Understanding Pedestrian Behaviors from Stationary Crowd Groups

Shuai Yi[1] Hongsheng Li[1,2] Xiaogang Wang[1]
[1]Department of Electronic Engineering, The Chinese University of Hong Kong
[2]School of Electronic Engineering, University of Electronic Science and Technology of China
syi@ee.cuhk.edu.hk lihongsheng@gmail.com xgwang@ee.cuhk.edu.hk

Abstract

Pedestrian behavior modeling and analysis is important for crowd scene understanding and has various applications in video surveillance. Stationary crowd groups are a key factor influencing pedestrian walking patterns but was largely ignored in literature. In this paper, a novel model is proposed for pedestrian behavior modeling by including stationary crowd groups as a key component. Through inference on the interactions between stationary crowd groups and pedestrians, our model can be used to investigate pedestrian behaviors. The effectiveness of the proposed model is demonstrated through multiple applications, including walking path prediction, destination prediction, personality classification, and abnormal event detection. To evaluate our model, a large pedestrian walking route dataset [1] is built. The walking routes of $12,684$ *pedestrians from a one-hour crowd surveillance video are manually annotated. It will be released to the public and benefit future research on pedestrian behavior analysis and crowd scene understanding.*

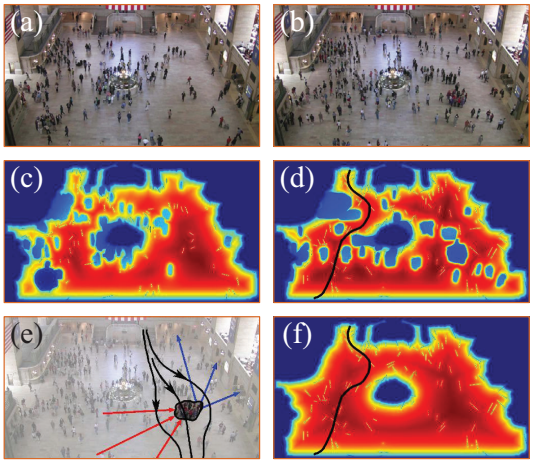


Figure 1. (a)-(b) Two video frames. (c)-(d) Energy maps calculated from (a) and (b) using the proposed model. Pedestrians are more likely to walk through regions with warm colors. (e) An illustration of multiple roles of a stationary crowd group. It can serve as source (blue lines), destination (red lines), and obstacle (black lines). (f) Energy map calculated from (b) without modeling the factor of stationary crowd groups.

1. Introduction

Pedestrian behavior modeling and analysis is important in video surveillance and has drawn increasing attentions in recent years. It can be used for various applications including pedestrian walking path prediction [4, 45], traffic flow segmentation [37, 41, 43], crowd counting and segmentation [40], and abnormal event detection [25, 27].

Pedestrian behavior modeling is challenging, especially for scenes with crowds [8, 21]. Previous studies [5, 12, 42, 45] have shown that the walking behavior of an individual can be influenced by a variety of factors including scene layout (*e.g.* entrances, exits, walls, and obstacles), pedestrian beliefs (the choice of source and destination), and interactions with other moving pedestrians. However, an important factor, *i.e.* stationary crowd groups, is missing in literature of modeling pedestrian behaviors.

We argue that stationary crowd groups have considerable influence on pedestrians and are crucial in pedestrian behavior modeling. As shown in Figure 1 (d), the walking path of a pedestrian (black curve) is affected by a stationary crowd group. However, without modeling the stationary crowd group, it is difficult to explain why the pedestrian detours when approaching the destination, as shown in Figure 1 (f).

Studies also show that stationary crowd groups have greater influences on pedestrian behaviors than moving crowds [28, 38, 39]. A pedestrian usually changes the walking speed rather than direction to avoid collision with other moving crowds. However, when moving crowds become stationary, the walking pedestrian is forced to changed his or her direction and the walking path is influenced significantly.

As shown in Figure 1 (e), stationary crowd groups can

[1]Available at http://www.ee.cuhk.edu.hk/∼syi/

serve as multiple roles for different pedestrians. For pedestrians that are leaving or joining a stationary crowd group, it can be regarded as the source or the destination (red and blue curves). For other pedestrians that are moving near the stationary crowd group, it can be regarded as an obstacle (black curves). Although both stationary crowd groups and fixed scene obstacles can block traffic, a pedestrian can choose to walk through the stationary crowd group or to detour from it, while scene obstacles are solid and cannot be penetrated. Moreover, as shown in Figure 1 (a)-(d), the spatial distribution of stationary crowd groups might change over time, which leads to the dynamic variations of traffic patterns. Therefore, static models cannot be used for stationary crowd group modeling.

In our work, the factor of stationary crowd groups is introduced for the first time to model pedestrian behaviors. Both walking through and walking bypass pedestrians can be well modeled. The proposed model can be dynamically updated with time to adapt the change of stationary crowd groups.

Based on our model, we can investigate the influence of stationary crowd groups on pedestrian behaviors. By learning model parameters, we observe that stationary crowd groups have greater influence on pedestrian walking paths than moving crowds, which shows the importance of monitoring stationary groups in a traffic control system. Moreover, by modeling the interactions among stationary groups and moving pedestrians, a personality attribute is proposed to classify pedestrians into different categories. This attribute is a key factor that makes each individual behave differently. One interesting observation is that people are more likely to behave in a conservative way when the scene is not that crowded. In contrast, a crowded scene leads to aggressive walking patterns because of the lack of space.

