Int J Comput Vis
DOI 10.1007/s11263-014-0735-3

Learning Collective Crowd Behaviors with Dynamic Pedestrian-Agents

Bolei Zhou · Xiaoou Tang · Xiaogang Wang

Received: 9 September 2013 / Accepted: 24 May 2014

Abstract Collective behaviors characterize the intrinsic dynamics of the crowds. Automatically understanding collective crowd behaviors has important applications to video surveillance, traffic management and crowd control, while it is closely related to scientific fields such as statistical physics and biology. In this paper, a new mixture model of dynamic pedestrian-Agents (MDA) is proposed to learn the collective behavior patterns of pedestrians in crowded scenes from video sequences. From agent-based modeling, each pedestrian in the crowd is driven by a dynamic pedestrian-agent, which is a linear dynamic system with initial and termination states reflecting the pedestrian's belief of the starting point and the destination. The whole crowd is then modeled as a mixture of dynamic pedestrian-agents. Once the model parameters are learned from the trajectories extracted from videos, MDA can simulate the crowd behaviors. It can also infer the past behaviors and predict the future behaviors of pedestrians given their partially observed trajectories, and classify them different pedestrian behaviors. The effectiveness of MDA and its applications are demonstrated by qualitative and quantitative experiments on various video surveillance sequences.

Keywords Crowd behavior analysis · Video surveillance · Motion analysis

Communicated by M. Hebert.

B. Zhou (✉)
Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, Cambridge, MA, USA
e-mail: bolei@mit.edu

X. Tang
Department of Information Engineering, The Chinese University of Hong Kong, Hong Kong, China
e-mail: xtang@ie.cuhk.edu.hk

X. Wang
Department of Electronic Engineering, The Chinese University of Hong Kong, Hong Kong, China
e-mail: xgwang@ee.cuhk.edu.hk

1 Introduction

Automatically understanding the behaviors of pedestrians in crowd from video sequences is of great interest to the computer vision community, and has drawn more and more attentions in recent years (Zhou et al. 2010). It has important applications to event recognition (Hospedales et al. 2011), traffic flow estimation (Wang et al. 2008b), behavior prediction (Antonini et al. 2006), abnormality detection (Mehran et al. 2009), and crowd simulation (Treuille et al. 2006). For example, in video surveillance, many places of security interest, such as shopping malls, train stations, and street intersections, are very crowded. Automatically detecting dangerous and abnormal behaviors in such environments plays an important role to ensure public safety. However, conventional video surveillance systems do not work well in crowded environments. In crowd control and traffic management, recognizing traffic patterns and estimating traffic flows provide valuable information to avoid congestion and to prevent potential crowd disasters (Moussaid et al. 2011). In civil engineering, long-term statistical information from crowd behavior analysis provides guidelines for planning and designing crowded public areas to increase safety and to optimize traffic capacity. One of the underlying challenges of these problems is to model and analyze the collective dynamics of pedestrians in crowd. The collective behaviors of crowds show striking analogies with some self-organization phenomena observed in social science and nat-

Published online: 14 June 2014

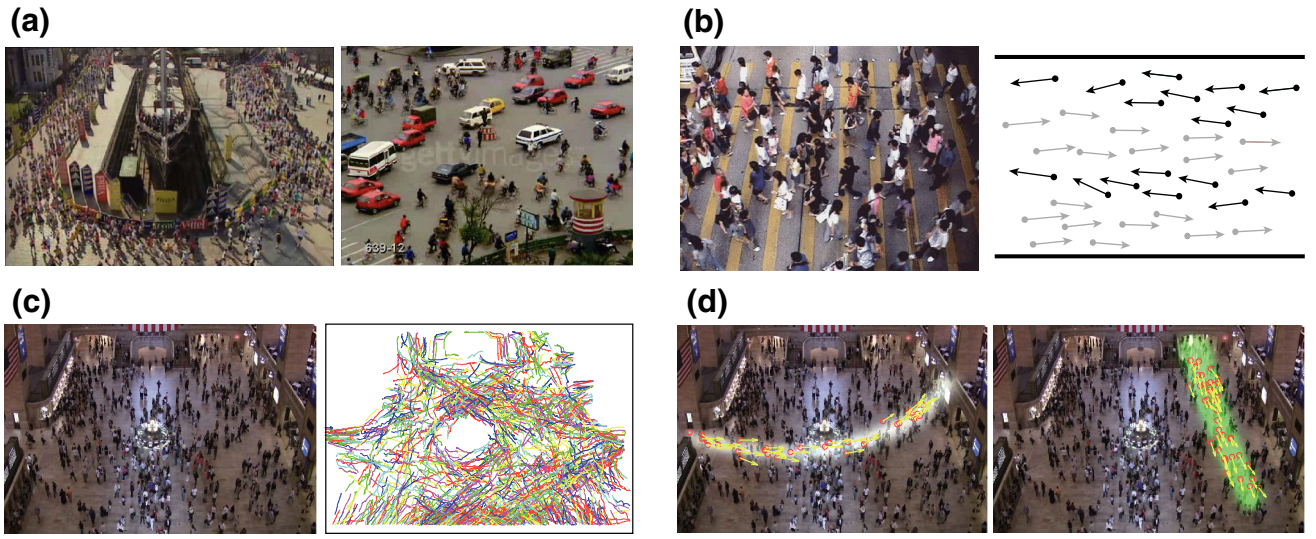

Fig. 1 (**a**) Marathon race at the street corner and traffic flow at the street intersection. (**b**) Bidirectional traffic of pedestrians crossing the street. Collective behaviors of forming lanes spontaneously emerge among the pedestrians. *Black* and *gray arrows* represent pedestrians walking in two opposite directions. (**c**) Crowd of pedestrians walking in a train station scene and the extracted trajectories of pedestrians. Since there are several entry and exit regions, the collective behaviors of the crowd in this scene become complicated. Pedestrians have different beliefs of the starting points and the destinations. These beliefs and other scene structures influence pedestrian behaviors. The shared beliefs and dynamics of movements also generate several major collective dynamic patterns. *Yellow arrows* indicate the moving directions of some exemplar pedestrians. Trajectories are highly fragmented because of frequent occlusions. (**d**) Two collective dynamic patterns of the crowd learned with MDA from these fragmented trajectories. The *colored densities* indicate the spatial distributions of the collective behaviors. Some pedestrians are simulated from MDA. *Red circles* and *yellow arrows* represent the current positions of simulated pedestrians and their velocities (Color figure online)

ural sciences such as physics and biology. Automatic crowd behavior analysis provides powerful tools for studying these related problems and leads to deep insight in these interdisciplinary fields.

Crowd behaviors have been studied in social science with a long history. French sociologist Le Bon (1841–1931) described collective crowd behaviors in his book The Crowd: A Study of the Popular Mind as, "*the crowd, an agglomeration of people, presents new characteristics very different from those of the individuals composing it, the sentiments and ideas of all the persons in the gathering take one and the same direction, and their conscious personality vanishes.* " It leads to the motivation of this work: crowd has its intrinsic collective dynamics. Although individuals in crowd might not acquaint with each other, their shared movements and destinations make them coordinate collectively and follow the paths commonly taken by others.

Collective crowd behaviors are driven by both external and self organization. In Fig. 1a, the collective behaviors of crowds are regularized by scene structures, such as athletic tracks, lane markers, and cross walks, while they are also controlled by traffic signals. Differently, in Fig. 1b, pedestrians are self-organized into several lanes spontaneously. These collective behaviors emerge without external or centralized control (Moussaid et al. 2009). As the scene structure becomes complicated, there will be a variety of collective crowd behaviors happening at the same time. As shown in Fig. 1c, since the train station has multiple entrances/exits and pedestrians have various destinations to reach, and the crowd forms multiple collective crowd behaviors with different dynamics and moving directions. The goal of this work is to statistically model and learn the collective dynamics of the crowd from its observations. This is a fundamental problem for understanding the collective crowd behaviors. It is quite challenging since detecting and tracking pedestrians fails frequently in crowded environments. Meanwhile, crowd behaviors involve a large number of objects, which increase the complexity of this problem.

In this paper, a new Mixture model of Dynamic pedestrian-Agents (MDA) is proposed to learn the collective dynamics of pedestrians from a large amount of observations without supervision. MDA is an agent-based model (Bonabeau 2002), which treats pedestrians as agents and models their process of deciding next actions based on current states. Therefore, MDA is suitable for simulating crowd behaviors once learned from real videos. Observations are trajectories of feature points on pedestrians obtained by a KLT tracker (Tomasi and Kanade 1991). Because of the frequent occlusions in crowded scenes and the tracking failures, most trajectories are highly fragmented with large portions of missing observations. The movement of a pedestrian is driven by

one of dynamic pedestrian-agents. Each dynamic pedestrian-agent is modeled as a linear dynamic system with initial and termination states reflecting pedestrians' beliefs of the starting point and the destination. The timings of pedestrians entering the scene with different dynamic patterns are modeled as Poisson processes. Thus, each dynamic pedestrian-agent represents one type of collective crowd behaviors. The collective dynamics of the whole crowd is further modeled as a mixture of dynamic pedestrian-agents. The effectiveness of MDA is demonstrated by multiple applications: simulating collective crowd behaviors, detecting semantic regions, estimating transition probabilities of traffic flows between entrance and exit regions, classifying collective behaviors, detecting abnormal behaviors, and predicting pedestrian behaviors [1].

The novelty and contributions of this work are summarized as follows. (1) Although there exist approaches (Hospedales et al. 2009; Wang et al. 2008b; Lin et al. 2009; Zhou et al. 2011) of learning motion patterns in crowded scenes, they did not explicitly model the dynamics of pedestrians. Many of them only took local location-velocity pairs as model input, while discarding the temporal order of trajectories which is important for both classification and simulation. Instead, MDA takes trajectories as model input, and it further considers the temporal generative process of trajectories. Therefore, it is much more natural for MDA to simulate collective crowd behaviors and to predict pedestrians' future behaviors, after the model parameters are learned from the real data. (2) Under MDA, pedestrians' beliefs, which strongly regularize their behaviors, are explicitly modeled and inferred from observations. In order to be robust to tracking failures, the states of missing observations on trajectories are modeled and inferred. Because of these two facts, MDA can well infer the past behaviors and predict the future behaviors of pedestrians given their partially observed trajectories. It also leads to better accuracy of recognizing the behaviors of pedestrians. (3) MDA is the first agent-based model to learn the global collective dynamics of crowd from videos. Based on the conference version of this work (Zhou et al. 2012b), more technical details on the model derivation, applications and experimental evaluations on more crowded scenes, and limitations of our model are provided in this paper. The effectiveness and limitation of our approach are evaluated on three video sequences from different scenes: Grand Central Train Station Scene (Zhou et al. 2012b), MIT Traffic Scene (Wang et al. 2008b), and Marathon Race Scene (Ali and Shah 2007).

[1] Datasets, demo videos and related materials are available from http://mmlab.ie.cuhk.edu.hk/project/dynamicagent/

2 Related Works

2.1 Crowd Behavior Analysis in Other Fields

Crowd behavior analysis is an interdisciplinary subject. Understanding the collective behaviors of crowd is a fundamental problem in social science. Social psychology studies (Le Bon 1897; Forsyth 2009) show that when an individual stays in crowd, he behaves differently to being alone. That's because other individuals in the crowd as well as the environment have a huge influence on his cognition and action. In biology, the collective behaviors of organisms such as fish school, flocking birds, and swarming ants have long attracted attentions over decades. People study the mechanism underlying the collective organization of individuals (Couzin 2009), the evolutionary origin of animal aggregation (Parrish and Edelstein-Keshet 1999) and the collective information processing in crowds (Moussaid et al. 2009) from both macroscopic and microscopic levels. Some important research topics such as self-organization, emergence, and phase transition in statistical physics have close relations with crowd behavior analysis. They study the physical laws governing the ways in which animals behave and organize themselves (Ball 2004).

In computer graphics, a number of models are proposed for crowd simulation. A compact survey could be found in (Zhou et al. 2010). Some simulation models come from the statistical fluid mechanics. For example, continuum-based pedestrian models (Hughes 2003; Treuille et al. 2006) treat the crowd motion as fluid, while the navigation fields are used to direct and control the virtual crowds (Patil et al. 2011). Another popular category is agent-based models (Bonabeau 2002), such as the social force model (Helbing and Molnar 1995), self-driven particle model (Vicsek et al. 1995), agent navigation model (Van den Berg et al. 2008), reciprocal velocity obstacles (Berg et al. 2008) and Couzin model (Couzin et al. 2002). Those models treat pedestrians as autonomous agents based on a set of defined rules and known scene structures. They require manually inputting parameters. Differently, under MDA the collective dynamics for crowd behavior simulation are automatically learned from the fragmented trajectories extracted from the real videos without knowing scene structures and without manually setting parameters.

