i to P_r is obtained as final result of the MG-SFM. Algorithm 1 reports the complete algorithm for the MG-SFM.

4. Experimental results

To validate the proposed method, we use the i-LIDS dataset from the London Gatwick airport [8] and we study the movement of people at the arrival terminal. We consider people that are visible when they walk out of the *passengers area*. The aim is to find where and when these people reappear in one of the next cameras in the *public area*. This is a challenging environment where people can potentially walk in many directions once they exit the camera view covering the passenger area. In addition to this, the movements may be constrained by barriers and people can not follow the shortest path to reach their desired goal. In the experiments, Camera 1 is the first observed region (C_1) and Camera 3 of the dataset is the second camera where people reappear (C_2). As we focus on the modeling of people movements in non-observed regions, in this paper we consider solved the single-camera tracking task. The top-view map $\mathcal{M}$ is shown in Fig. 1[3]. Table 2 summarizes the parameters used in the evaluation. A_B is set high and B_B is set to 1 in order to implement barrier avoidance while letting people move in the environment without strong influence. Using Eq. 4 it can

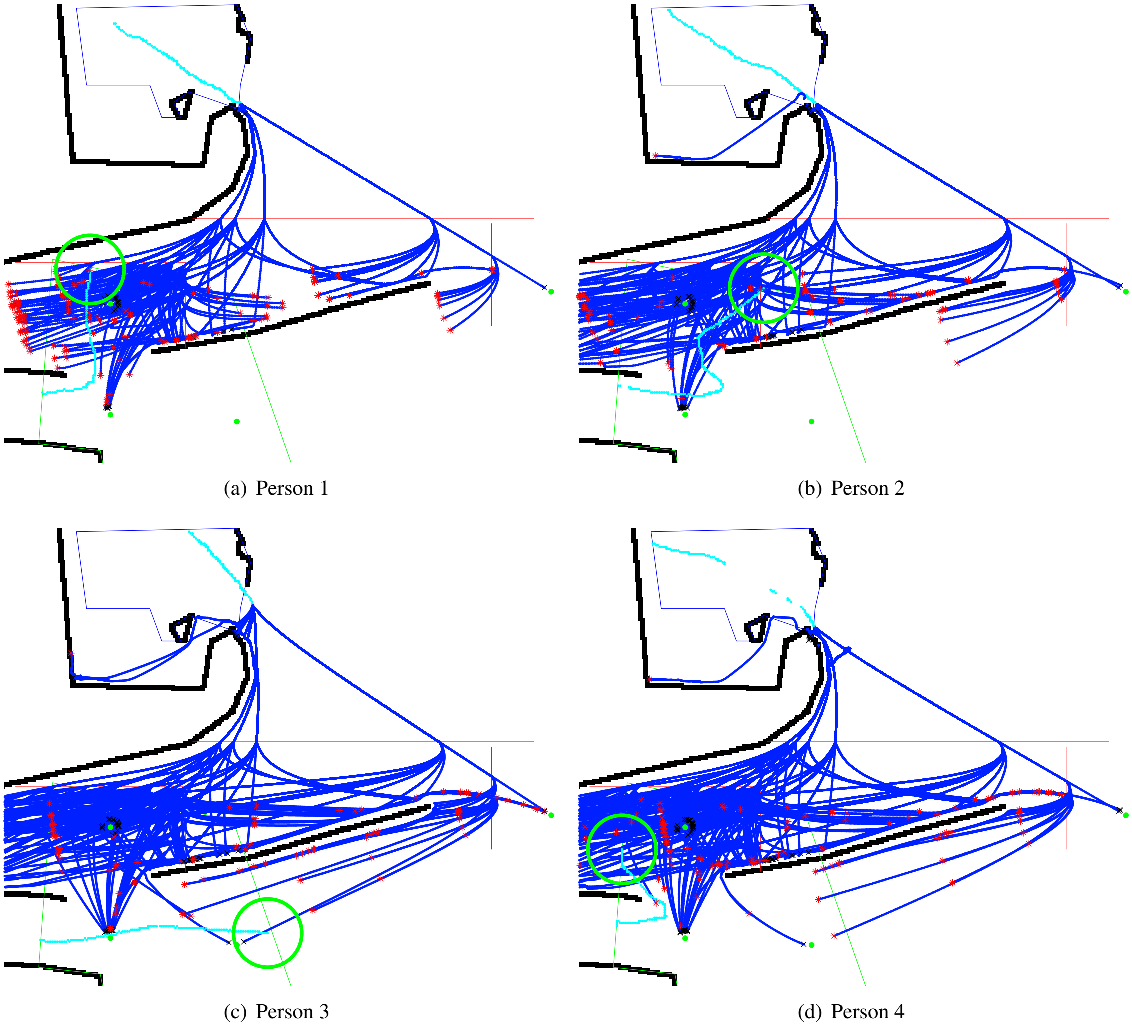
[3] Part of the map has been created using information from the London Gatwick airport website http://www.gatwickairport.com/.