The contribution of this work is summarized as below. (1) A novel model is proposed for pedestrian behavior modeling by including stationary crowd groups as a key component. Through inference based on the interactions between stationary crowd groups and pedestrians, our model can be used to investigate pedestrian behaviors. (2) A large pedestrian walking path dataset is built. The walking routes of more than $12,000$ pedestrians from a one-hour crowd video are annotated. (3) The effectiveness of the proposed model is demonstrated by multiple applications on the proposed dataset, including pedestrian walking path prediction (Section 5.1), pedestrian destination prediction (Section 5.2), pedestrian personality estimation and classification (Section 5.3), and abnormal event detection (Section 5.5).

2. Related Work

A lot of works have been done on modeling crowd motion patterns and segmenting traffic flows. Lagrangian coherent structures [2] and Lie algebra representation [22, 23] were used for flow field computation and segmentation. Topic models have been widely used [18, 37] for crowd flow modeling and estimation. Spatio-temporal dependency on motion patterns could be included in topic models [7, 13, 14]. Motion patterns could also be discovered through clustering trajectories [15, 16, 24, 26, 36, 44]. Shao *et al*. [33, 34] characterized the generic properties of crowd systems by modeling the coherent motion crowd groups.

Agent based models [5] are in a different category, in the sense that they model the decision making process of individuals. A typical example is the social force model [12], which was originally proposed for crowd simulation [11], and then was used in tracking [29], interaction analysis [31], and abnormal event detection [25].

Our proposed model is also agent-based. Existing agent based models use pre-defined rules to control each individual's walking behavior, which can be used for simulation and prediction. However, they have three main shortcomings compared with our model. First, stationary crowd groups are ignored in all these models. A lot of research works have been done on analyzing moving social groups [6, 9, 19, 20]. The stationary crowd group is lack of attention, although it has great influence on pedestrian behaviors. Second, most of these agent based models are static models which cannot be dynamically updated with time. However, the influence factors are changing and pedestrian interactions also need to be updated with time. Third, most existing methods cannot model personality, which is a key factor that makes each individual behave differently [10, 30].

Recently, Alahi *et al*. [1] built a large scale crowd dataset for forecasting pedestrian destinations. However, that dataset only provides trajectories of moving pedestrians without video frames, and cannot be used to study the influence of stationary crowds on pedestrian behaviors. Therefore, we built a new crowd dataset with both manually annotated trajectories and video frames.

3. Pedestrian Behavior Modeling

Human walking path selection is similar to water flow. A pedestrian usually selects the most convenient and efficient path for reaching the destination. Based on this assumption, a general scene energy map $\mathcal{M}$ is proposed to model the traveling difficulty of every location of the scene.

Regions with higher energy values denote that pedestrians are energetic at these locations and can travel through these locations more easily. More pedestrians tend to choose their walking paths through, and therefore the probability of observing pedestrians at these locations should be higher. Lower energy values indicate locations with lower occurrence probability of pedestrians. For example, areas near an obstacle or inside a stationary crowd group are difficult to walk through. The probability of observing pedestrians at these locations is lower.

In our model, scene layout, moving pedestrians, and stationary groups are included. Different factors may have different effects on pedestrian decision making. Their influence weights are learned from training data and reflect the importance of these factors.

Personalized energy maps $\mathcal{M}_P$ are generated based on the general energy map $\mathcal{M}$ and a personality parameter P. $\mathcal{M}_P$ can be viewed as different pedestrians' interpretations of the general map $\mathcal{M}$. Given a source and a destination, the fast marching algorithm [17, 32] is used to generate an optimal walking path in the energy map.

3.1. General energy map modeling

A general energy map $\mathcal{M}$ can be modeled with three channels calculated based on Scene Layout, Moving Pedestrians, and Stationary Groups. These channels are represented by f_{SL}, f_{MP}, and f_{SG}. $\mathcal{M}(x)$ can be pixel-wisely modeled by combining the channels,

$$\mathcal{M}(x;\Theta) = f_{SL}(x;\theta_1)f_{MP}(x;\theta_2)f_{SG}(x;\theta_3,\theta_4), \quad (1)$$

where $\Theta = [\theta_1,\theta_2,\theta_3,\theta_4]^T$ are weighting parameters for different terms. $\mathcal{M}$ is also a probability map and can be used as the probability of pedestrian appearing at each location. It can be extended by including new channels.

3.2. Scene layout factor

Pedestrian's walking behavior is constrained by scene layout. Pedestrians cannot walk freely in a scene due to the constraints of walls and other static obstacles, and therefore they cannot be observed at some locations. More over, people tend to keep a distance from these obstacles and are not likely to walk very close to them, and thus the probability of observing a pedestrian decreases when getting close to the obstacle regions.

The Scene Layout influence map is therefore modeled as

$$f_{SL}(x;\theta_1) = \exp\left(-\frac{\theta_1}{d_1(x,SL)}\right), \quad (2)$$

where SL is a set of locations occupied by scene obstacles which are unreachable, $d_1(x,SL) = \min_{y\in SL}||x-y||_2^2$ measures the distance from the current location x to its nearest scene obstacle location y, and θ_1 is a parameter indicating the influence bandwidth (which also can be viewed as the importance) of the scene layout term.

If $x \in SL$, there is an obstacle at location x, and $d_1(x,SL) = 0$. In this case, $f_{SL}(x;\Theta)$ is equal to 0, which means that pedestrians cannot appear at location x. When $x \notin SL$, $d_1(x,SL) > 0$. $f_{SL}(x;\Theta)$ gets close to 0 when the current location x approaches to scene obstacles. An example of a scene layout map is shown in Figure 2.

3.3. Influence of moving pedestrians

The interaction with other moving pedestrians is another factor to be considered. A pedestrian tends to keep certain distance from others. As a result, there is a probability drop around the regions occupied by pedestrians.