2.2 Crowd Behavior Analysis in Computer Vision

In computer vision, a lot of work focuses on learning global motion patterns (Ali and Shah 2008; Lin et al. 2009, 2010; Mehran et al. 2010; Wang et al. 2008b; Li et al. 2008; Hospedales et al. 2009, 2011; Emonet et al. 2011; Loy et al. 2009; Yang et al. 2009; Kuettel et al. 2010; Makris and Ellis 2005; Wang et al. 2008a, 2011; Kim et al. 2011; Zhou

et al. 2011; Zen and Ricci 2011; Saleemi et al. 2010), modeling local spatio-temporal variations (Mahadevan et al. 2010; Kratz and Nishino 2009; Rodriguez et al. 2009; Wu et al. 2010; Saligrama and Chen 2012), analyzing interactions among individuals (Pellegrini et al. 2009; Mehran et al. 2009; Scovanner and Tappen 2009), and detecting group behaviors (Zhou et al. 2012a; Ge et al. 2011; Moussaid et al. 2010; Choi et al. 2011; Yamaguchi et al. 2011; Pellegrini et al. 2010; Lan et al. 2011, 2012). The learned models of crowd behaviors are also used as priors to improve detection and tracking (Rodriguez et al. 2011; Chang et al. 2011; Zhao and Medioni 2011; Yamaguchi et al. 2011). A brief review is given below.

There has been significant amount of work on learning the motion patterns of crowd. Ali and Shah (2007) and Lin et al. (2009,2010) computed flow fields and segmented crowd flows with Lagrangian coherent structures or Lie algebra. Mehran et al. (2010) proposed a streakline representation for crowd flows. With topic models, Wang et al. (2008b) explored the co-occurrence of moving pixels without tracking objects to learn the motion patterns in crowd. Topic models were augmented by adding spatio-temporal dependency among motion patterns (Hospedales et al. 2009, 2011; Emonet et al. 2011) or supervision (Kuettel et al. 2010). These approaches took the local location-velocity pairs as input while ignoring the temporal order of observations in order to be robust to tracking failures. The beliefs of pedestrians were not considered either. Some approaches (Makris and Ellis 2005; Hu et al. 2007; Wang et al. 2008a, 2011; Morris and Trvedi 2011; Kim et al. 2011; Zhou et al. 2011) learn motion patterns through clustering trajectories, and face the challenge of fragmentation of trajectories in crowded scenes.

Different from MDA, none of the above methods used agent-based models, which could model the process of pedestrians making decisions based on the current states. It is also difficult for them to simulate or predict collective crowd behaviors.

Detecting collective motions and abnormal behaviors in crowd is of great interests for surveillance and crowd management. Zhou et al (2012a). proposed a graph-based method to detect coherent motions from tracklets. Collectivenss, defined as the the degree of individuals acting as a union, was used to measure and detect collective motion patterns (Zhou et al. 2014, 2013). Some approaches were proposed to model the local spatio-temporal variations for abnormality detection with dynamic texture (Mahadevan et al. 2010; Chan and Vasconcelos 2008), HMM (Kratz and Nishino 2009), distributions of spatio-temporal oriented energy (Rodriguez et al. 2009), chaotic invariants (Wu et al. 2010), and local motion descriptors (Saligrama and Chen 2012).

To analyze interactions among pedestrians, the social force model, first proposed by Helbing and Molnar (1995,2000) for crowd simulation, was introduced to the computer vision community recently and was applied to multi-target tracking (Scovanner and Tappen 2009; Pellegrini et al. 2009), abnormality detection (Mehran et al. 2009), and interaction analysis (Scovanner and Tappen 2009). It is also an agent-based model and assumes that pedestrians' movements for the next step are influenced by their destinations, the states of neighbors, and the borders of buildings, streets, and obstacles. It is complementary to MDA, since it models local interactive dynamics among pedestrians but requires the scene structures and the beliefs of pedestrians to be known in advance. MDA better models the global collective dynamics, automatically learns the regularization added by scene structures, and infers the beliefs of pedestrians. Both MDA and the social force model are agent-based models so that they have the potential to be well combined. It would be very interesting to integrate both collective dynamics and interactive dynamics which characterize the crowd behaviors from different perspectives into a single model in the future work. Some individuals with closer relationships form social groups in crowd. They have different interactions than individuals outside the groups. Ge et al. (2011) proposed a hierarchical clustering method to detect groups and Chang et al. (2011) proposed a probabilistic strategy to softly assign individuals into groups. Moussaid et al. (2010) modified the social force model to account for the influence of social groups. Lan et al. (2011,2012) analyzed individual behaviors considering the context of social groups with hierarchical models.

3 MDA Model

A crowd is an agglomeration of pedestrian. Collective crowd behaviors would emerge when enough pedestrians' behaviors are observed, because pedestrians in a specific scene share the common dynamics and beliefs, while their behaviors are also regularized by the same scene structures. These shared movement patterns could be abstracted as different dynamic pedestrian-agents with various *dynamics* and *beliefs*. Each dynamic pedestrian agent represents one type of collective crowd behaviors. In a complex scene, there are multiple types of collective crowd behaviors happening simultaneously. Therefore, a mixture model of dynamic pedestrian-agents is needed. In our model, *dynamics* and *beliefs* of pedestrians are modeled as two key modules D and B in the agent system. Since pedestrians of each dynamic pedestrian-agent emerge from an entrance with a certain frequency, we augment a dynamic pedestrian-agent with another module, *timing* of emerging. Thus, the crowd in a scene is formulated as a mixture model of dynamic pedestrian-agents as shown in Fig. 2. In the following subsections, each module will be explained in details.

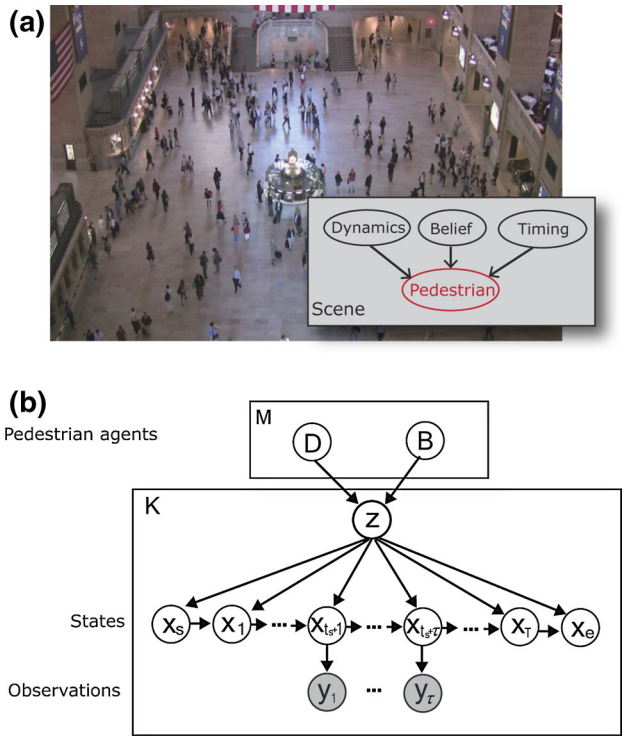

Fig. 2 (**a**) The behavior of a pedestrian in crowd is described with three key components, the dynamics of movements, the belief of the starting point and the destination, and the timing of entering the scene. (**b**) Graphical representation of MDA. The *shadowed* variables are partial observations of the hidden states due to frequent tracking failures in crowded environment

3.1 Modeling Collective Dynamics

Trajectories are time-series observations of pedestrian dynamics. If we treat a pedestrian as a dynamic agent system which actively senses the environment and makes decisions, the trajectory is a set of observations of the hidden dynamic states of this system. The dynamics of a pedestrian-agent is modeled as a linear dynamic system:

$$\mathbf{x}_t = \mathbf{A}\mathbf{x}_{t-1} + v_t, \tag{1}$$

$$\mathbf{y}_t = \mathbf{C}\mathbf{x}_t + w_t. \tag{2}$$

$\mathbf{x}_t = [x_t^1, x_t^2, 1]^\top$ is the current state of the agent and represents its position in homogeneous coordinates. $\mathbf{y}_t \in \mathcal{R}^m$ is the observation of $\mathbf{x}_t$. $\mathbf{A} \in \mathcal{R}^{3\times 3}$ is the state transition matrix and $\mathbf{C} \in \mathcal{R}^{m\times 3}$ is observation matrix. v_t is system noise and w_t is observation noise. Since observations are also positions of agents, $m = 3$ and $\mathbf{C}$ is a identity matrix. The conditional distributions of states and observations are

$$p(\mathbf{x}_t|\mathbf{x}_{t-1}) = \mathcal{N}(\mathbf{x}_t|\mathbf{A}\mathbf{x}_{t-1}, Q), \tag{3}$$

$$p(\mathbf{y}_t|\mathbf{x}_t) = \mathcal{N}(\mathbf{y}_t|\mathbf{x}_t, R), \tag{4}$$

where $\mathcal{N}$ is a 3-dimensional multivariate Gaussian distribution, Q and R are covariance matrices. We denote $D = (\mathbf{A}, Q, R)$ as the *dynamics* parameters to be learned.

Under homogeneous coordinates, $\mathbf{A}$ can be expressed as,

$$\mathbf{A} = \begin{bmatrix} \mathbf{M} & \mathbf{b} \\ \mathbf{0} & 1 \end{bmatrix}, \tag{5}$$

and $\hat{\mathbf{x}}_t = \mathbf{M}\hat{\mathbf{x}}_{t-1} + \mathbf{b}$, where $\hat{\mathbf{x}}_t = [x_t^1, x_t^2]^\top$, $\mathbf{M}$ is a linear transformation matrix, and $\mathbf{b}$ is a translation vector. Therefore, $\mathbf{A}$ is an affine transformation matrix and the dynamics of a dynamic pedestrian-agent is modeled as an affine transform. An important advantage of using homogeneous coordinates is that the multiplication of any two affine transform matrices is also an affine transform matrix. Work (Schneider and Eberly 2003) shows that many important 2D geometric transforms such as translation, geometric contraction, expansion, dilation, rotation, shear, and their combinations are all affine transforms. Thus, $\mathbf{A}$ in Eq. (1) has good generalization capability of learning complex affine transforms from real data.

3.2 Modeling Pedestrian Beliefs

A pedestrian normally has a belief of the starting point and the destination when walking in a scene. This *belief* is a key factor driving the overall behavior of the pedestrian, while it is also considered as the source and sink of the scene (Stauffer 2003; Zhou et al. 2011). We model it as the initial state $\mathbf{x}_s$ and the termination state $\mathbf{x}_e$ of the agent system. For a trajectory k, the joint distribution of the system states and observations is

$$\begin{aligned} &p\left(\mathbf{y}^k, \mathbf{x}^k, \mathbf{x}_e^k, t_s^k, t_e^k\right) \\ &= p\left(t_s^k\right) p\left(t_e^k\right) p\left(\mathbf{x}_s^k\right) p\left(\mathbf{x}_1^k|\mathbf{x}_s^k\right) p\left(\mathbf{x}_e^k|\mathbf{x}_{T^k}^k\right) \\ &\quad \prod_{t=2}^{T^k} p(\mathbf{x}_t^k|\mathbf{x}_{t-1}^k) \prod_{t=1}^{\tau^k} p(\mathbf{y}_t^k|\mathbf{x}_{t_s^k+t}^k). \end{aligned} \tag{6}$$

$\mathbf{x}^k = \left(\{\mathbf{x}_t^k\}_{t=1}^{T^k}, \mathbf{x}_s^k, \mathbf{x}_e^k\right)$ and $\mathbf{y}^k = \{\mathbf{y}_t^k\}_{t=1}^{\tau^k}$. $\mathbf{y}^k$ are the partial observations of the whole set of states $\mathbf{x}^k$. In crowd, the trajectories of objects are highly fragmented due to occlusions. Therefore, most trajectories are only partially observed. We assume that trajectory k is only observed from step $t_s^k + 1$ to $t_s^k + \tau^k$. t_s^k is the number of steps with missing observations between the initial state $\mathbf{x}_s^k$ and $\mathbf{x}_{t_s^k+1}^k$, and t_e^k is the number of steps with missing observations between $\mathbf{x}_{t_s^k+\tau_k}^k$ and the termination state $\mathbf{x}_e^k$ ($T^k = t_e^k + t_s^k + \tau^k$). If $t_s^k = 0$ and $t_e^k = 0$, the complete trajectory is observed. Here we assume the priors of $p(t_s^k)$ and $p(t_e^k)$ are uniform distributions over $[0, H]$, where H is the upper bound of t_s^k and t_e^k to make their

priors proper. Section 4 shows that the choice of H does not affect the learning and inference of MDA as long as it is large enough (e.g. $H = 10,000$). We do not adopt other priors such as truncated Gaussian or exponential distributions, because of lack knowledge on typical distributions of t_s^k and t_e^k.