Algorithm 1 MG-SFM for camera pairs

Define: map $\mathcal{M}$; set B of barriers position; goals $\mathbf{g} \in G$; set of key regions $\mathcal{K}$; parameters ϵ_v and ϵ_g; T: set of considered time steps; I: set of walking people; T_p: frames to consider for actual velocity; $\Delta_t = 2 * T_p$: frame interval for re-identification.
C_1: first observed region; $[T_{start_i^1}, T_{end_i^1}]$: time interval when person P_i is within the FOV of Camera 1;
$\mathbf{p}_i^1(t)$: position of person P_i at time t within Camera 1; v_i^0: desired speed of person P_i;
C_2: second observed region; $[T_{start_i^2}, T_{end_i^2}]$: time interval when person P_i is within the FOV of Camera 2;
$\mathbf{p}_i^2(t)$: position of person P_i at time t within Camera 2;
$\mathbf{p}_i^{*j}(t)$: predicted position of person P_i toward goal $\mathbf{g}_i^j$ at time t, $\mathbf{g}_i^j \in G$;
$\mathrm{D}\big(\mathbf{a}, \mathbf{b}\big)$: Euclidean distance between $\mathbf{a}$ and $\mathbf{b}$; $\min\big(a, b\big)$: minimum value between a and b;
$\mathbf{p}(1 \rightarrow t)$: positions from time 1 to time t;

```
for all $t \in T$ do
    for all $i|P_i \in I$ do
        if $t \in [T_{start_i^1}, T_{end_i^1}]$ then                                   ▷ First observed region
            obtain $\mathbf{p}_i^1(t)$ by single-camera tracking
        else                                                                           ▷ Non-observed regions
            if $t = T_{end_i^1} + 1$ then                                                ▷ Initialization of $\Psi_i^*$
                initialize $\Psi_i^* = \big\{\mathbf{p}_i^1(t)\big\}$
                $\Psi_i^*$ =ADDBRANCHES$\big(\Psi_i^*, G\big)$
            end if
            for all $j\,|\,\mathbf{p}_i^{*j}(t) \in \Psi_i^*$ do                                      ▷ Prediction step
                apply Eq. 2 to $\mathbf{p}_i^{*j}(t)$ (toward $\mathbf{g}_i^j$)
                $v_i^{*j}(t)$ = speed of $\mathbf{p}_i^{*j}(t)$
                if $\Big(t > T_{end_i^2} + T_p \wedge v_i^{*j}(t) < \epsilon_v * v_i^0\Big) \vee \Big(D\big(\mathbf{p}_i^{*j}(t), \mathbf{g}_i^j\big) < \epsilon_g\Big)$ then     ▷ Check for non-valid predictions
                    $\Psi_i^* = \Psi_i^* / \big\{\mathbf{p}_i^{*j}(t)\big\}$
                end if
                if $\mathbf{p}_i^{*j}(t)$ within $\mathcal{K}$ then
                    $\Psi_i^*$ =ADDBRANCHES$\Big(\Psi_i^*, G / \big\{\mathbf{g}_i^j\big\}\Big)$
                end if
            end for
        end if
    end for
    initialize $j_r = 1, \overline{\Psi}_i^* = \emptyset$
    for all $r|P_r \in I$ do                                                           ▷ Second observed region
        for all $t^* \in [T_{start_r^2} - \Delta_t, T_{start_r^2} + \Delta_t]\,|\,\exists \mathbf{p}_i^{*j}(t^*) \in \Psi_i^*$ do
            $T_{proj} = \min\Big(T_{end_r^2} - T_{start_r^2} + 1, T_p\Big)$
            for all $j\,|\,\mathbf{p}_i^{*j}(t^*) \in \Psi_i^*$ do
                $\overline{\mathbf{p}}_i^{*j_r}(1 \rightarrow t^*) = \mathbf{p}_i^{*j}(1 \rightarrow t^*)$
                $\overline{\Psi}_i^* = \overline{\Psi}_i^* \cup \big\{\overline{\mathbf{p}}_i^{*j_r}(1 \rightarrow t^*)\big\}$
                for all $t_r \in [0, T_{proj} - 1]$ do
                    apply Eq. 2 to $\overline{\mathbf{p}}_i^{*j_r}(t^* + t_r)$ (toward $\mathbf{p}_r^2(T_{start_r^2} + t_r)$)
                end for
                $j_r = j_r + 1$
            end for
        end for
    end for
    apply Eq. 6 to $\overline{\Psi}_i^*$
end for

procedure $\Psi$ = ADDBRANCHES($\Psi$,$G$)                  ▷ Procedure to add new branches for trajectory prediction
    $\Psi$: set of trajectory predictions; $G$: set of goal positions
    for all $\mathbf{p} \in \Psi$ do
        for all $\mathbf{g} \in G$ do
            create new $\overline{\mathbf{p}} = \mathbf{p}$
            associate $\overline{\mathbf{p}}$ to the goal $\mathbf{g}$
            $\Psi = \Psi \cup \overline{\mathbf{p}}$
        end for
    end for
end procedure
```