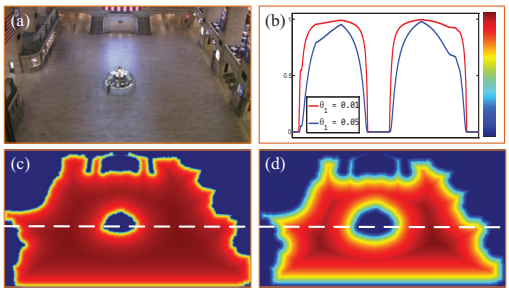


Figure 2. Example of scene layout influence maps. (a) Scene background. (b) The energy values along the white horizontal lines in (c) and (d). The color bar indicates the energy values displayed in (c) and (d). (c)-(d) Two scene influence maps calculated by setting θ_1 as 0.01 and 0.05, respectively. Energy drops near the scene boundaries.

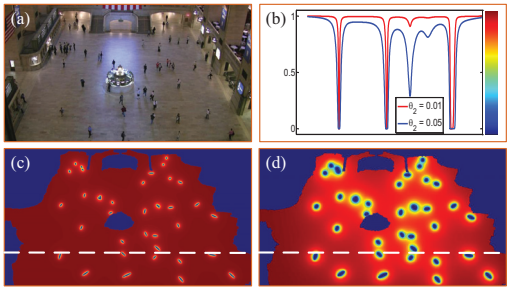


Figure 3. Example of moving pedestrian influence maps. (a) A video frame. (b) The energy values along the white horizontal lines in (c) and (d). (c)-(d) Two moving pedestrian influence maps calculated by setting θ_2 as 0.01 and 0.05, respectively. Energy drops around moving pedestrian.

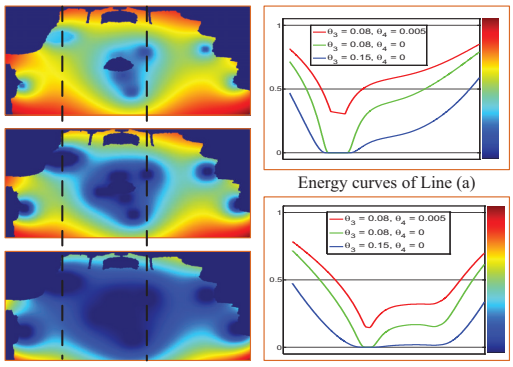
The Moving Pedestrian influence map is modeled as

$$f_{MP}(x;\theta_2) = \exp\left(-\sum_{i=1}^{m}\frac{\theta_2}{d_2(x,MP_i)}\right), \quad (3)$$

where MP_i ($i \in [1,m]$) is the ith moving pedestrian, $x_t^{MP_i}$ is the spatial location of MP_i at current time t, $x_{t+1}^{MP_i}$ is used to estimate the spatial location of MP_i at time $t+1$, $d_2(x,MP_i) = (||x-x_t^{MP_i}|| + ||x-x_{t+1}^{MP_i}||)^2 - (||x_t^{MP_i} - x_{t+1}^{MP_i}||)^2$ measures the distance from the current location x to the moving pedestrian MP_i, and θ_2 is the influence bandwidth of the moving pedestrian term. We use the same distance metric as the social force model [11]. An example of a moving pedestrian influence map is shown in Figure 3.

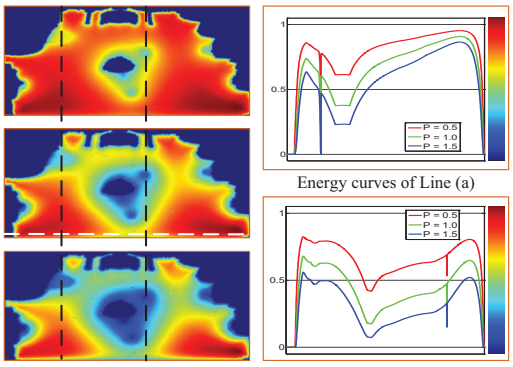
3.4. Influence of stationary crowd groups

Stationary crowd groups are modeled in two aspects. First, for pedestrians that bypass a stationary crowd group, this stationary crowd group acts similarly as a scene obstacle. The group has a repulsive force around the group region to keep moving pedestrians away. Second, for pedestrians



(a) (b)

Figure 4. Example of stationary crowd group influence maps. (Left) Three stationary crowd group influence maps calculated from the same frame by using different θ_3 and θ_4. (Right) The energy values alone two vertical lines (a) and (b) in (Left). Comparing the two red curves, we notice that the stationary group regions may have non-zero energy values by setting a nonzero θ_4. Different groups may have different energy values due to the density differences. By setting $\theta_4 = 0$, the differences disappear and the energy values inside the groups turn to zero.



(a) (b)

Figure 5. Example of a personalized maps. (Left) Three personalized maps calculated from the same frame using different P (P is set as 0.5, 1.0, 1.5 from top to bottom). (Right) The energy values alone line (a) and (b) in (Left).

that walk through a stationary crowd group, there should be a penalty inside the group region. This is the key difference with the scene layout factor, where obstacles cannot be penetrated. The penalty is related to crowd density. It is more difficult to walk through denser stationary crowds.

The Stationary Group influence map is modeled as

$$f_{SG}(x;\theta_3,\theta_4) = \exp\left(-\sum_{i=1}^{n}\frac{\theta_3}{d_3(x,SG_i)+\theta_4 d_4(SG_i)}\right), \quad (4)$$

where $SG_i (i \in [1,n])$ is the ith stationary crowd group region automatically detected by using the approach proposed in [39], $d_3(x,SG_i) = \min_{y\in SG_i} ||x-y||_2^2$ measures the distance from x to the stationary crowd group region SG_i, θ_3 is the influence bandwidth of the stationary crowd group term, and $d_4(SG_i) \in (0,+\infty)$ is used to measure the sparsity of stationary crowd group region SG_i. d_4 is calculated as the average distance among group members. Larger d_4 represents lower crowd density. The weight θ_4 controls the influence of group sparsity on estimation result.