The initial state is sampled from a Gaussian distribution,

$$p\left(\mathbf{x}_s^k\right) = \mathcal{N}\left(\mathbf{x}_s^k|\mu_s, \Phi_s\right), \tag{7}$$

where μ_s and Φ_s are the mean and covariance matrix of the entry region. The termination state $\mathbf{x}_e$ conditioned on its previous state $\mathbf{x}_{T^k}^k$ is sampled from

$$\begin{aligned} p\left(\mathbf{x}_e^k|\mathbf{x}_{T^k}^k\right) \equiv & p\left(\mathbf{x}_e^k, \mathbf{y}_e^k = \mu_e|\mathbf{x}_{T^k}^k\right) \\ = & p\left(\mathbf{x}_e^k|\mathbf{A}\mathbf{x}_{T^k}^k\right) p\left(\mathbf{y}_e^k = \mu_e|\mathbf{x}_{T^k}^k\right) \\ = & \mathcal{N}\left(\mathbf{x}_e^k|\mathbf{A}\mathbf{x}_{T^k}^k, Q\right)\mathcal{N}\left(\mathbf{x}_e^k|\mu_e, \Phi_e\right) \end{aligned} \tag{8}$$

where μ_e and Φ_e are the mean and covariance matrix of the exit region, and $p\left(\mathbf{y}_e^k|\mathbf{x}_{T^k}^k\right) \equiv \mathcal{N}\left(\mathbf{y}_e^k|\mathbf{x}_e, \Phi_e\right)$. Here to constrain the termination state we introduce a dummy variable $\mathbf{y}_e^k$ and let $\mathbf{y}_e^k = \mu_e$. $p\left(\mathbf{x}_e^k|\mathbf{x}_{T^k}^k\right)$ is then the product of two Gaussian distributions, which is also a Gaussian distribution. Thus the sampling of the termination state is regularized by $\mathbf{x}_{T^k}^k$ and also the center of the exit region. We denote $B = (\mu_s, \Phi_s, \mu_e, \Phi_e)$ as the *belief* parameters. Other conditional distributions such as $p\left(\mathbf{x}_1^k|\mathbf{x}_s^k\right)$, $p\left(\mathbf{x}_t^k|\mathbf{x}_{t-1}^k\right)$, and $p\left(\mathbf{y}_t^k|\mathbf{x}_{t_s^k+t}^k\right)$ are given by Eqs. (3) and (4). $\mathbf{x}^k$, t_s^k and t_e^k are all hidden variables, to be inferred from the model. The initial/termination states and the states of missing observations have to be estimated from the model.

3.3 Mixture of Dynamic Pedestrian-Agents

Numerous pedestrians in a scene have various *dynamics* and *beliefs*. To model the diversity, we extend the single agent system described above to a mixture system with M possible dynamics and beliefs $(D_1, B_1), \ldots, (D_M, B_M)$. A hidden variable $z^k = 1, \ldots, M$ indicates the pedestrian-agent from which a trajectory k is sampled. The prior $p(z^k)$ is a discrete distribution parameterized by $(\pi_1, \ldots, \pi_M)$, *i.e.* $p(z^k = m) = \pi_m$. The joint distribution is

$$\begin{aligned} & p\left(\mathbf{x}^k, \mathbf{y}^k, t_s^k, t_e^k, z^k\right) \\ & = p\left(z^k\right) p\left(t_s^k\right) p\left(t_e^k\right) p\left(\mathbf{x}_s^k|z^k\right) p\left(\mathbf{x}_1^k|\mathbf{x}_s^k, z^k\right) \\ & \quad p\left(\mathbf{x}_e^k|\mathbf{x}_{T^k}^k, z^k\right) \prod_{t=2}^{T^k} p\left(\mathbf{x}_t^k|\mathbf{x}_{t-1}^k, z^k\right) \prod_{t=1}^{\tau^k} p\left(\mathbf{y}_t^k|\mathbf{x}_{t_s^k+t}^k, z^k\right). \end{aligned} \tag{9}$$

3.4 Discussion

Linear dynamic systems (LDS) (Doretto and Chiuso 2003; Oh et al. 2005) and mixture of LDS (Chan and Vasconcelos 2008) have been successfully used to solve computer vision problems in literature. We are inspired by these works and apply mixture of LDS as an agent-based model for pedestrian behavior analysis. Besides the major difference on the targeting problems, our model is different than existing LDS in several other aspects. (1) $\{\mathbf{y}^k\}$ are only partially observed in MDA, but fully observed in other LDS. (2) The temporal length is known in other LDS, while it is a hidden variable in MDA. (3) We model entry and exit regions which add regularization on the initial and termination states.

4 Model Learning and Inference

Given trajectories $\{\mathbf{y}^k\}_{k=1}^K$, we learn the model parameters $\Theta = \{(D_1, B_1, \pi_1), \ldots, (D_M, B_M, \pi_M)\}$ by maximizing the likelihood of observations,

$$\Theta^* = \arg\max_{\Theta} \sum_{k=1}^{K} \log p(\mathbf{y}^k; \Theta). \tag{10}$$

There are three types of hidden variables: (1) the index z^k of assigning a trajectory k to a mixture component; (2) the complete sequence of states $\mathbf{x}^k$ that produce the partial observation $\mathbf{y}^k$; and (3) the numbers of steps with missing observations, *i.e.* t_s^k and t_e^k. We apply the EM algorithm to estimate parameters. Each iteration of EM consists of

$$\begin{aligned} & \textbf{E-step:} \mathcal{Q}(\Theta; \hat{\Theta}) = E_{\mathbf{X},\mathbf{T},\mathbf{Z}|\mathbf{Y};\hat{\Theta}}(\log p(\mathbf{X}, \mathbf{Y}, \mathbf{T}, \mathbf{Z}; \Theta)), \\ & \textbf{M-step:} \hat{\Theta}^* = \arg\max_{\Theta} \mathcal{Q}(\Theta; \hat{\Theta}). \end{aligned}$$

$p(\mathbf{X}, \mathbf{Y}, \mathbf{T}, \mathbf{Z}; \Theta)$ is the complete-data likelihood of the partial observations $\mathbf{Y}$, complete hidden states $\mathbf{X}$ (including the initial and termination states), the numbers of steps with missing observations $\mathbf{T}$, and hidden assignment variables $\mathbf{Z}$.

4.1 Initialization

To initialize the estimation of parameters, we roughly draw the boundaries of entry/exit regions in a scene as shown in Fig. 4a. For every agent component m, its entry/exit region is randomly chosen from these regions (entry and exit regions cannot be the same). For initialization, we let points $\mathbf{y}$ of trajectories which start/end within the source/sink regions of component m be equal to their hidden states $\mathbf{x}$, and then use $\mathbf{x}$ to estimate the dynamics parameters $\mathbf{A}_m$ and Q_m with maximum likelihood estimation. R_m is initialized as [0.1 0 0; 0 0.1 0; 0 0 0]. The starting/ending points

of trajectories which start/end within the entry/exit regions are used to initialize the estimation of belief parameters $(\mu_m^s, \Phi_m^s, \mu_m^e, \Phi_m^e)$. π_m is initialized as $1/M$.

4.2 Expectation Step

The posterior probabilities and the expectation of complete-data likelihood under current estimated parameters $\hat{\Theta}$ are,

$$\begin{aligned}\mathcal{Q} &= E_{\mathbf{X},\mathbf{T},\mathbf{Z}|\mathbf{Y};\hat{\Theta}}(\log p(\mathbf{X},\mathbf{Y},\mathbf{T},\mathbf{Z};\Theta)) \\ &= E_{\mathbf{Z},\mathbf{T}|\mathbf{Y};\hat{\Theta}}(E_{\mathbf{X}|\mathbf{Y},\mathbf{Z},\mathbf{T};\hat{\Theta}}(\log p(\mathbf{X},\mathbf{Y},\mathbf{T},\mathbf{Z};\Theta))) \\ &= \sum_{k,m,g,h} \gamma_k(m,g,h) E_{\mathbf{x}^k|\mathbf{y}^k,z^k=m,t_s^k=g,t_e^k=h} \\ &\quad \left(\log p\left(\mathbf{x}^k,\mathbf{y}^k,t_s^k,t_e^k,z^k\right)\right)\end{aligned}$$

where $\gamma_k(m,g,h)$ is defined as

$$\begin{aligned}\gamma_k(m,g,h) &= p(z^k=m,t_s^k=g,t_e^k=h \mid \mathbf{y}^k) \\ &= \frac{\pi_m p(\mathbf{y}^k|z^k=m,t_s^k=g,t_e^k=h)}{\sum_{m'=1}^{M}\sum_{g',h'}\pi_{m'}p(\mathbf{y}^k|z^k=m',t_s^k=g',t_e^k=h')}.\end{aligned} \tag{11}$$

The priors of $p(t_s^k)$ and $p(t_e^k)$ are uniform. The likelihood of observations $p(\mathbf{y}^k|z^k=m,t_s^k=g,t_e^k=h)$ is computed with the modified Kalman smoothing filter in Sect. 4.4.

$\gamma_k(m,g,h)$ has three discrete variables. It is time consuming to compute all their possible combinations in the range of $[1,M]\times[0,H]^2$. For most (g,h), $\gamma_k(m,g,h)$ are approximately 0. We first estimate the most plausible $\hat{g}$ and $\hat{h}$ for fragmented trajectory k by optimization,

$$\begin{aligned}\hat{h} &= \arg\min_t \parallel \mu_e^m - \mathbf{A}_m^t \mathbf{y}_\tau^k \parallel, \\ \hat{g} &= \arg\min_t \parallel \mu_s^m - \mathbf{A}_m^{-t} \mathbf{y}_1^k \parallel,\end{aligned} \tag{12}$$

where $\mathbf{y}_\tau^k$ and $\mathbf{y}_1^k$ are the last and first points of $\mathbf{y}^k$ and $\mathbf{A}^t$ refer to t matrix power of $\mathbf{A}$. Equation (12) is to find a candidate $\hat{g}$ ($\hat{h}$), the starting point $\hat{\mathbf{y}}_s^k$ (ending point $\hat{\mathbf{y}}_e^k$) predicted according to which is closest to the source center μ_s^m (sink center μ_e^m). Since a starting (ending) point is regularized by the source (sink) with a Gaussian distribution, when g (h) is largely different from $\hat{g}$ ($\hat{h}$), the predicted starting (ending) point is far away from μ_s^m (μ_e^m). Then $\gamma_k(m,g,h)$ is close to zero and can be ignored. The one-dimensional discrete search problems in Eq. (12) can be solved efficiently. The illustration of this optimization is shown in Fig. 3, we extend the trajectory and search the nearest points to the mean of initial state and termination state. Then we limit the plausible set of t_k^s as $[\hat{g}-\Delta,\hat{g}-\Delta+1,\ldots,\hat{g},\ldots,\hat{g}+\Delta-1,\hat{g}+\Delta]$, and

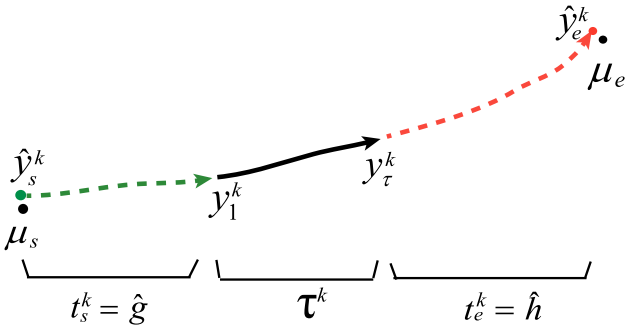

Fig. 3 Estimate the most possible $\hat{g}$ and $\hat{h}$ from Eq. (12) as the numbers of steps generating the nearest points $\hat{\mathbf{y}}_s^k$ and $\hat{\mathbf{y}}_e^k$ to μ_s and μ_e respectively. The *black curve* is a fragmented trajectory, the *dashed curves* are the estimation of missing states

the plausible set of t_k^e as $[\hat{h}-\Delta,\hat{h}-\Delta+1,\ldots,\hat{h},\ldots,\hat{h}+\Delta-1,\hat{h}+\Delta]$, where Δ is an integer and empirically determined. When it is out of the plausible range, $\gamma_k(m,g,h)$ is approximated as 0. So there are $(2\Delta+1)^2$ combinations of (t_s^k,t_e^k).

4.3 Maximization Step

New parameters Θ^* are estimated by maximizing $\mathcal{Q}$. We first recursively estimate the expectations of hidden states and their products, *i.e.*

$$\hat{\mathbf{x}}^k = E_{\mathbf{x}^k|\mathbf{y}^k,z^k=m,t_s^k=g,t_e^k=h}(\mathbf{x}^k), \tag{13}$$

$$P_{t,t}^k = E_{\mathbf{x}^k|\mathbf{y}^k,z^k=m,t_s^k=g,t_e^k=h}(\mathbf{x}_t^k \mathbf{x}_t^{k\top}), \tag{14}$$

$$P_{t,t-1}^k = E_{\mathbf{x}^k|\mathbf{y}^k,z^k=m,t_s^k=g,t_e^k=h}(\mathbf{x}_t^k \mathbf{x}_{t-1}^{k\top}), \tag{15}$$

from partial observations with the modified Kalman smoothing filter (Palma 2007; Shumway and Stoffer 1982), whose details are summarized in Sect. 4.4. The values of $\hat{\mathbf{x}}^k$, $P_{t,t}^k$ and $P_{t,t-1}^k$ depend on the choice of m, g, and h. We do not include the indices of m, g and h in notations to simplify the equations here and in Sect. 4.4. Then Θ^* are updated are as follows,

$$\begin{aligned}\mathbf{A}_m^* = &\left(\sum_{k,g,h}\gamma_k(m,g,h)\sum_{t=1}^{T^k+1}P_{t,t-1}^k\right) \\ &\left(\sum_{k,g,h}\gamma_k(m,g,h)\sum_{t=1}^{T^k+1}P_{t-1,t-1}^k\right)^{-1},\end{aligned} \tag{16}$$

$$\mathcal{Q}_m^* = \frac{\sum_{k,g,h}\gamma_k(m,g,h)\left(\sum_{t=1}^{T^k+1}P_{t,t}^k - \mathbf{A}_m^*\sum_{t=1}^{T^k+1}P_{t,t-1}^k\right)}{\sum_{k,g,h}\gamma_k(m,g,h)(T^k+1)}, \tag{17}$$

$$R_m^* = \frac{\sum_{k,g,h}\gamma_k(m,g,h)\sum_{t=1}^{\tau^k}(\mathbf{y}_t^k\mathbf{y}_t^{k\top} - \hat{\mathbf{x}}_t^k\mathbf{y}_t^{k\top} - \mathbf{y}^k\hat{\mathbf{x}}_t^{k\top} + P_{t,t}^k)}{\sum_{k,g,h}\gamma_k(m,g,h)\tau^k}, \tag{18}$$

$$\mu_s^{m*} = \frac{\sum_{k,g,h}\gamma_k(m,g,h)\hat{\mathbf{x}}_s^k}{\sum_{k,g,h}\gamma_k(m,g,h)}, \tag{19}$$

$$\Phi_s^{m*} = \frac{\sum_{k,g,h} \gamma_k(m,g,h)(\hat{\mathbf{x}}_s^k - \mu_s^m)(\hat{\mathbf{x}}_s^k - \mu_s^m)^\top}{\sum_{k,g,h} \gamma_k(m,g,h)}, \tag{20}$$

$$\mu_e^{m*} = \frac{\sum_{k,g,h} \gamma_k(m,g,h)\hat{\mathbf{x}}_e^k}{\sum_{k,g,h} \gamma_k(m,g,h)}, \tag{21}$$

$$\Phi_e^{m*} = \frac{\sum_{k,g,h} \gamma_k(m,g,h)(\hat{\mathbf{x}}_e^k - \mu_e^m)(\hat{\mathbf{x}}_e^k - \mu_e^m)^\top}{\sum_{k,g,h} \gamma_k(m,g,h)}, \tag{22}$$

$$\pi_m^* = \frac{\sum_{k,g,h} \gamma_k(m,g,h)}{\sum_{m'=1}^{M} \sum_{k,g,h} \gamma_k(m',g,h)}. \tag{23}$$