(a) Person 1

(b) Person 2

(c) Person 3

(d) Person 4

Figure 5: Examples of trajectory prediction for four people walking from Camera 1 (C_1) to Camera 3 (C_2) at the London Gatwick airport. Cyan: trajectory in the observed regions. Blue: predicted trajectories using MG-SFM-MAX25 (see text for details). Red star: predicted trajectories at the time step when the person reappears in C_2. Black cross: predicted trajectories that stop because they have reached the goal or their speed is too small. Red segment: definition of the key regions for splitting the predictions. Black segment: barriers. Green dot: goals. Green circle: 20 units of radius centered in the first observation in C_2.

be seen that the influence of the barriers on a person is negligible at a distance of about 10 units. We consider the mass m_i of each person to have the same value [19] and we set it to 70 Kg [27]. The actual velocity is calculated during the last 1 second of video (25 frames).

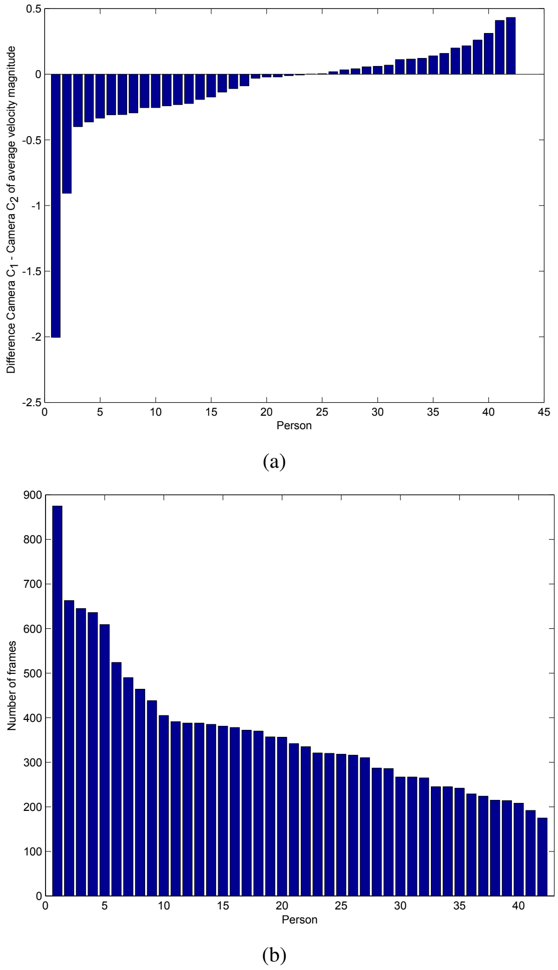
In order to understand how much variation there is in people movements, Fig. 6(a) shows the difference of the average speed (velocity magnitude) registered in C_1 and C_2 using ground-truth information on the top-view from 42 people. In addition to this, Fig. 6(b) shows the traveling time to go from C_1 to C_2. From the graphs we can see that people move at substantially different speeds in the two camera views and their traveling time can go from 7 seconds to 35 seconds.