If $x \in SG_i$, the location x is inside SG_i, and $d_3(x,SG_i) = 0$. $f_{SG}(x;\Theta)$ at locations $x \in SG_i$ inside the group is constant and is positively correlated with group sparsity $d_4(SG_i)$. $f_{SG}(x;\Theta)$ is in the range of $(0,1)$, which means that the probability of observing a pedestrian walking through the group region decreases because of the influence of the stationary group, but it is still larger than 0. If $x \notin SG_i$, x is outside SG_i, and $d_3(x,SG_i) > 0$. The influence value increases from group boundary to faraway locations. An example of a stationary crowd group influence map is shown in Figure 4.

3.5. Personalized energy map modeling

People might behave differently under the same situation. It is modeled by a personality parameter P. Different personalized energy maps $\mathcal{M}_P$ are generated based on the general energy map $\mathcal{M}$ with different P values,

$$\mathcal{M}_P(x;\Theta) = \exp\left(P \times \ln \mathcal{M}(x;\Theta)\right). \quad (5)$$

If P is large for a pedestrian, the influence bandwidth of all the terms $(\theta_1,\theta_2,\theta_3)$ would equivalently increase for this individual. The energy values are small at locations near obstacles and stationary crowd groups. It denotes that this pedestrian cares more about these influence factors and is likely to walk a longer way to avoid close contact with these obstacles. In contrast, smaller P means that the pedestrian is walking aggressively and cares less about obstacles. An example of a personalized map is shown in Figure 5.

3.6. Path generation

To generate pedestrian walking paths, Fast Marching [17, 32] is used. Given the source x_s and the destination x_d, an optimal path $\widehat{T}$ is calculated based on the energy map $\mathcal{M}$ or $\mathcal{M}_P$:

$$\widehat{T} = f_{FM}(\mathcal{M}(\mathcal{M}_P), x_s, x_d), \quad (6)$$

where $\widehat{T}$ is the most efficient and probable route from x_s to x_d according to the current energy map $\mathcal{M}$ or $\mathcal{M}_P$. Several examples are shown in Figure 6. When using a personalized map $\mathcal{M}_P$, the optimal path is just for the specific individual. When using a general map $\mathcal{M}$, the optimal path can be regarded as an average path for ordinary pedestrians.

3.7. Model learning

For a scene, model parameters Θ need to be estimated from training data. The values in an energy map $\mathcal{M}$ represent the probabilities of pedestrian appearing at locations.

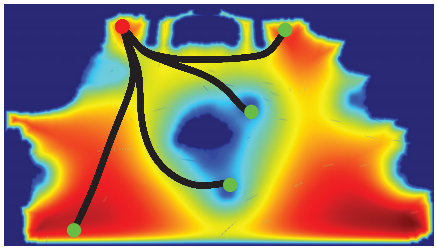


Figure 6. Example of path generation. The red point is a source. Green points are destinations. Black curves are optimal walking routes calculated by Equation (6).

Therefore, model parameters can be optimized by maximizing likelihood on the training data. A general energy map $\mathcal{M}$ is built based on Equation (1). By dividing a marginalization term, $Z(\Theta)$, the energy map $\mathcal{M}(x;\Theta)$ can be transformed to a probability distribution,

$$p(x;\Theta) = \frac{1}{Z(\Theta)}\mathcal{M}(x;\Theta), \tag{7}$$

where $Z(\Theta) = \int \mathcal{M}(x;\Theta)dx$.

Given $X = \{x_1, ..., x_k, ..., x_K\}$ as K independent observations of x, the likelihood of these observations is

$$p(X;\Theta) = \prod_{k=1}^{K} \frac{1}{Z(\Theta)}\mathcal{M}(x_k;\Theta). \tag{8}$$

Parameter Θ can then be optimized as

$$\widehat{\Theta} = \arg\max \log p(X;\Theta). \tag{9}$$

Gradient descent is used for updating parameters.

$$\Theta_{new} = \Theta_{old} + \eta \frac{\partial \log p(X;\Theta_{old})}{\partial \Theta_{old}}. \tag{10}$$

4. Pedestrian walking route dataset

4.1. Dataset details

Pedestrian walking route data with accurate annotation can be used for model learning and evaluation. However, automatically tracking, especially in crowded scenes, is not accurate. Several existing datasets [1, 3, 35] have limitations and cannot be used in our study. Most of these datasets are not long enough, not crowded enough, or do not contain enough pedestrians. The dataset proposed by [1] is large, but contains only trajectories without video frames, and thus cannot be used to detect and analyze stationary crowds.

A new large scale pedestrian walking route dataset [2] is built in this work. Accurate pedestrian walking routes from a one-hour crowd video were manually annotated as ground truth. The video was from the dataset released in [44]. The details of the dataset are summarized in Table 1. Twelve example pedestrian walking routes are shown in Figure 7.

[2] Available at http://www.ee.cuhk.edu.hk/∼syi/

Resolution (pixel)	1,920 × 1,080
Total frame number	100,000
Frame rate (fps)	25
Annotated frame number	5,000
Annotated frame rate (fps)	1.25
Annotated pedestrian number	12,684
Average pedestrian number per frame	123
Max pedestrian number per frame	332

Table 1. The details of the proposed dataset.

Figure 7. Twelve examples of annotated walking routes.