4.4 Modified Kalman Smoothing Filter

Kalman smoothing filter (Shumway and Stoffer 1982; Palma 2007) is used to estimate the means and covariances (in Eqs. (13)–(15)) of the states $\mathbf{x}$ of a LDS conditioned on the observations $\{\mathbf{y}_t\}_{t=1}^T$ ($T = t_s + t_e + \tau$). It is assumed that $\mathbf{y}_t$ is observed at steps $t_s + 1$ to $t_s + \tau$ and missed at steps 1 to t_s and $t_s + \tau + 1$ to T. Kalman filter is also used to compute the likelihood of observations in Eq. (11). Detailed discussion and proof on the modifications on the Kalman filter in order to account for the missing observations can be found in (Palma 2007).

Denote the expectations conditioned on the observed sequence $\mathbf{y}_1, \ldots, \mathbf{y}_n$ as

$$\mathbf{x}_t^n = E_{\mathbf{x}|\mathbf{y}_1,\ldots,\mathbf{y}_n}(\mathbf{x}_t), \tag{24}$$

$$V_t^n = E_{\mathbf{x}|\mathbf{y}_1,\ldots,\mathbf{y}_n}((\mathbf{x}_t - \mathbf{x}_t^n)(\mathbf{x}_t - \mathbf{x}_t^n)^\top), \tag{25}$$

$$V_{t,t-1}^n = E_{\mathbf{x}|\mathbf{y}_1,\ldots,\mathbf{y}_n}\left((\mathbf{x}_t - \mathbf{x}_t^n)(\mathbf{x}_{t-1} - \mathbf{x}_{t-1}^n)^\top\right). \tag{26}$$

For $t = 1, \ldots, T$, we obtain the Kalman forward recursions:

$$\mathbf{x}_t^{t-1} = \mathbf{A}\mathbf{x}_{t-1}^{t-1},$$
$$V_t^{t-1} = \mathbf{A}V_{t-1}^{t-1}\mathbf{A}^\top + Q,$$
$$K_t = V_t^{t-1}(V_t^{t-1} + R)^{-1},$$
$$\mathbf{x}_t^t = \begin{cases} \mathbf{x}_t^{t-1} + K_t\left(\mathbf{y}_t - \mathbf{x}_t^{t-1}\right) & \text{if } \mathbf{y}_t \text{ observed,} \\ \mathbf{x}_t^{t-1} & \text{if } \mathbf{y}_t \text{ missed,} \end{cases}$$
$$V_t^t = V_t^{t-1} - K_t V_t^{t-1},$$

where $\mathbf{x}_1^0 = \mu_s$ and $V_1^0 = \Phi_s$. When $t = T + 1$, it reaches termination state, one more Kalman forward recursion with the constraint of exit region distribution is

$$\mathbf{x}_T^{T+1} = \mathbf{A}\mathbf{x}_T^T,$$
$$V_T^{T+1} = \mathbf{A}V_T^T\mathbf{A}^\top + Q,$$
$$K_{T+1} = V_{T+1}^T(V_{T+1}^T + \Phi_e)^{-1},$$
$$\mathbf{x}_{T+1}^{T+1} = \mathbf{x}_T^{T+1} + K_{T+1}(\mu_e - \mathbf{x}_{T+1}^T),$$
$$V_{T+1}^{T+1} = V_T^{T+1} - K_{T+1}V_{T+1}^T.$$

Then $\hat{\mathbf{x}}_e = \mathbf{x}_{T+1}^{T+1}$. To further compute $\hat{\mathbf{x}}_t \equiv \mathbf{x}_t^{T+1}$ and $P_{t,t} \equiv V_t^{T+1} + \mathbf{x}_t^{T+1}\mathbf{x}_t^{T+1\top}$, we perform backward recursions from $t = T + 1, \ldots, 1$ using

$$J_{t-1} = V_{t-1}^{t-1}\mathbf{A}^\top\left(V_t^{t-1}\right)^{-1},$$
$$\mathbf{x}_{t-1}^{T+1} = \mathbf{x}_{t-1}^{t-1} + J_{t-1}\left(\mathbf{x}_t^{T+1} - \mathbf{A}\mathbf{x}_{t-1}^{t-1}\right),$$
$$V_{t-1}^{T+1} = V_{t-1}^{t-1} + J_{t-1}\left(V_t^{T+1} - V_t^{t-1}\right)J_{t-1}^\top.$$

Here $\hat{\mathbf{x}}_s = \mathbf{x}_0^{T+1}$. To compute $P_{t,t-1} \equiv V_{t,t-1}^{T+1} + \mathbf{x}_t^{T+1}\mathbf{x}_{t-1}^{T+1\top}$, one performs the backward recursions from $t = T + 1, \ldots, 2$

$$V_{t-1,t-2}^{T+1} = V_{t-1}^{t-1}J_{t-2}^\top + J_{t-1}\left(V_{t,t-1}^{T+1} - \mathbf{A}V_{t-1}^{t-1}\right)J_{t-2}^\top,$$

with initial condition $V_{T+1,T}^{T+1} = (I - K_{T+1})\mathbf{A}V_T^T$.

To compute the log-likelihood of observation $\mathbf{y}$, we use the innovations form (Shumway and Stoffer 1982),

$$\log p(\mathbf{y}) = \sum_{t=1}^{\tau} \log p(\mathbf{y}_t \mid \mathbf{y}_1^{t-1}) = \sum_{t=1}^{\tau} \log \mathcal{N}(\mathbf{y}_t | \hat{\mathbf{x}}_t^{t-1}, V_t^{t-1} + R). \tag{27}$$

Then $\gamma(m,g,h)$ can be computed from $p(\mathbf{y}|z = m, t_s = g, t_e = h)$ in Eq. (11).

5 Simulation and Prediction

5.1 Crowd Behavior Simulation

To simulate crowd behaviors by sampling trajectories, we also model the frequency of new pedestrians entering the scene over time, and integrate this module into MDA.

We assume the timings of pedestrians emerging in an entrance region follows a homogeneous Poisson process, whose underlying distribution is a Poisson distribution

$$p(N(t + \Delta t) - N(t) = n) = \frac{(\lambda \Delta t)^n e^{-\lambda \Delta t}}{n!}, \tag{28}$$

where n is the number of emerging pedestrians during time interval $(t, t + \Delta t)$. λ is the rate parameter and indicates the expected number of emerging pedestrians per time interval.

After $\{(D_1, B_1), \ldots, (D_M, B_M)\}$ are learned, every trajectory k has the most likely z^k, and its emerging time can also be estimated. Thus we can count the number of emerging pedestrians in each time interval Δt (here Δt is 5 seconds), and estimate λ for each dynamic pedestrian-agent by

maximum likelihood estimation. Since a pedestrian may be lost and found again during the tracking process, it causes the multiple counting problem. Therefore a trajectory k is counted only if its first observation $\mathbf{y}_1^i$ is within the source region, which is an ellipse specified by $(\mu_s^{z^k}, 4\Phi_s^{z^k})$, such that the bias can be reduced to some extent.

Once MDA is learned, crowd behaviors can be simulated as follows: firstly obtaining the temporal order of emerging by sampling from the Poisson process, and then at each time step of emerging generating the trajectory by sequentially sampling from the linear dynamic system with an initial state and a termination state. Since constraint on termination state is not considered during sequentially sampling, there is a resampling step in the end. The procedure of sampling one trajectory from a pedestrian-agent is listed in Algorithm 1.

Algorithm 1 Model sampling

INPUT: time length L, resampling number N, pedestrian-agent m.
OUTPUT: simulated trajectories.
01:sample temporal order δ_{1-H} from $Poisson P(\lambda_m)$
02:**for** $\omega = 1 : L$
03: **if** $\delta_\omega == 1$
04: **for** $n = 1 : N$
05: sample $\mathbf{x}_s^n$ from $p_m(\mathbf{x}_s^n)$
07: $T^n = \arg\min_t \parallel \mu_m^e - \mathbf{A}_m^t \mathbf{x}_s^n \parallel$.
08: generate trajectory $\mathbf{y}^n = \{\mathbf{y}_t^n\}_{t=1}^{T^n}$ by sequentially sampling $p_m(\mathbf{x}_t^n|\mathbf{x}_{t-1}^n)$ and $p_m(\mathbf{y}_t^n|\mathbf{x}_t^n)$.
sample $\mathbf{x}_e^n$ from $p_m(\mathbf{x}_e^n|\mathbf{x}_{T^n}^n)$, then compute $l_n = p_m(\mathbf{x}_e^n)$.
09: **end for**
10: resample one trajectory $\mathbf{y}$ out of the N simulated trajectories $\{\mathbf{y}^n\}$ according to normalized distribution $\{l_1, \ldots, l_N\}$.
11: **end if**
12:**end for**

5.2 Pedestrian Behavior Prediction

After MDA is learned, given a fragmented trajectory of a pedestrian, our model can fit it to the optimal pedestrian-agent z^* and predict the pedestrian's past and future paths with the corresponding state transition matrix $\mathbf{A}_{z^*}$, as well as the the starting point and the destination with the corresponding belief parameters B_{z^*}. The procedure of fitting a pedestrian-agent is listed in Algorithm 2.

Algorithm 2 Model fitting

INPUT: trajectory k from any tracker.
OUTPUT: the optimal fitted z^*.
01: **for** $m = 1 : M$ **do**
02: compute $\gamma(z^k = m) = \sum_{g,h} \gamma_k(m, g, h)$
03: **end for**
04:$z^* = \arg\max_m \gamma(z^k = m)$
05:compute the future states or past states with $\mathbf{A}_{z^*}$; predict the belief with B_{z^*}.

6 Experiments and Applications

Most experimental results are reported on a 15-minutes video collected from the New York Grand Central Station at 24fps with a resolution of 480 × 720. A KLT keypoint tracker (Tomasi and Kanade 1991) is used to extract trajectories. Tracking terminates when ambiguities caused by occlusions and clutters arise, and new tracks are initialized later. After filtering short and stationary trajectories, around 20,000 trajectories are extracted and shown in Fig. 4a. The histogram of trajectory lengths in Fig. 4b shows that most trajectories are highly fragmented and short. More experimental results on the MIT traffic dataset (Wang et al. 2008b) and the marathon race video (Ali and Shah 2007) are reported in Sects. 6.8 and 6.9.

6.1 Model Learning

To initialize the parameters of MDA, we first roughly label 8 entry/exit regions with ellipses indexed by 1–8 in Fig. 4a. Parameters are initialized according to Sect. 4.1. It takes around one hour for EM to converge, running on a computer with 3GHz Core Quad CPU and 4GB RAM with Matlab implementation. Totally $M = 20$ agent components are learned. In this work, M is chosen empirically, but it also could be estimated with Dirichlet process (Wang et al. 2008b). The model learning is not sensitive to initialization.