We compare the MG-SFM for person re-identification with two methods based on the average traveling time of people from C_1 to C_2. Let TTALL be the first method that calculates the average traveling time of all people that go from C_1 to C_2, and considers this average as the expected traveling time of each person. This method is similar to the one proposed in [7] where people traveling time is used to make hypotheses for multi-camera tracking. Let TT4REG be the second method that divides C_2 in four entrance regions and calculates the average traveling time of people that only enter in the specific region. Similarly to TTALL, in TT4REG the average traveling time in

Parameter	size($\mathcal{M}$)	m_i	A_B	B_B	w	$\|G\|$
Value	577×961 *units*	70 Kg	60000 N	1 *unit*	0.3	8

Parameter	K	τ_i	T_p	ϵ_v	ϵ_g
Value	3	1 *frame*	25 *frames*	0.1	5 *units*

Table 2: Parameters of the proposed Multi-Goal Social Force Model (MG-SFM). $\mathcal{M}$: top-view map; m_i: person mass; A_B: weight associated to the barrier force; B_B: barrier interaction range; w: weight for actual velocity; $|G|$: number of goals; K: number of key regions; T_p: number of previous frames to calculate actual velocity; ϵ_v: value for low velocity thresholding; ϵ_g: number of units for goal reached thresholding.



(a)

(b)

Figure 6: Variations of people walking speeds from Camera 1 (C_1) and Camera 3 (C_2) at the London Gatwick airport [8] calculated on the top-view map. (a) Average speed difference in the two cameras; (b) traveling time to go from C_1 to C_2.

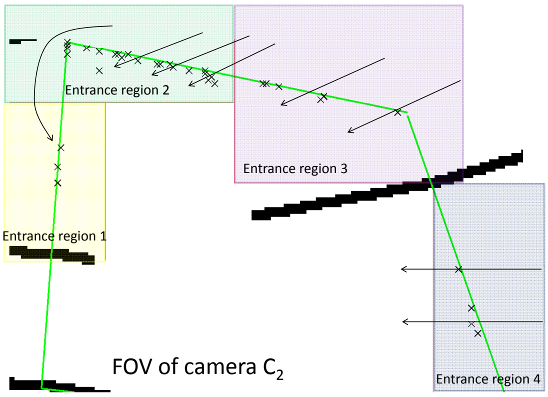


Figure 7: Field of view of camera C_2 (London Gatwick airport) and corresponding segmentation in entrance regions for TT4REG (see text for details). Crosses: people reappearance position. Arrows: possible motion direction.

each region is the expected traveling time of people that enter in that region. Figure 7 shows the four regions: crosses correspond to reappearance position of people and arrows correspond to possible direction of motion. Note that TT4REG is trained assuming known the region of reappearance of each person. For both TTALL and TT4REG, the traveling time is calculated by the difference between the frame when a person reappears in C_2 with the last frame when the same person is visible in C_1. We perform a ranking for person re-identification by calculating the absolute time difference between the time step when a person reappears and the results of TTALL and TT4REG. Since in the MG-SFM we consider only predictions within a time interval of $\pm\Delta_t$, in order to make a fair comparison we consider valid only (absolute) time differences lower than Δ_t (let us call the corresponding methods TTALL-50 and TT4REG-50) and, to make the comparison more challenging, lower than $2*\Delta_t$ (let us call the corresponding methods TTALL-100 and TT4REG-100).

For the MG-SFM we calculate the distance between the predictions of each person going out from C_1 and the trajectories that appear over time in C_2 (Sec. 3). Using Eq. 6 we then generate a ranking of the predictions for the person re-identification task.