The proposed dataset has several advantages compared with existing ones. First, our dataset is much longer than any existing one with ground truth on tracking. Long-term traffic flow change can be observed from our dataset and it contains rich information to train a complex model of pedestrian behaviors. Second, it is a crowd surveillance dataset which is difficult and challenging for vision tasks. An average of 123 pedestrians can be observed in each frame. The most crowed frame contains 332 pedestrians. Complex crowd behaviors can be observed in this dataset. Third, this dataset is well annotated. All the $12,684$ pedestrians in this video are manually annotated. For each individual, the complete trajectory from the time point he/she enters the scene to the time he/she leaves is labeled. The large amount of data with accurate annotation is crucial for comprehensive evaluation and convincing statistical analysis. Besides pedestrian behavior modeling, our dataset can be used in various research areas, such as pedestrian detection, individual tracking, crowd segmentation, density estimation, and pedestrian counting.

4.2. Statistical analysis of the annotated data

A lot of statistical information can be obtained from this dataset, and such information is valuable for the design of the supervised model. The influence of stationary crowds on pedestrian walking efficiency is analyzed in Figure 8. We record the dynamic changes of (a) the percentage of stationary pedestrians and (b)-(c) two efficiency measurements. Larger values in (b)-(c) indicate lower efficiency. The strong correlations between (a) and (b)-(c) indicate that stationary crowd is a key factor that decreases traffic efficiency. In contrast, the correlations between total crowd density and (b)-(c) are much weaker. If every pedestrian is

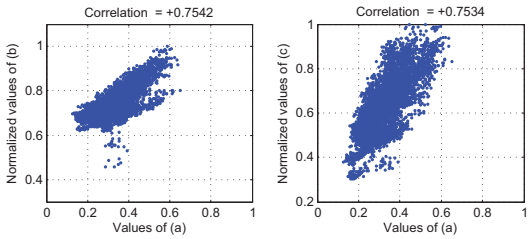


Figure 8. Correlations between stationary crowds and traffic efficiency. (a) Percentage of stationary pedestrians. (b) Average walking path length. (c) Average traveling time. The correlations between (a) and (b)-(c) are $+0.754$, and $+0.753$. The correlations between total crowd density and (b)-(c) are -0.061, and -0.121.

Parameter	Meaning	Learned value
θ_1	Scene layout weight	0.016
θ_2	Moving pedestrian weight	0.023
θ_3	stationary crowd group weight	0.390
θ_4	Group density weight	0.006

Table 2. Learned parameter values for the proposed dataset.

moving, traffic flow is smooth and efficient even when the scene is very crowded. However, when stationary crowd groups appear, the traffic efficiency might be dramatically reduced.

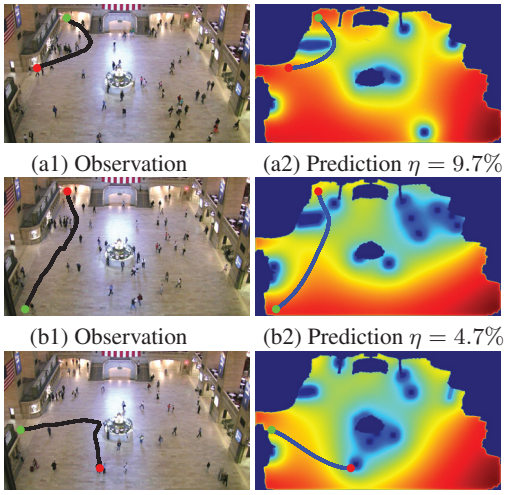
4.3. Learning result

The trajectories of moving pedestrians are used as training samples to learn model parameters. The optimized parameters are shown in Table 2. Comparing θ_3 with θ_1 and θ_2, we observe that the stationary crowd groups have greater influence on pedestrian walking behaviors than scene layout and moving pedestrians. The learned θ_4 is greater than zero, which indicates that the stationary crowd group density does influence pedestrian behaviors.

A pedestrian is not sensitive to scene obstacles, because scene obstacles can never move and he/she does not need to consider possible collisions with these obstacles. A pedestrian might prefer to adjust walking speed rather than change pre-decided walking direction to avoid close contact with other moving pedestrians. The walking path might be slightly changed but the influence is not obvious. When stationary crowds emerge in front of a pedestrian, he/she has to change his/her walking route to bypass the stationary crowds. This is the reason why stationary crowd group influence weight θ_3 is much larger than scene layout weight θ_1 and moving pedestrian weight θ_2.

5. Applications

Based on the proposed model, inference and learning algorithms, various applications can be implemented and interesting characteristics about human walking behaviors can be revealed.



(a1) Observation (a2) Prediction $\eta = 9.7\%$

(b1) Observation (b2) Prediction $\eta = 4.7\%$

(c1) Observation (c2) Prediction $\eta = 74.6\%$

Figure 9. Examples of prediction results. (a)-(b) Predictions well match the observations and η is small. (c) A pedestrian walks in an unexpected path. The walking cost of the observation is much higher than that of the predicted path from optimization. Therefore, the over-cost is high ($\eta = 74.6\%$). This activity can be regarded as abnormal and η can be used for abnormal detection.

5.1. Prediction on pedestrian walking paths

Given a source x_s and a destination x_d, we predict an optimal walking route as $\widehat{T} = f_{FM}(\mathcal{M}, x_s, x_d)$ by minimizing Equation (6). In this application, we assume $P = 1$ as no prior on personality of pedestrians is given.

An over-cost value η is proposed to evaluate whether predictions match observations. For the optimized route $\widehat{T}$ and the observed route T_O, the walking costs are calculated based on the energy map $\mathcal{M}$. Over-cost η is defined as

$$\eta = \frac{\mathcal{C}(T_O, \mathcal{M}) - \mathcal{C}(\widehat{T}, \mathcal{M})}{\mathcal{C}(\widehat{T}, \mathcal{M})}, \tag{11}$$

where $\mathcal{C}(T_O, \mathcal{M})$ is the walking cost of the observed route T_O based on the current map $\mathcal{M}$, and $\mathcal{C}(\widehat{T}, \mathcal{M})$ is the cost of the optimized route $\widehat{T}$. η should be nonnegative because $\mathcal{C}(T_O, \mathcal{M})$ is no smaller than $\mathcal{C}(\widehat{T}, \mathcal{M})$, and smaller η indicates better match. Examples of prediction results are shown in Figure 9.