Figure 5a illustrates ten representative dynamic pedestrian-agents. Trajectories are sampled from each pedestrian-agent using Algorithm 1. Results show that the learned dynamic pedestrian-agents have different dynamics, beliefs and timings of emerging, so that they characterize various types of collective behaviors. The learned distributions of initial/termination sates are more accurate than the initialized entry/exit regions. For example, region 8 in Fig. 4a corresponds to multiple smaller initial/termination state distributions in Fig. 5. The entry and exit regions of a dynamic pedestrian-agent are randomly selected in initialization. However, if there is no commonly taken path between them, the dynamic pedestrian-agent will diminish and switch to other entry/exit regions during EM learning. For example, there is no path connecting regions 8 and 7, regions 2 and 3 among the learned dynamic pedestrian-agents. Some paths are deformed by the information booth at the center of the scene. By densely sampling, MDA can also estimate the velocity flow field for each pedestrian-agent as shown in Fig. 5b. For comparison, the representative flow fields by LAB-FM (Lin et al. 2009), which learned motion patterns using Lie algebra, are shown in Fig. 5c. MDA performs better in terms of capturing long-range collective behaviors and separating different collective behaviors. For example, some flow fields learned with LAB-FM are locally distributed,without

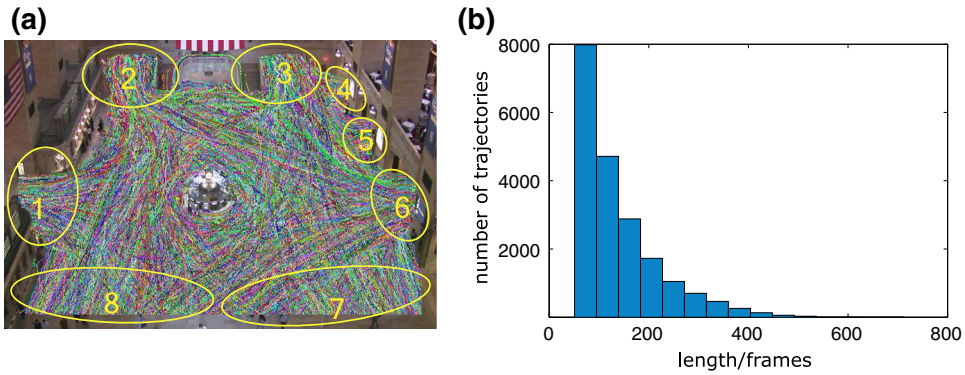

Fig. 4 (**a**) Extracted trajectories and entry/exit regions indicated by *yellow ellipses*. The colors of trajectories are randomly assigned. (**b**) Histogram of the trajectory lengths. Most of the trajectories are short and fragmented (Color figure online)

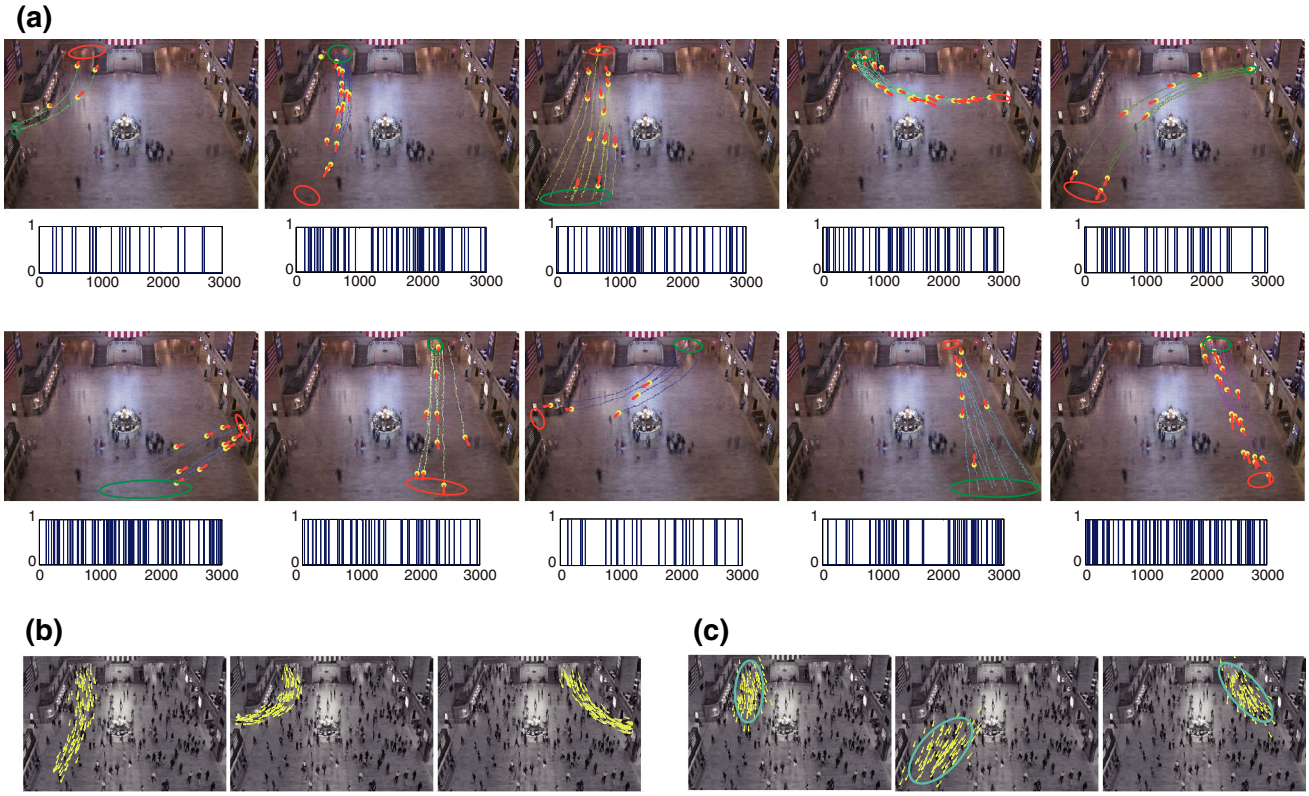

Fig. 5 (**a**) Ten representative dynamic pedestrian-agents with their simulated pedestrians. *Green* and *red circles* indicate the learned distributions of initial/termination states for each pedestrian-agent. *Yellow circles* indicate the current positions of the simulated pedestrians along with their trajectories, and *red arrows* indicate current velocities. The timings of pedestrians entering the scene sampled from the Poisson process are shown below. One impulse indicates a new pedestrian entering the scene, who is driven by the corresponding dynamic pedestrian-agent. (**b**) Flow fields generated from dynamic pedestrian-agents. (**c**) Flow fields learned by LAB-FM (Lin et al. 2009) (Color figure online)

covering the complete paths. The upper parts of the first two flow fields in Fig. 5b, which represent two different collective behaviors, are merged by LAB-FM as shown in the first flow field in Fig. 5c. This is due to the facts that (1) MDA better models the shared beliefs of pedestrians and states of missing observations, and takes the whole trajectories instead of local position-velocity pairs as input, and (2) LAB-FM assumes that the spatial distributions of the flow fields are Gaussian (indicated by cyan ellipses).

6.2 Collective Crowd Behavior Simulation

Compared with other approaches (Hospedales et al. 2009; Wang et al. 2008b; Zhou et al. 2011) of modeling global motion patterns in crowded scenes, a distinctive feature of MDA is to simulate collective crowd behaviors once the model parameters are learned from observations. According to the superposition property of Poisson process (Kingman 1993), the timings of overall pedestrians emerging in

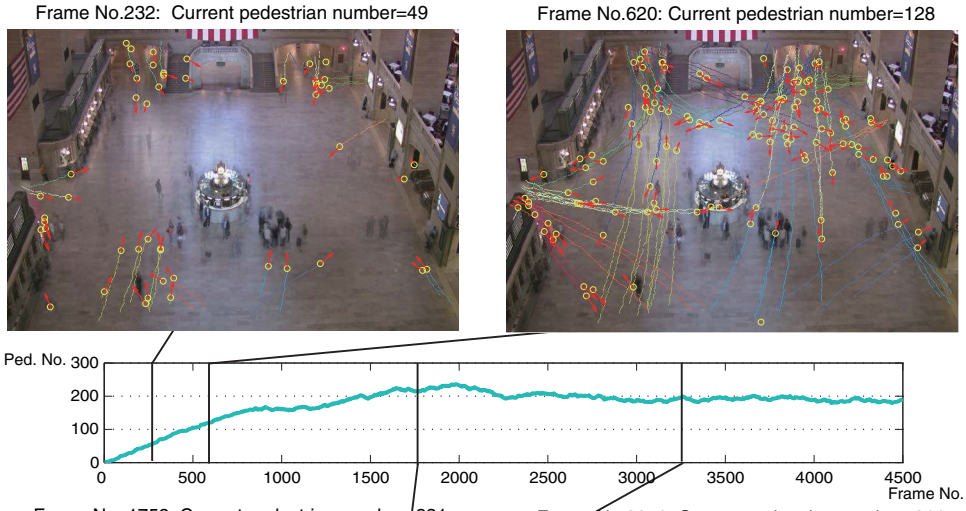

Fig. 6 Four exemplar frames from the crowd behavior simulation. Simulated trajectories are colored according to the indices of their dynamic pedestrian-agents. The *middle plots* the population of pedestrians over time

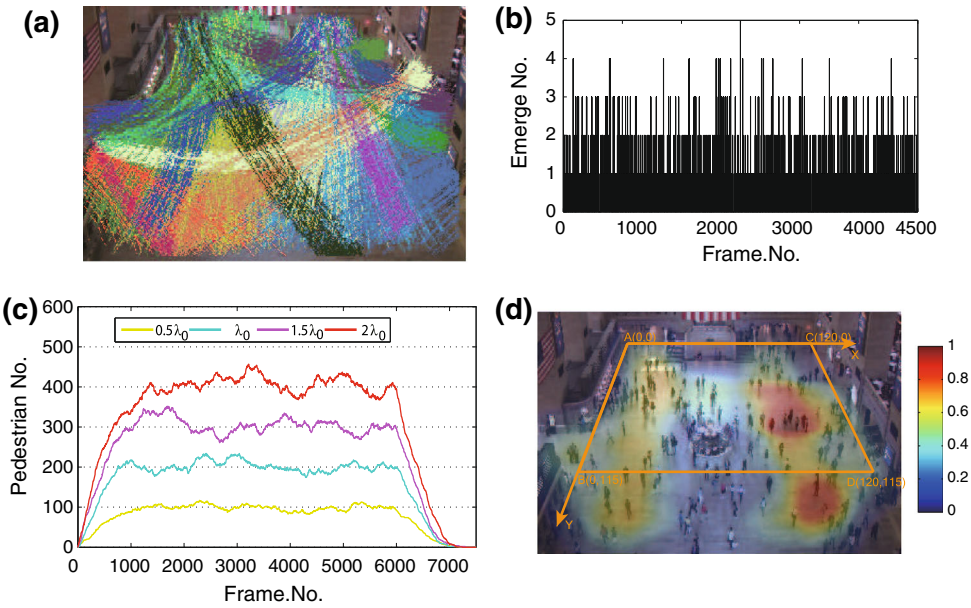

Fig. 7 (**a**) All the simulated trajectories. Colors of trajectories are assigned according to pedestrian-agent indices. (**b**) The numbers of pedestrians entering the scene at different frames. (**c**) The population of the scene with $\lambda = 0.5\lambda_0, \lambda_0, 1.5\lambda_0, 2\lambda_0$ in simulation, where λ_0 is the value learned from data. (**d**) Relative population density map computed from the crowd simulation. It is normalized to the perspective distortion labeled by the *orange* polygon (Color figure online)

the scene also follow a Poisson process with $\lambda = \sum_{m=1}^{M} \lambda_m$. To simulate the overall crowd, we first sample the temporal order series from the Poisson process with λ. Then for each newly emerging pedestrian, its pedestrian-agent index is first sampled from the discrete distribution $(\pi_1, \ldots, \pi_M)$, then its trajectory is sampled from the dynamic pedestrian-agent using Algorithm 1.

Figure 6 shows four exemplar frames of the simulated crowd behaviors. At the first frame pedestrians begin to enter the empty scene. After 1500 frames the crowd reaches the equilibrium population with around 200 pedestrians.