Figure 8 shows the final results as a cumulative frequency graph (the ideal result is a horizontal line at value 1 that corresponds to having correct predictions for all people). Results are generated using the three different strategies for the calculation of people desired speed reported in Sec. 3.2: MG-SFM-AVG, MG-SFM-MAX50, and MG-SFM-MAX25. It is important to note here that the results are obtained *without* using appearance matching of the targets across cameras. MG-SFM-MAX25 outperforms TTALL and TT4REG, while MG-SFM-MAX50 outperforms them starting at position rank 2. With MG-SFM-MAX25 we obtain 50% of correctly re-identified people, compared to 41% of TT4REG-100, and 29% of MG-SFM-MAX50 and MG-SFM-AVG. Furthermore, if we consider the first 4 positions in the ranking we have 88% and 83% of correct re-identifications for MG-SFM-MAX25 and MG-SFM-MAX50, respectively. On the other hand, MG-SFM-MAX25 never reaches 100% in the re-identifications task because the method can not predict the behavior of a person who travels at an average speed in C_1, and then takes a long time to reappear in C_2 (more than 31% of the average traveling time of their reappearance region). In general, MG-SFM-MAX50 and MG-SFM-MAX25 better model people's desired speed as compared to MG-SFM-AVG. In fact, it is likely that the registered highest speed well describes the desired speed that a person would maintain if there would not be any constraints in the environment.

Finally, Fig. 9 shows the confusion matrix obtained with MG-SFM-MAX25, reporting the distances resulting from Eq. 6. It is interesting to note that person 12 and person 14 are re-identified with rank 2, and the distance between the best prediction and reappearance position is less than 3 units, hence very close to the correct re-identification (the green circle in Fig. 5 is 20 units). Difficult cases for our motion modeling are when people exit C_1 at roughly the same position and time step.

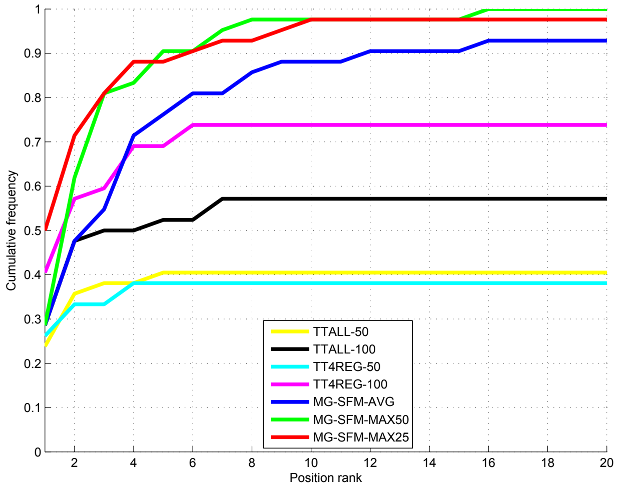


Figure 8: Cumulative frequency graph for person re-identification. X-axis: person re-identification ranking. Y-axis: frequency accumulation of the correct person re-identification ranking.

An example is person 21, 22, and 23. However, since these people exit at different velocities, we can still have rank 2 and 1 for person 21 and 22, respectively, because our model creates different predictions for each of them. A second example are person 37 and 38 that walk and exit together C_1 at approximately the same velocity, and reappear in the same region in C_2. In this case, the distance between predictions and observed trajectory is less than 7 units between the two and over 81 units from person 42: a wrong hypothesis for the re-identification. Furthermore, there is only one person (number 7) out of ranking because too far away in time and only two people (number 18 and 25) with the correct ranking values over 20 units. These results show how MG-SFM can well predict people movements in non-observed regions for the re-identification problem, and even in cases when the method can not perfectly solve the re-identification problem, it can give reasonable hypotheses on the position and the time of reappearance of a person.

5. Conclusion and future work

We presented a method to estimate people movements in non-observed regions between camera views and demonstrated it on a people re-identification problem without using appearance features on a real surveillance scenario. The method is based on a modification of the Social Force Model and takes into account barrier avoidance constraints as well as the desired motion toward specific goals in the scene. Unlike existing methods that assume a linear motion between cameras we only assume that a person will maintain roughly the same speed when traveling across cameras, but can change directions as a function of local goals and barriers. We showed that the proposed method outperformed an algorithm based on the average traveling time of people between cameras [7] on a standard challenging dataset.

As future work we will integrate a person tracking algorithm in the observed regions and we will test the proposed method on different scenarios with different types of obstacles in the non-observed areas.