Two baselines are used to investigate the effectiveness of the proposed model considering stationary crowd groups. For the first baseline, we set $\theta_3 = 0$, which denotes that the stationary crowd group factor is removed. For the second one, we set $\theta_4 = 0$, which denotes that the stationary crowd groups are simply regarded as solid obstacles. The η distributions of our method and the baselines are shown in Figure 10. The average over-cost of our method is 5.86%, while the over-costs of the baselines are much high, *i.e.* 15.44% and 16.22%, respectively. Because abnormal pedestrians may have relatively high η, we also calculate the average over-

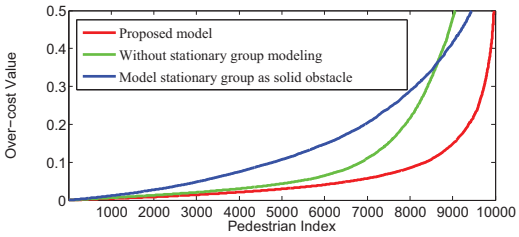


Figure 10. Distributions of over-cost η. Over-cost values of $10,000$ pedestrians are calculated with our model and the two baselines. They are sorted in an increasing order for each method.

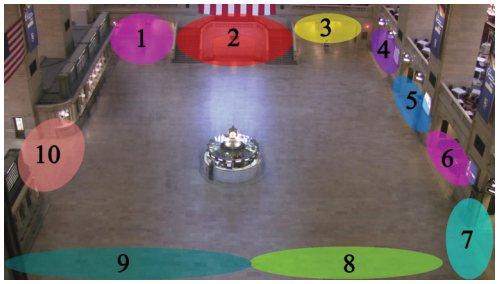


Figure 11. Ten source / destination regions of the scene.

cost of 80% pedestrians with smaller η. For these pedestrians, the average over-cost of our method is 2.70%, while over-costs of the baselines are 4.79% and 9.46%. From the results, we can conclude that including the influence of stationary crowd groups is necessary when modeling pedestrian behaviors, and the stationary crowd groups should be modeled differently from scene obstacles.

5.2. Prediction of pedestrian destinations

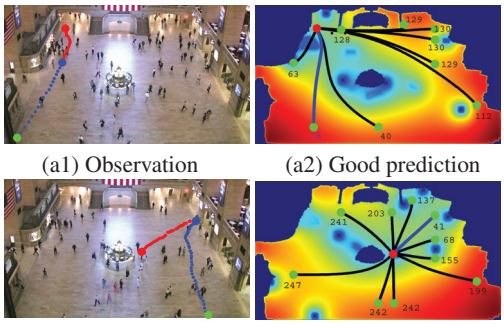
The source x_s, the destination x_d, and the walking path T are the three basic elements of pedestrian behaviors. In Section 5.1, we predict T based on x_s and x_d. Given x_s and part of the walking path, we can also predict the destination of this pedestrian.

Ten source/destination regions $\mathcal{S}_i (i \in [1, 10])$ are manually labeled as shown in Figure 11. The first half of observed trajectory $T_{0.5}$ is used as input in this experiment. Given x_s and $T_{0.5}$, the task is to estimate the destination index $i \in [1, 10]$. For each destination region $\mathcal{S}_i$, $\mathcal{L}(i)$ is calculated as $\mathcal{L}(i) = \min_{x'_d \in \mathcal{S}_i} \mathcal{D}(T_{0.5}, \widehat{T_{0.5}}(x'_d))$, where $\widehat{T_{0.5}}(x'_d)$ is the first half of $\widehat{T}(x'_d)$, $\widehat{T}(x'_d) = f_{FM}(\mathcal{M}, x_s, x'_d)$ is the optimized route ended with x'_d, and $\mathcal{D}(\cdot, \cdot)$ represents the distance between the two half trajectories. Smaller $\mathcal{L}(i)$ indicates that the pedestrian is more likely to go to the destination $\mathcal{S}_i$. Then the index of estimated destination is obtained as $\widehat{i} = \arg\min_{i \in [1,10]} \mathcal{L}(i)$.

The *top N accuracy* (ground truth is within the top N predictions) is adopted for evaluation. The MDA model [45], together with the two baselines introduced in Section 5.1 are used for comparison. Estimation results are shown in Table 3. The destinations of 48% pedestrians can be successfully predicted at the first trial using our algorithm.

N	1	2	3	4	5
Proposed	**48%**	69%	83%	90%	93%
$\theta_3 = 0$	38%	68%	82%	90%	93%
$\theta_4 = 0$	33%	57%	71%	80%	87%
MDA [45]	43%	-	-	-	-

Table 3. Accuracy of destination prediction. *Top N accuracy* is calculated. Our method can obtain better results than others.



(a1) Observation (a2) Good prediction

(b1) Observation (b2) Abnormal case

Figure 12. Two examples of destination prediction. (a1)-(b1) Observed walking routes. The first half (red part) of the observed walking path is used as input. (a2)-(b2) Prediction results. The values of $\mathcal{L}_i$ for different destinations are shown. The estimated destination is chosen as the one with minimal $\mathcal{L}_i$ (blue curve). In example (b), our model makes wrong prediction due to the sudden and unexpected turning of the pedestrian.

Two examples of destination prediction are shown in Figure 12. For some pedestrians, the walking paths are unusual and it is difficult to estimate the correct destinations at first several trials. These walking routes can be regarded as abnormal. In this way, destination prediction can be used for abnormal behavior detection.