Figure 7a plots all the simulated trajectories over 4500 frames. Figure 7b shows the numbers of new pedestrians entering the scene over time. The crowd simulation with

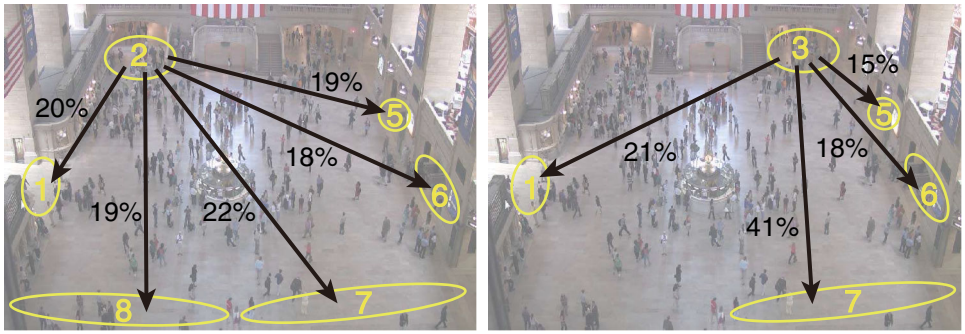

Fig. 8 The transition ratios of pedestrian flows from entries 2 and 3 to other exits

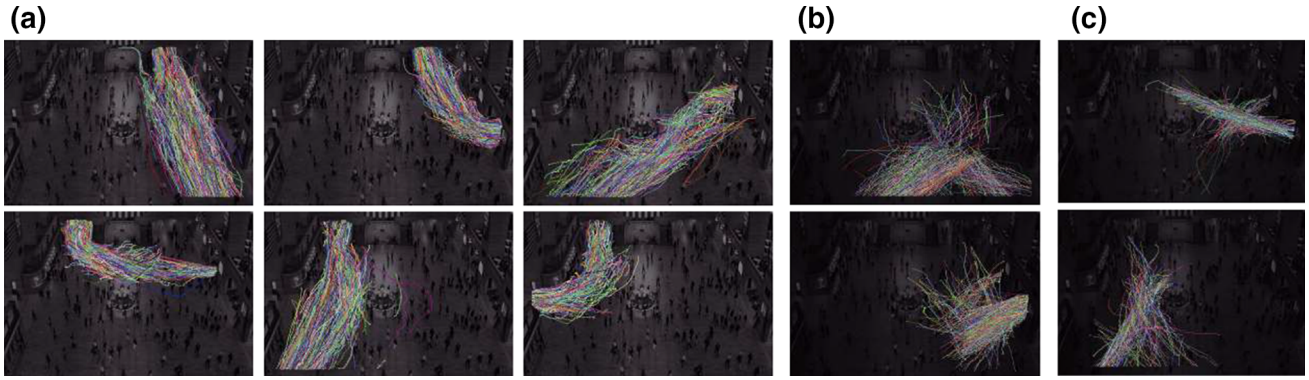

Fig. 9 Representative clusters of trajectories obtained with (**a**) MDA, (**b**) Spectral Clustering (Wang et al. 2006) and (**c**) HDP (Wang et al. 2008a)

MDA can provide valuable information about the dynamics of the crowd. For example, in Fig. 7c, we investigate the relationship between the different rate parameter λ and the population of the scene, where pedestrians begin and stop to enter the scene at the Frame 1 and 6000 respectively. As pedestrians keep entering the scene with a constant birth rate, the scene reaches its equilibrium state. When $\lambda = \lambda_0$, which is learned from data, the system reaches its equilibrium state after 1500 frames with around 200 pedestrians in the scene. And the equilibrium state changes with different birth rates. In Fig. 7d we compute the averaged population density map when $\lambda = \lambda_0$. In this scene, the crossing regions of multiple paths and the entrance/exit regions have higher population density. These crowded areas deserve more attention of security since accidents would most likely happen there when panic or abnormal event strikes. These types of information are very useful for crowd management and public facility optimization.

6.3 Flow Transitions between Sources and Sinks

We can compute the transition ratios of pedestrian flows from the simulation data. Figure 8 shows the transition probabilities from regions 2 and 3 to the other exits. We can observe that the pedestrian flow from source 2 goes roughly equally to the other five exit regions; but differently 41 % pedestrians from region 3 go to region 7. Explaining the difference requires knowledge on the infrastructure and the transportation schedule of the train station. These statistics of the pedestrian flow provide useful information for crowd control and management.

6.4 Collective Behavior Classification

MDA can be used to cluster trajectories of pedestrians into different collective motion patterns. Here we simply take the inferred z^k from Algorithm 2 as the cluster index of that trajectory. A lot of works have been done on trajectory clustering (Hu et al. 2004; Morris and Trivedi 2008). This problem is especially challenging in crowded scenes because trajectories are highly fragmented with many missing observations. Generally speaking, existing approaches are in two categories: distance-based (Wang et al. 2006; Hu et al. 2007) and model-based (Wang et al. 2008a; Morris and Trvedi 2011). We choose one representative approach from each category for comparison: Hausdorff distance-based spectral clustering (Wang et al. 2006) and hierarchical Dirichlet processes (HDP) (Wang et al. 2008a). Figure 9a shows some representative clusters obtained by MDA. Even though most trajectories are fragmented and are far away from each other in space, they are still well grouped into one cluster because they share the same collective dynamics. Figure 9b and c show the representative clusters obtained by spectral clustering (Wang et al. 2006) and HDP (Wang et al. 2008a). They are all in short spatial range and it is hard to interpret their

Table 1 Completeness and correctness of MDA, and Spectral Clustering (Spectral) (Wang et al. 2006), with different numbers of clusters. The result of HDP (Wang et al. 2008a) is also shown. HDP automatically finds the number of clusters from data as 22

Cluster Number		2	5	8	11	14	17	20	25	30
MDA	*Completeness*	0.82	0.71	0.73	0.52	0.51	0.61	0.70	0.62	0.59
	Correctness	0.21	0.75	0.80	0.92	0.92	0.91	0.92	0.95	0.97
Spectral	*Completeness*	0.83	0.60	0.50	0.43	0.42	0.39	0.36	0.27	0.26
	Correctness	0.51	0.80	0.87	0.89	0.90	0.91	0.91	0.92	0.95
HDP	*Completeness*	0.45(cluster number is 22)								
	Correctness	0.82(cluster number is 22)								

semantic meanings, because they cannot well handle fragmentation of trajectories.

We use *correctness* and *completeness* introduced in (Moberts et al. 2005) to measure clustering accuracy. Correctness is the accuracy that two trajectories, which belong to different collective behaviors based on the ground truth, are also grouped into different clusters by the algorithm. Completeness is the accuracy that two trajectories, which belong to the same collective behavior, are also grouped into the same cluster by the algorithm. If all the trajectories are grouped into one cluster, the completeness is 100 % while the correctness is 0 %; if every trajectories is put into a different cluster, the completeness is 0 % while the correctness is 100 %. A good clustering algorithm should have both high correctness and high completeness. Rand index (Rand 1971) is another commonly used measure of similarity between clustering results, and can be viewed as a linear combination of correctness and completeness. We choose to report both correctness and completeness scores, such that readers can have a comprehensive understanding of clustering quality.

To measure correctness (completeness), we manually label 2000 (1500) pairs of trajectories and each pair of trajectories belong to different (the same) collective behavior categories (category) as ground truth. The accuracies of correctness and completeness for MDA, HDP (Wang et al. 2008a) and spectral clustering (Wang et al. 2006) are reported in Table 1. MDA achieves the best performance in terms of both correctness and completeness when the cluster number is chosen as 20, and outperforms HDP and spectral clustering. Note that the correctness is low when the cluster number is 2, since many trajectories of different collective behaviors have to be put into one cluster. The completeness is low when the cluster number is large, since trajectories of the same collective behaviors are divided into different clusters.

6.5 Abnormality Detection

We detect abnormal behaviors by measuring the likelihoods of trajectories with MDA, which are normalized by the lengths of trajectories. Figure 10a displays the top 50 abnormal trajectories with low normalized likelihoods. Two concrete examples of the detected abnormal behaviors are shown in Fig. 10b in zoom-in views. Their starting and ending points are marked with red and blue crosses in Fig. 10a. The detected abnormal trajectories are mainly in two categories. (1) Pedestrians change their destinations in the middle way or loiter, such that their trajectories globally deviate from typical paths. (2) Pedestrians have abnormal speed. Our model has tolerance on the change of moving directions and speed in local regions, since linear dynamic systems allow Gaussian noise, whose covariance matrices are learned from data. Abnormality is detected only when significant global deviation happens. The abnormality detection results are reasonable given the proposed MDA. MDA does not model the interactions among pedestrians, therefore the abnormal behaviors caused by interactions cannot be detected. On the other hand, the approaches (Pellegrini et al. 2009; Saligrama and Chen 2012) of only modeling interactions of pedestrians cannot detect global abnormal behaviors. It is an interesting topic to integrate both types of models. Evaluating the abnormality detection results also depends on applications scenarios.

6.6 Semantic Region Generation

In video surveillance, there are a lot of works on learning semantic regions (Wang et al. 2008a; Makris and Ellis 2005; Wang et al. 2008b; Zhou et al. 2011). Semantic regions correspond to paths commonly taken by objects, thus activities observed in the same semantic region have similar semantic interpretation. Semantic regions could be used to improve object detection, classification and tracking (Kaucic et al. 2005; Wang and Wang 2011).

From the perspective of behavior analysis, semantic regions can be interpreted as *the temporal and spatial accumulation of trajectories generated by objects with shared belief and common movement dynamics in the scene*. MDA well describes the generative process of semantic regions. Figure 11 shows the density distributions of ten semantic regions estimated from 1000 trajectories sampled from the corresponding dynamic pedestrian-agents respectively. The distributions of paths converge and become denser towards

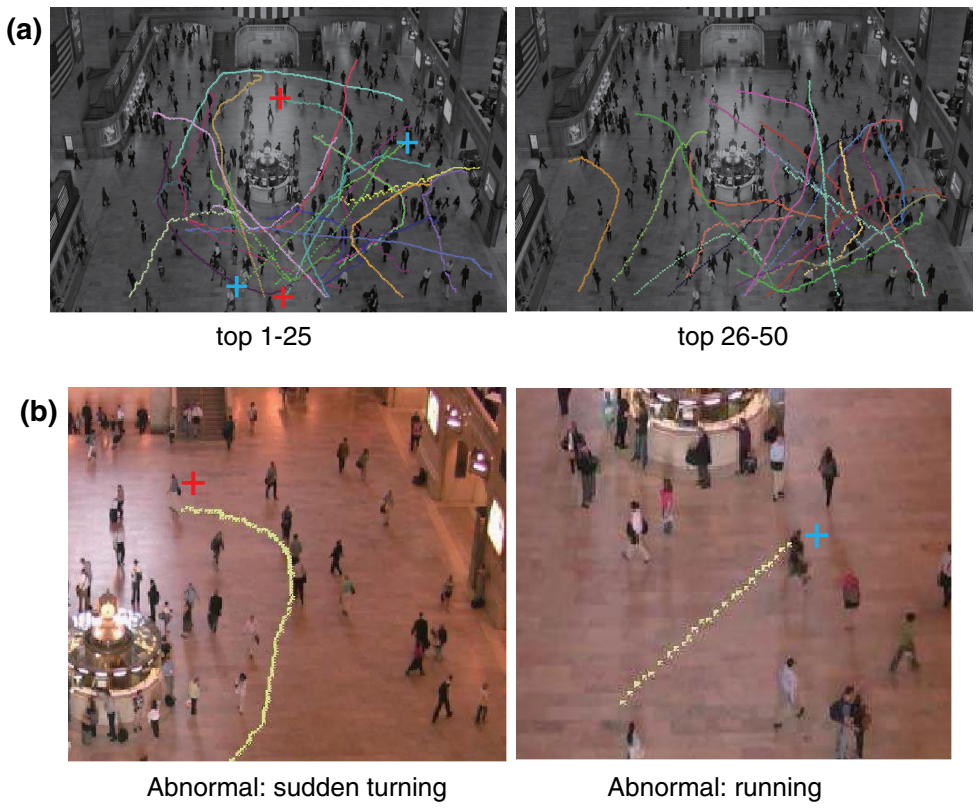

Fig. 10 (**a**) Top 50 abnormal trajectories. (**b**) Examples of abnormal behaviors in zoom-in views. *Left* a pedestrian changes his mind and moves towards a different destination than his original plan. *Right* a pedestrian is running, with dynamics quite different with other walking pedestrians

Fig. 11 Density distributions of ten exemplar semantic regions estimated from trajectories sampled from MDA

the entry/exit regions in the back of the scene. They reflect the perspective distortion of the scene.

6.7 Behavior Prediction

MDA can predict pedestrians' behaviors given that their trajectories are only partially observed. We manually label 30 trajectories of pedestrians as ground-truth. For each ground-truth trajectory, we use the observations of the first 20 frames to estimate its pedestrian-agent index z using the Algorithm 2. Then, the model of the selected pedestrian-agent is used to recursively generate the following states as the predicted future trajectory. The performance is measured by the averaged prediction error, *i.e.* the mean deviation between the predicted trajectories and the ground-truth trajectories. Two baseline methods are used for comparison. In the first comparison method (referred as ConVelocity), a constant velocity which is estimated as the averaged velocity of the past observations, is used to predict future positions. In the second comparison method LAB-FM (Lin et al. 2009), the learned flow field which best fit the first 20 frame observations, is used to predict future positions. Figure 12 show that MDA has better prediction performance.