References

[1] O. Javed, K. Shafique, Z. Rasheed, M. Shah, Modeling inter-camera space-time and appearance relationships for tracking across non-overlapping views, Computer Vision and Image Understanding 109 (2008) 146–162.

[2] B. Prosser, W.-S. Zheng, S. Gong, T. Xiang, Person re-identification by support vector ranking, in: Proc. of the British Machine Vision Conf., Aberystwyth, UK, pp. 21.1–21.11.

[3] C.-H. Kuo, C. Huang, R. Nevatia, Inter-camera association of multi-target tracks by on-line learned appearance affinity models, in: Proc. of Europ. Conf. on Computer Vision, Hersonissos, Crete, Greece, pp. 383–396.

[4] V. Kettnaker, R. Zabih, Bayesian multi-camera surveillance, in: Proc. of IEEE Int. Conf. on Computer Vision and Pattern Recognition, volume 2, Fort Collins, CO, USA, pp. 252–259.

[5] O. Javed, Z. Rasheed, O. Alatas, M. Shah, Knight^M: A real time surveillance system for multiple overlapping and non-overlapping cameras, in: IEEE Conf. on Multimedia and Expo, Baltimore, MD, USA, pp. 6–9.

[6] O. Javed, Z. Rasheed, K. Shafique, M. Shah, Tracking across multiple cameras with disjoint views, in: Proc. of IEEE Int. Conf. on Computer Vision, Nice, France, pp. 952–957.

[7] R. Bowden, P. KaewTraKulPong, Towards automated wide area visual surveillance: tracking objects between spatially-separated, uncalibrated views, IEE Proc. on Vision, Image and Signal Processing 152 (2005) 213–223.

[8] iLIDS, Home Office multiple camera tracking scenario definition (UK)., 2008.

[9] A. Turner, A. Penn, Encoding natural movement as an agent-based system: an investigation into human pedestrian behaviour in the built environment, Environment and Planning B: Planning and Design 29 (2002) 473–490.

[10] A. Johansson, D. Helbing, P. K. Shukla, Specification of a microscopic pedestrian model by evolutionary adjustment to video tracking data, Advances in Complex Systems 10 (2007) 271–288.

[11] B. Zhan, D. N. Monekosso, P. Remagnino, S. A. Velastin, L.-Q. Xu, Crowd analysis: a survey, Machine Vision and Applications 19 (2008) 345–357.

[12] R. L. Hughes, The flow of human crowds, Annual Review of Fluid Mechanics 35 (2003) 169–182.

[13] D. Bauer, S. Seer, N. Brndle, Macroscopic pedestrian flow simulation for designing crowd control measures in public transport after special events, in: Summer Computer Simulation Conference, San Diego, CA, USA, pp. 1035–1042.

[14] M. Rodriguez, S. Ali, T. Kanade, Tracking in unstructured crowded scenes, in: Proc. of IEEE Int. Conf. on Computer Vision, Kyoto, Japan, pp. 1389–1396.

[15] S. Ali, M. Shah, Floor fields for tracking in high density crowd scenes, in: Proc. of Europ. Conf. on Computer Vision, Marseille, France, pp. 1–14.

[16] A. Lerner, Y. Chrysanthou, D. Lischinski, Crowds by example, Computer Graphics Forum (Proc. of Eurographics) 26 (2007) 655–664.

[17] G. Antonini, S. V. Martinez, M. Bierlaire, J. P. Thiran, Behavioral priors for detection and tracking of pedestrians in video sequences, Int. Journal of Computer Vision 96 (2006) 159–180.

[18] D. Helbing, P. Molnar, Social force model for pedestrian dynamics, Physical Review E 51 (1995) 4282–4286.

[19] D. Helbing, I. Farkas, T. Vicsek, Simulating dynamical features of escape panic, Nature 407 (2000) 487–490.

[20] E. L. Andrade, R. B. Fisher, Simulation of crowd problems for computer vision, in: First Int. Workshop on Crowd Simulation, Lausanne, Switzerland, pp. 71–80.

[21] R. Mehran, A. Oyama, M. Shah, Abnormal crowd behavior detection using social force model, in: Proc. of IEEE Int. Conf. on Computer Vision and Pattern Recognition, Miami Beach, FL, USA, pp. 935–942.