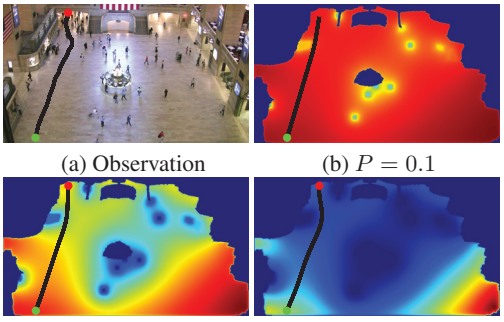
5.3. Personality attribute estimation

The personality P of each individual can be estimated by

$$\widehat{P} = \arg\min_{P} \mathcal{D}(T_O, \widehat{T}(P)), \tag{12}$$

where T_O is the observed trajectory of current pedestrian, $\widehat{T}(P) = f_{FM}(\mathcal{M}_P(P), x_s, x_d)$ is the optimal walking path calculated using personalized energy map $\mathcal{M}_P$ in Equation (6), and $\mathcal{D}(\cdot, \cdot)$ represents the distance between the two trajectories. The estimated personality parameter $\widehat{P}$ minimizes the difference between observation T_O and optimal path $\widehat{T}(P)$. An example is shown in Figure 13.

The distribution of P is shown in Figure 14. All the pedestrians can be classified into three categories based on their walking behaviors: aggressive, conservative, and abnormal. The peak A in Figure 14 represents aggressive pedestrians who prefer to walk directly to their destinations. Conservative pedestrians are represented by the peak



(a) Observation (b) $P = 0.1$

(c) $P = 1.0$ (d) $P = 1.5$

Figure 13. (b)-(d) The predicted walking paths and their corresponding personalized energy maps $\mathcal{M}_P$. A smaller P may lead to a more straight walking route while a larger P may lead to a longer walking path that keeps away from stationary crowds. This pedestrian is walking in a conservative manner because the observation shown in (a) is more similar to the walking path in (d) with large P.

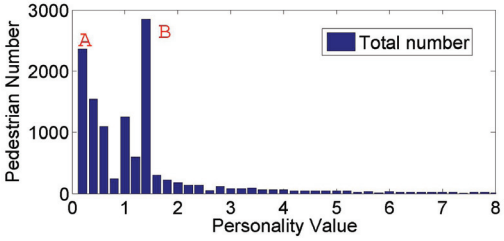


Figure 14. Personality value distribution. Two peaks A and B, and the long tail can be observed in this figure. The peak B is larger than 1 and the peak A is smaller than 1. When $P = 1$, the personalized map is degenerated into the general energy map.

B. They prefer to walk in a longer way to avoid close contact with others. The long tail of the distribution of P represents pedestrians that take a long route to their destination. Conservativeness is no longer proper to describe these pedestrians and we define these behaviors as abnormal.

5.4. Further investigation on personality

Personality attribute P can be used for pedestrian classification as different P may lead to different walking behaviors. All the pedestrians are annotated into three categories as ground truth, *i.e.* aggressive, conservative, and abnormal. Bayesian classifiers (that minimize classification errors) are used to classify these three categories. The leave-one-out evaluation results are shown in Table 4. Among all the annotated pedestrians, 87.43% are correctly classified using P as the feature value.

We also explore the relationship between the personality value P and the scene population density. The quantitative correlation between the two values is -0.44, and the dynamic changes of the two quantities are shown in Figure 15. The negative correlation shows that the personality value P is negatively related to the scene population density. This finding is reasonable. When the scene is too crowded, the

Category	Aggressive	Conservative	Abnormal
Total number	8,062	4,322	1,572
Correctly classified	7,123	3,668	1,411
Accuracy(%)	88.35	84.87	89.76

Table 4. Pedestrian classification results based on personality.

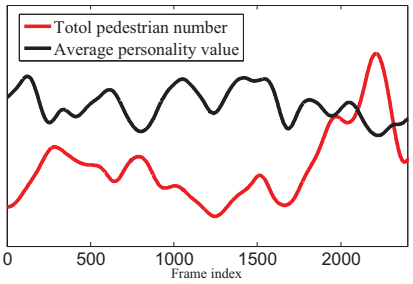


Figure 15. The relation between the scene population and average personality value. The black curve is the average personality value at each time. It is anti-related to the scene population density.

walking behaviors of pedestrians are constrained and there is no enough space for conservative walking patterns. In order to reach destinations, close contact with each other is unavoidable.

5.5. Abnormal behavior detection

Abnormal behaviors can be defined as unexpected observations which are significantly different from our predictions. Walking path prediction (Section 5.1) and destination prediction (Section 5.2) can both be used for abnormal detection. Examples have been shown in Figure 9(c) and 12(b). Moreover, personality estimation based abnormal detection has been introduced in Section 5.4.

6. Conclusion

In this paper, a novel pedestrian behavior model is proposed and the stationary crowd group influence is included as a key component. It is applied to various applications, including walking path prediction, destination prediction, personality estimation, and abnormal event detection. A new pedestrian walking route dataset is proposed and will benefit future studies on pedestrian behavior analysis.

Acknowledgment

This work is partially supported by the General Research Fund sponsored by the Research Grants Council of Hong Kong (Project Nos. CUHK 419412, CUHK 417011, and CUHK 14207814), Hong Kong Innovation and Technology Support Programme (Project reference ITS/221/13FP), Shenzhen Basic Research Program (JCYJ20130402113127496), NSFC (Project No. 61301269) and Sichuan High Tech R&D Program (No. 2014GZX0009).

References

[1] A. Alahi, V. Ramanathan, and L. Fei-Fei. Socially-aware large-scale crowd forecasting. In *Proc. CVPR*. IEEE, 2014.

[2] S. Ali and M. Shah. A lagrangian particle dynamics approach for crowd flow segmentation and stability analysis. In *Proc. CVPR*. IEEE, 2007.