6.8 Results on the MIT Traffic Dataset

We further test MDA on the MIT traffic Dataset (Wang et al. 2008b). The 25 minutes long video is 25 fps with a resolu-

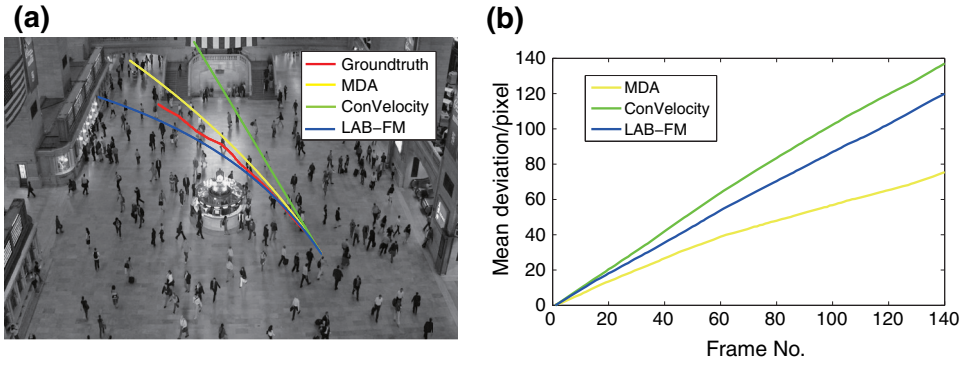

Fig. 12 (**a**) An example of predicting behaviors with different methods. (**b**) The averaged prediction errors with different methods tested on 30 trajectories

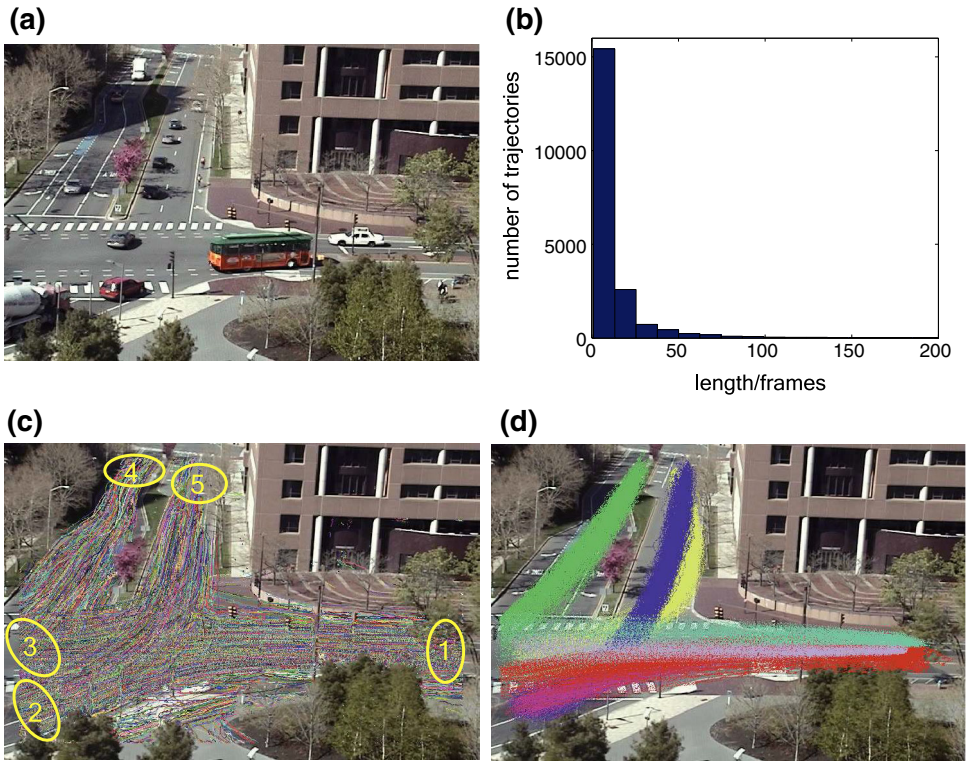

Fig. 13 (**a**) MIT Traffic Dataset. (**b**) Histogram of trajectory lengths. (**c**) Extracted trajectories and initialized entry/exit regions. (**d**) Trajectories simulated from 7 agent models learned from fragmented trajectories

tion of 480 × 720. 43, 389 trajectories are extracted by the KLT tracker. Since majority of moving objects in the scene are vehicles, dynamic agents are vehicles moving on different roads in different directions. As shown in Fig. 13b and c, most trajectories are short and highly fragmented due to occlusion and scene clutters. In Fig. 13c, we label 5 entry/exit regions for initialization. Seven agent models are learned from data. Figure 13d shows trajectories simulated from the 7 agent components. They represent dominant motion patterns of vehicles. Figure 14a shows semantic regions estimated from MDA. Figure 14b, c and d show representative clusters obtained by MDA, spectral clustering (Wang et al. 2006), and HDP (Wang et al. 2008a). The clusters obtained by MDA better reflects the collective motion patterns in the scene.

MDA has some robustness to the initialization for entry/exit regions. In Fig. 15a we label 3 instead of 5 regions: regions 2 and 3 in Fig. 15a are the superset of regions 2, 3 and regions 4, 5 in Fig. 13c respectively. MDA can reasonably cluster trajectories into different dynamics as shown in the first two clusters of Fig. 15a. However, rough labeling of entry/exit regions may also lead to merging similar motion patterns. The last cluster in Fig. 15a shows that trajectories from two dynamic agents are clustered together by MDA.

6.9 Limitation and Extension of MDA

MDA assumes affine transform. Although it can represent many important geometric transforms as discussed in Sect.

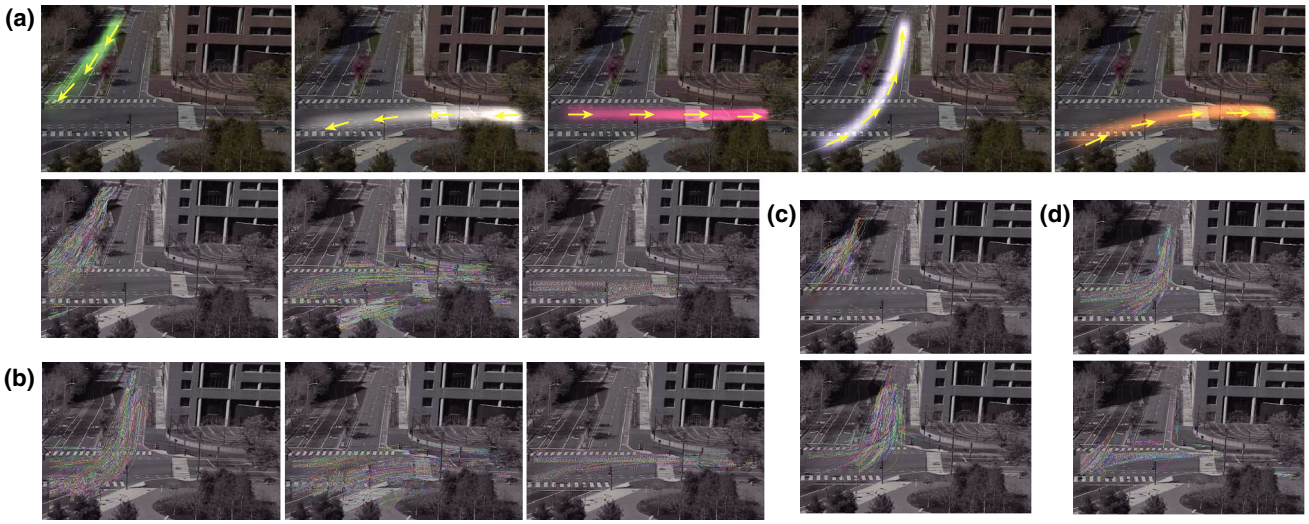

Fig. 14 (**a**) Density distributions of five semantic regions estimated from the learned MDA model. Representative clusters of trajectories obtained with (**b**) MDA model, (**c**) Spectral Clustering (Wang et al. 2006) and (**d**) HDP (Wang et al. 2008a)

3.1, MDA does have difficulty on some complex shapes such as u-turn or s-turn. In Fig. 15b, we learn a single dynamic pedestrian-agent from the trajectories of a marathon race video (Ali and Shah 2007). The trajectories simulated from the learned MDA cannot well fit the real motion pattern indicated by the red trace, since this u-turn shape is not an affine motion. One possible extension is to decompose the complex motion pattern into multiple connected linear dynamic systems with different affine motions. In Fig. 15c, we label 4 entry/exit regions, and 3 dynamic pedestrian-agent components are learned with shared starting and terminating locations. The simulated trajectories from the three connected agent components well fit the real motion of Marathon race, if they can be connected. However, in order to use multiple agents to generate one trajectory, significant modification on MDA has to be made. This extension is related to switching linear dynamic models (Pavlovic et al. 1999), where multiple state transition parameters ($\mathbf{A}$ and Q) are selected via a separate Markovian switching variable as time progresses in a single dynamic system. In our future work we would extend MDA to switching linear systems to model more complex motion patterns.

The number of agent components are empirically decided by the result in Table 1. We can also consider the number of agent components as one of the model parameters so that it can be automatically decided in the model inference. One way is to model the number of mixture as Dirichlet process then the joint distribution becomes the nonparametric bayesian model. The most likely mixture number could be inferred by Markov Chain Monte Carlo or variational inference (Wang et al. 2008b, 2011; Hospedales et al. 2009).

MDA does not model local interactive dynamics among pedestrians, which is also an essential component of describing behaviors in crowd. MDA can be integrated with the social force models (Helbing and Molnar 1995; Pellegrini et al. 2009) to characterize both the collective dynamics and interactive dynamics of crowd behaviors at both macroscopic and microscopic levels, since both are agent-based models. The local interactions from neighbors can be added to the dynamic state transition process in Eq. (1). It could lead to better accuracies on object tracking, behavior classification, simulation, and prediction.

MDA assumes that a pedestrian has a clear belief. Some pedestrians in the New York Grand Central Station simply wait or loiter in the scene without clear destination. Those behaviors cannot be well modeled with MDA.

7 Conclusion

In this paper, we propose a Mixture model of Dynamic Pedestrian-Agent to learn the collective dynamics from video sequences in crowded scenes. The collective dynamics of pedestrians are modeled as linear dynamic systems to capture long range moving patterns. Through modeling the beliefs of pedestrians and the missing states of observations, MDA can be well learned from highly fragmented trajectories caused by frequent tracking failures. Therefore, it is suitable for behavior analysis in crowded environments. By modeling the process of pedestrians making decisions on actions, it can not only classify collective behaviors, but also simulate and predict collective crowd behaviors. Various statistics valu-

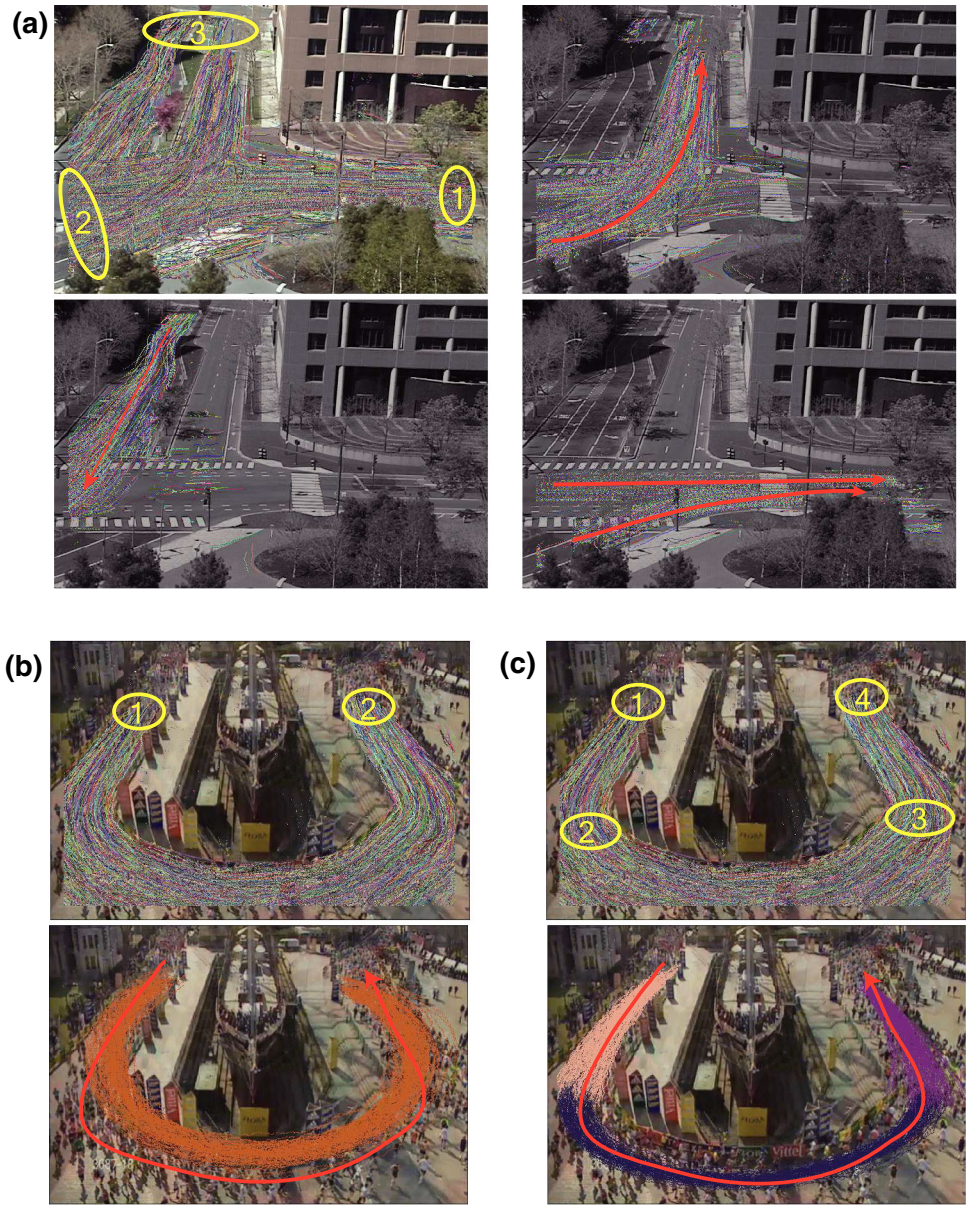

Fig. 15 (**a**) With only 3 initialized entry/exit regions in the MIT traffic scene, MDA still can reasonably cluster trajectories. Three representative clusters are shown. (**b**) A single dynamic pedestrian-agent cannot well fit the real complex motion in the scene due to the limitation of affine transformation. (**c**) 3 connected dynamic pedestrian-agents with different dynamics and shared starting and termination locations can well fit the real motion. Simulated trajectories from the same agent model are in the same *color*, and *red* trace indicates the real motion pattern (Color figure online)

able for traffic management and crowd control, such as flow fields, population density maps,

flow transitions between sources and sinks, and semantic regions can be well estimated from simulation results of the learned MDA.