[22] P. Scovanner, M. F. Tappen, Learning pedestrian dynamics from the real

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42
1	2.3	-	0.1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
2	-	7.2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	49.3	-	-	-	-	-	-	-	-	-	-	11.9	9.5	49.0	48.4	-	-	-	-	-	-	-	-
3	23.6	-	2.0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
4	17.6	-	17.9	0.4	0.3	0.2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
5	-	-	65.8	20.4	19.3	29.8	18.6	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
6	14.5	-	14.6	1.2	0.8	0.5	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
7	-	-	-	-	-	-	-	30.5	67.3	5.9	2.3	0.6	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
8	-	-	-	36.1	38.9	34.6	31.0	0.2	1.8	41.7	53.2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-										
9	-	-	-	22.8	21.5	-	14.9	10.3	0.1	6.9	1.4	45.9	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-										
10	-	-	-	-	-	-	80.0	4.2	27.2	3.7	5.4	8.4	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-				-	-	-	-	-	-	-
11	-	-	-	-	-	-	79.5	8.5	31.1	8.8	2.7	10.5	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
12	-	-	-	-	-	-	36.1	48.7	39.8	22.6	1.0	2.3	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
13	-	-	-	-	-	-	-	-	-	74.5	28.9	-	2.1	37.4	113.1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
14	-	-	-	-	-	-	-	-	-	-	-	-	7.1	1.8	0.1	34.4	46.1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
15	-	-	-	-	-	-	-	-	-	-	92.4	-	28.3	1.9	0.3	43.8	54.8	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
16	-	-	-	-	-	-	-	-	-	-	-	-	57.5	26.7	9.0	3.1	14.6	50.7	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
17	-	-	-	-	-	-	-	-	-	-	-	-	46.1	11.5	10.2	2.4	18.2	27.5	90.6	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
18	-	-	-	-	-	-	-	-	-	-	-	-	46.9	48.3	7.6	26.2	21.7	46.8	5.8	-	65.6	66.6	43.1	76.3	78.8	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
19	-	-	-	-	-	-	-	-	-	-	-	-	95.3	23.3	23.1	22.8	22.6	20.8	0.1	-	1.2	0.4	0.5	0.5	1.3	4.4	71.9	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
20	-	4.0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	2.8	-	-	-	-	-	-	-	-	1.3	-	3.0	8.5	4.1	1.3	95.1	-	-	-	-	-	-	-
21	-	-	-	-	-	-	-	-	-	-	-	-	42.3	44.9	42.7	29.2	19.8	3.1	15.1	-	0.2	0.9	24.1	0.1	19.7	0.2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
22	-	-	-	-	-	-	-	-	-	-	-	-	-	20.9	20.3	21.1	20.4	4.1	12.7	-	1.5	0.1	0.7	0.5	22.0	2.4	0.5	20.1	-	73.1	-	-	-	-	-	-	-	-	-	-	-	-
23	-	-	-	-	-	-	-	-	-	-	-	-	-	11.5	12.7	11.4	12.0	1.8	10.8	-	4.0	1.4	10.3	5.4	8.4	1.3	14.1	63.0	-	-	-	-	-	-	-	-	-	-	-	-	-	-
24	-	-	-	-	-	-	-	-	-	-	-	-	-	2.3	3.4	2.6	1.9	0.9	14.8	-	1.0	0.0	11.3	3.6	11.6	3.3	0.3	5.3	-	61.0	-	-	-	-	-	-	-	-	-	-	-	-
25	-	-	-	-	-	-	-	-	-	-	-	-	-	21.4	31.0	27.0	20.5	19.9	21.6	-	4.2	0.1	22.7	3.0	21.6	0.5	0.3	40.7	-	-	-	-	-	-	-	-	-	-	-	-	-	-
26	-	-	-	-	-	-	-	-	-	-	-	-	-	1.3	26.7	23.9	12.8	11.3	3.2	-	1.7	0.1	0.8	1.9	2.8	0.1	2.8	42.0	-	-	-	-	-	-	-	-	-	-	-	-	-	-
27	-	-	-	-	-	-	-	-	-	-	-	-	-	12.5	12.6	31.9	21.0	18.4	11.7	-	0.4	10.8	12.5	0.3	23.8	12.6	0.6	0.4	-	14.8	-	-	-	-	-	-	-	-	-	-	-	-
28	-	-	-	-	-	-	-	-	-	-	-	-	-	4.0	6.7	5.3	3.8	2.5	2.4	-	4.3	1.1	5.3	2.8	1.8	2.3	0.8	0.6	49.5	0.1	-	-	-	-	-	-	-	-	-	-	-	-
29	-	-	-	-	-	-	-	-	-	-	-	-	-	68.3	14.2	10.2	10.7	11.2	11.2	-	12.5	15.5	13.1	12.6	12.5	12.0	0.3	2.3	0.1	0.4	-	-	-	-	-	-	-	-	-	-	-	-
30	-	-	-	-	-	-	-	-	-	-	-	-	-	-	11.3	12.1	13.4	61.4	8.2	-	26.8	9.4	6.4	20.0	6.4	10.0	6.8	0.1	0.1	0.4	84.0	-	-	-	-	-	-	-	-	-	-	-
31	-	77.0	-	-	-	-	-	-	-	-	-	-	-	-	91.3	102.8	-	-	-	-	23.9	16.7	-	13.6	13.4	-	10.9	16.8	12.6	21.0	0.0	1.2	-	-	-	-	-	-	-	-	-	-
32	-	0.6	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	8.0	4.2	-	4.3	69.2	-	5.4	9.8	7.9	6.6	2.5	0.3	-	-	-	-	-	-	-	-	-	-
33	-	30.9	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0.5	-	-	-	-	-	-	-	-	63.9	-	25.9	26.2	0.1	0.7	50.4	-	-	-	-	-	-	-
34	-	12.5	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0.9	-	-	-	-	-	-	-	-	92.0	-	9.1	9.3	0.3	0.5	23.4	-	-	-	-	-	-	-
35	-	41.1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	2.9	-	-	-	-	-	-	-	-	-	-	27.7	26.5	0.2	3.2	9.3	-	-	-	-	-	-	-
36	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	96.2	-	-	4.4	-	-	-	-	-	-
37	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	4.6	0.6	-	-	-	82.9
38	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	6.6	1.5	-	-	-	87.5
39	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	6.2	65.5	2.1	22.7
40	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	25.4	5.1	30.5	8.6
41	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	8.2	5.2	10.2	11.3
42	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1.8	4.0	15.0	1.0