MDA has more potential applications and extensions to be explored. In this work, we did not study the application of MDA to object tracking in crowd. Since MDA has the capability to predict future behaviors of objects based on partial observations, it can be used as prior for object tracking.

Instead of being fixed, the dynamics parameters D and belief parameters B can also be dynamically updated over time by modeling their temporal dependency. Meanwhile, as the variation of crowd density influences crowd behaviors, it is interesting to investigate how the dynamic and belief parameters change with crowd density.

Acknowledgments This work is partially supported by the General Research Fund sponsored by the Research Grants Council of Hong Kong (Project No. CUHK417110, CUHK417011, and CUHK 429412).

References

Ali, S., Shah, M. (2007). A lagrangian particle dynamics approach for crowd flow segmentation and stability analysis. In *Proceedinds CVPR*.

Ali, S., Shah, M. (2008). Floor fields for tracking in high density crowd scenes. In *Proceedings ECCV*.

Antonini, G., Martinez, S., Bierlaire, M., & Thiran, J. (2006). Behavioral priors for detection and tracking of pedestrians in video sequences. *International Journal of Computer Vision*, *69*, 159–180.

Ball, P. (2004). *Critical mass: How one thing leads to another*. New York: Farrar Straus & Giroux.

Van den Berg, J., Lin, M., Manocha, D. (2008). Reciprocal velocity obstacles for real-time multi-agent navigation. In *Proceedings ICRA*.

Berg, J., Lin, M., Manocha, D. (2008). Reciprocal velocity obstacles for real-time multi-agent navigation. In *Proceedings of IEEE International Conference on Robotics and Automation*.

Bonabeau, E. (2002). Agent-based modeling: Methods and techniques for simulating human systems. *PNAS*, *14*, 7280–7287.

Chan, A. B., & Vasconcelos, N. (2008). Modeling, clustering, and segmenting video with mixtures of dynamic textures. *IEEE Transactions on PAMI*, *30*, 909–926.

Chang, M., Krahnstoever, N., Ge, W. (2011). Probabilistic group-level motion analysis and scenario recognition. In *Proceedings ICCV*.

Choi, W., Shahid, K., Savarese, S. (2011). Learning context for collective activity recognition. In *Proceedings CVPR*.

Couzin, I. (2009). Collective cognition in animal groups. *Trends in Cognitive Sciences*, *13*(1), 36–43.

Couzin, I., Krause, J., James, R., Ruxton, G., & Franks, N. (2002). Collective memory and spatial sorting in animal groups. *Journal of Theoretical Biology*, *218*(1), 1–11.

Doretto, G., & Chiuso, A. (2003). Dynamic textures. *International Journal of Computer Vision*, *51*, 91–109.

Emonet, R., Varadarajan, J., Odobez, J. (2011). Extracting and locating temporal motifs in video scenes using a hierarchical non parametric bayesian model. In *Proceedings CVPR*.

Forsyth, D. (2009). *Group dynamics*. Belmont: Wadsworth Pub Co.

Ge, W., Collins, R., & Ruback, R. (2011). Vision-based analysis of small groups in pedestrian crowds. *IEEE Transactions on PAMI*, *34*(5), 1003–1016.

Helbing, D., & Molnar, P. (1995). Social force model for pedestrian dynamics. *Physical Review E*, *51*(5), 4282–4286.

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

Hospedales, T., Gong, S., Xiang, T. (2009). A markov clustering topic model for mining behaviour in video. In *Proceedings ICCV*.

Hospedales, T., Li, J., Gong, S., & Xiang, T. (2011). Identifying rare and subtle behaviours: A weakly supervised joint topic model. *IEEE Transactions on PAMI*, *33*(12), 2451–2464.

Hu, W., Tan, T., Wang, L., & Maybank, S. (2004). A survey on visual surveillance of object motion and behaviors. *IEEE Transactions on SMC - Part C*, *34*(3), 334–352.

Hu, W., Xie, D., Fu, Z., Zeng, W., & Maybank, S. (2007). Semantic-based surveillance video retrieval. *IEEE Transactions on Image Processing*, *16*(4), 1168–1181.

Hughes, R. (2003). The flow of human crowds. *Annual Review of Fluid Mechanics*, *35*(1), 169–182.

Kaucic, R., Perera, A., Brooksby, G., Kaufhold, J., Hoogs, A. (2005). A unified framework for tracking through occlusions and across sensor gaps. In *Proceedings CVPR*.

Kim, K., Lee, D., Essa, I. (2011). Gaussian process regression flow for analysis of motion trajectories. In *Proceedings ICCV*.

Kingman, J. (1993). *Poisson processes*. Oxford: Oxford University Press.

Kratz, L., Nishino, K. (2009). Anomaly detection in extremely crowded scenes using spatio-temporal motion pattern models. In *Proceedings CVPR*.

Kuettel, D., Breitenstein, M., Van Gool, L., Ferrari, V. (2010). What's going on? discovering spatio-temporal dependencies in dynamic scenes. In *Proceedings CVPR*.

Lan, T., Wang, Y., Yang, W., Robinovitch, S. N., & Mori, G. (2011). Discriminative latent models for recognizing contextual group activities. *IEEE Transactions on PAMI*, *34*(8), 1549–1562.

Lan, T., Sigal, L., Mori, G. (2012). Social roles in hierarchical models for human activity recognition. In *Proceedings CVPR*.

Le Bon, G. (1897). *The crowd: A study of the popular mind*. New York: The Macmillan Co.

Li, J., Gong, S., Xiang, T. (2008). Scene segmentation for behaviour correlation. In *Proceedings ECCV*.

Lin, D., Grimson, E., Fisher, J. (2009). Learning visual flows: A Lie algebraic approach. In *Proceedings CVPR*.

Lin, D., Grimson, E., Fisher, J. (2010). Modeling and estimating persistent motion with geometric flows. In *Proceedings CVPR*.

Loy, C., Xiang, T., Gong, S. (2009). Multi-camera activity correlation analysis. In *Proceedings CVPR*.

Mahadevan, V., Li, W., Bhalodia, V., Vasconcelos, N. (2010). Anomaly detection in crowded scenes. In *Proceedings CVPR*.

Makris, D., & Ellis, T. (2005). Learning semantic scene models from observing activity in visual surveillance. *IEEE Transactions on SMC - Part B*, *35*(3), 397–408.

Mehran, R., Oyama, A., Shah, M. (2009). Abnormal crowd behavior detection using social force model. In *Proceedings CVPR*.

Mehran, R., Moore, B., Shah, M. (2010). A streakline representation of flow in crowded scenes. In *Proceedings ECCV*.

Moberts, B., Vilanova, A., Jake, J.W. (2005). Evaluation of fiber clustering methods for diffusion tensor imaging. In *Proceedings of IEEE Visualization*.

Morris, B. T., & Trivedi, M. M. (2008). A survey of vision-based trajectory learning and analysis for surveillance. *IEEE Transactions on CSVT*, *18*, 1114–1127.

Morris, T. B., & Trvedi, M. M. (2011). Trajectory learning for activity understanding: Unsupervised, multilevel, and long-term adaptive approach. *IEEE Transactions on PAMI*, *33*, 2287–2301.

Moussaid, M., Garnier, S., Theraulaz, G., & Helbing, D. (2009). Collective information processing and pattern formation in swarms, flocks, and crowds. *Topics in Cognitive Science*, *1*(3), 469–497.

Moussaid, M., Perozo, N., Garnier, S., Helbing, D., & Theraulaz, G. (2010). The walking behaviour of pedestrian social groups and its impact on crowd dynamics. *PLoS One*, *5*(4), e10047.

Moussaid, M., Helbing, D., & Theraulaz, G. (2011). How simple rules determine pedestrian behavior and crowd disasters. *PNAS*, *108*, 6884–6888.

Oh, S.M., Rehg, J.M., Balch, T., Dellaert. F. (2005). Learning and inference in parametric switching linear synamic systems. In *Proceedings ICCV*.

Palma, W. (2007). *Long-memory time series: Theory and methods*. Hoboken: Wiley-Blackwell.

Parrish, J., & Edelstein-Keshet, L. (1999). Complexity, pattern, and evolutionary trade-offs in animal aggregation. *Science*, *284*, 99–101.

Patil, S., Van Den Berg, J., Curtis, S., Lin, M. C., & Manocha, D. (2011). Directing crowd simulations using navigation fields. *IEEE Transactions on Visualization and Computer Graphics*, *17*(2), 244–254.

Pavlovic, V., Frey, B., Huang, T. (1999). Time-series classification using mixed-state dynamic bayesian networks. In *Proceedings CVPR*.

Pellegrini, S., Ess, A., Schindler, K., Van Gool, L. (2009). You'll never walk alone: Modeling social behavior for multi-target tracking. In *Proceedings ICCV*.

Pellegrini, S., Ess, A., Van Gool, L. (2010). Improving data association by joint modeling of pedestrian trajectories and groupings. In *Proceedings ECCV*.

Rand, W. M. (1971). Objective criteria for the evaluation of clustering methods. *Journal of the American Statistical Association*, *66*, 846–850.

Rodriguez, M., Ali, S., Kanade, T. (2009). Tracking in unstructured crowded scenes. In *Proceedings ICCV*.

Rodriguez, M., Sivic, J., Laptev, I., Audibert, J. (2011). Data-driven crowd analysis in videos. In *Proceedings ICCV*.

Saleemi, I., Hartung, L., Shah, M. (2010). Scene understanding by statistical modeling of motion patterns. In *Proceedings CVPR*.
Saligrama, V., Chen, Z. (2012). Video anomaly detection based on local statistical aggregates. In *Proceedings CVPR*.
Schneider, P., & Eberly, D. H. (2003). *Geometric Tools for Computer Graphics*. San Francisco: Morgan Kaufmann.
Scovanner, P., Tappen, M. (2009). Learning pedestrian dynamics from the real world. In *Proceedings ICCV*.
Shumway, R., & Stoffer, D. (1982). An approach to time series smoothing and forecasting using the EM algorithm. *Journal of time series analysis*, *3*(4), 253–264.
Stauffer, C. (2003). Estimating tracking sources and sinks. In *Proceedings CVPR Workshop*.
Tomasi, C., Kanade, T. (1991). Detection and Tracking of Point Features. *International Journal of Computer Vision*.
Treuille, A., Cooper, S., & Popović, Z. (2006). Continuum crowds. *ACM SIGGRAPH*, *25*(3), 1160–1168.
Vicsek, T., Czirók, A., Ben-Jacob, E., Cohen, I., & Shochet, O. (1995). Novel type of phase transition in a system of self-driven particles. *Physical Review Letters*, *75*(6), 1226–1229.
Wang, M., Wang, X. (2011). Automatic adaptation of a generic pedestrian detector to a specific traffic scene. In *Proceedings CVPR*.
Wang, X., Tieu, K., Grimson, W. (2006). Learning semantic scene models by trajectory analysis. In *Proceedings ECCV*.
Wang, X., Ma, K., Ng, G., Grimson, W. (2008a). Trajectory analysis and semantic region modeling using a nonparametric bayesian model. In *Proceedings CVPR*.
Wang, X., Ma, X., & Grimson, W. (2008b). Unsupervised activity perception in crowded and complicated scenes using hierarchical bayesian models. *IEEE Transactions on PAMI*, *31*(3), 539–555.
Wang, X., Ma, K., Ng, G., & Grimson, W. (2011). Trajectory analysis and semantic region modeling using nonparametric hierarchical bayesian models. *International Journal of Computer Vision*, *95*(3), 287–312.
Wu, S., Moore, B.E., Shah, M. (2010). Chaotic invariants of lagrangian particle trajectories for anomaly detection in crowded scenes. In *Proceedings CVPR*.
Yamaguchi, K., Berg, A.C., Ortiz, L.E., Berg, T.L. (2011). Who are you with and where are you going? In *Proceedings CVPR*.
Yang, Y., Liu, J., Shah, M. (2009). Video scene understanding using multi-scale analysis. In *Proceedings ICCV*.
Zen, G., Ricci, E. (2011). Earth mover's prototypes: a convex learning approach for discovering activity patterns in dynamic scenes. In *Proceedings ICCV*.
Zhao, X., Medioni, G. (2011). Robust unsupervised motion pattern inference from video and application. In *Proceedings ICCV*.
Zhou, B., Wang, X., Tang, X. (2011). Random field topic model for semantic region analysis in crowded scenes from tracklets. In *Proceedings CVPR*.
Zhou, B., Tang, X., Wang, X. (2012a). Coherent filtering: detecting coherent motions from crowd clutters. In *Proceedings ECCV*.
Zhou, B., Wang, X., Tang, X. (2012b). Understanding collective crowd behaviors: Learning a mixture model of dynamic pedestrian-agents. In *Proceedings CVPR*.
Zhou, B., Tang, X., Wang, X. (2013). Measuring crowd collectiveness. In *Proceedings CVPR*.
Zhou, B., Tang, X., Zhang, H., Wang, X. (2014). Measuring crowd collectiveness. *IEEE Transactions on PAMI*.
Zhou, S., Chen, D., Cai, W., Lyo, L., Yoke, M., Hean, L., et al. (2010). Crowd modeling and simulation technologies. *ACM Transactions on Modeling and Computer Simulation*, *20*(4), 20.