1 rank
2 rank
3 rank

Correct re-identification
Missed re-identification

Figure 9: Confusion matrix for person re-identification. Each row corresponds to a person P_i to be re-identified. Each column corresponds to possible candidates P_r for re-identification. Each cell contains the minimum distance between the closest predicted trajectory and the trajectory in the observed region (calculated with Eq. 6). Red cell: missed re-identification ranking. Colored cells: different person re-identification ranking. Red-bordered cells: diagonal of the original confusion matrix (in the ideal case it contains the minimum distance). Cells with ’-’: the predicted trajectories are too far away in time to be considered and therefore removed from the candidate list.

world, in: Proc. of IEEE Int. Conf. on Computer Vision, Kyoto, Japan, pp. 381–388.

[23] S. Pellegrini, A. Ess, K. Schindler, L. V. Gool, You'll never walk alone: Modeling social behavior for multi-target tracking, in: Proc. of IEEE Int. Conf. on Computer Vision, Kyoto, Japan, pp. 261–268.

[24] S. Pellegrini, A. Ess, M. Tanaskovic, L. V. Gool, Wrong turn - no dead end: A stochastic pedestrian motion model, in: Computer Vision and Pattern Recognition Workshops, San Francisco, CA, USA, pp. 15–22.

[25] D. Vasquez, T. Fraichard, C. Laugier, Incremental learning of statistical motion patterns with growing hidden markov models, IEEE Trans. on Intelligent Transportation System 10 (2009) 403–416.

[26] R. Hartley, A. Zisserman, Multiple view geometry in computer vision, Second ed. Cambridge University Press (UK), 2004.

[27] G. A. Frank, C. O. Dorso, Room evacuation in the presence of an obstacle, Physica A: Statistical Mechanics and its Applications 390 (2011) 2135–3145.