[3] S. Ali and M. Shah. Floor fields for tracking in high density crowd scenes. In *Proc. ECCV*. Springer, 2008.

[4] G. Antonini, S. V. Martinez, M. Bierlaire, and J. P. Thiran. Behavioral priors for detection and tracking of pedestrians in video sequences. *International Journal of Computer Vision*, 69(2):159–180, 2006.

[5] E. Bonabeau. Agent-based modeling: Methods and techniques for simulating human systems. *Proceedings of the National Academy of Sciences of the United States of America*, 99(Suppl 3):7280–7287, 2002.

[6] M.-C. Chang, N. Krahnstoever, and W. Ge. Probabilistic group-level motion analysis and scenario recognition. In *Proc. ICCV*. IEEE, 2011.

[7] R. Emonet, J. Varadarajan, and J.-M. Odobez. Extracting and locating temporal motifs in video scenes using a hierarchical non parametric bayesian model. In *Proc. CVPR*. IEEE, 2011.

[8] D. Forsyth. *Group dynamics*. Cengage Learning, 2009.

[9] W. Ge, R. T. Collins, and R. B. Ruback. Vision-based analysis of small groups in pedestrian crowds. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 34(5):1003–1016, 2012.

[10] S. J. Guy, S. Kim, M. C. Lin, and D. Manocha. Simulating heterogeneous crowd behaviors using personality trait theory. In *Proceedings of the 2011 ACM SIGGRAPH/Eurographics Symposium on Computer Animation*. ACM, 2011.

[11] D. Helbing, I. Farkas, and T. Vicsek. Simulating dynamical features of escape panic. *Nature*, 407(6803):487–490, 2000.

[12] D. Helbing and P. Molnar. Social force model for pedestrian dynamics. *Physical review E*, 51(5):4282, 1995.

[13] T. Hospedales, S. Gong, and T. Xiang. A markov clustering topic model for mining behaviour in video. In *Proc. ICCV*. IEEE, 2009.

[14] T. M. Hospedales, J. Li, S. Gong, and T. Xiang. Identifying rare and subtle behaviors: A weakly supervised joint topic model. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 33(12):2451–2464, 2011.

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

[16] K. Kim, D. Lee, and I. Essa. Gaussian process regression flow for analysis of motion trajectories. In *Proc. ICCV*. IEEE, 2011.

[17] R. Kimmel, A. Amir, and A. M. Bruckstein. Finding shortest paths on surfaces using level sets propagation. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 17(6):635–640, 1995.

[18] D. Kuettel, M. D. Breitenstein, L. Van Gool, and V. Ferrari. What's going on? Discovering spatio-temporal dependencies in dynamic scenes. In *Proc. CVPR*. IEEE, 2010.

[19] T. Lan, L. Sigal, and G. Mori. Social roles in hierarchical models for human activity recognition. In *Proc. CVPR*. IEEE, 2012.

[20] T. Lan, Y. Wang, W. Yang, S. N. Robinovitch, and G. Mori. Discriminative latent models for recognizing contextual group activities. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 34(8):1549–1562, 2012.

[21] G. Le Bon. *The crowd: A study of the popular mind*. Macmillian, 1897.

[22] D. Lin, E. Grimson, and J. Fisher. Learning visual flows: A lie algebraic approach. In *Proc. CVPR*. IEEE, 2009.

[23] D. Lin, E. Grimson, and J. Fisher. Modeling and estimating persistent motion with geometric flows. In *Proc. CVPR*. IEEE, 2010.

[24] D. Makris and T. Ellis. Learning semantic scene models from observing activity in visual surveillance. *Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on*, 35(3):397–408, 2005.

[25] R. Mehran, A. Oyama, and M. Shah. Abnormal crowd behavior detection using social force model. In *Proc. CVPR*. IEEE, 2009.

[26] B. T. Morris and M. M. Trivedi. Trajectory learning for activity understanding: Unsupervised, multilevel, and long-term adaptive approach. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 33(11):2287–2301, 2011.

[27] M. Moussaïd, D. Helbing, and G. Theraulaz. How simple rules determine pedestrian behavior and crowd disasters. *Proceedings of the National Academy of Sciences*, 108(17):6884–6888, 2011.

[28] M. Moussaïd, N. Perozo, S. Garnier, D. Helbing, and G. Theraulaz. The walking behaviour of pedestrian social groups and its impact on crowd dynamics. *PloS one*, 5(4):e10047, 2010.

[29] S. Pellegrini, A. Ess, K. Schindler, and L. Van Gool. You'll never walk alone: Modeling social behavior for multi-target tracking. In *Proc. ICCV*. IEEE, 2009.

[30] L. A. Pervin. *The science of personality*. Oxford University Press, 2003.

[31] P. Scovanner and M. F. Tappen. Learning pedestrian dynamics from the real world. In *Proc. ICCV*. IEEE, 2009.

[32] J. A. Sethian. A fast marching level set method for monotonically advancing fronts. *Proceedings of the National Academy of Sciences*, 93(4):1591–1595, 1996.

[33] J. Shao, K. Kang, C. C. Loy, and X. Wang. Deeply learned attributes for crowded scene understanding. In *Proc. CVPR*. IEEE, 2015.

[34] J. Shao, C. C. Loy, and X. Wang. Scene independent group profiling in crowd. In *Proc. CVPR*. IEEE, 2014.

[35] G. Shu, A. Dehghan, and M. Shah. Improving an object detector and extracting regions using superpixels. In *Proc. CVPR*. IEEE, 2013.

[36] X. Wang, K. T. Ma, G.-W. Ng, and W. E. L. Grimson. Trajectory analysis and semantic region modeling using nonparametric hierarchical bayesian models. *International Journal of Computer Vision*, 95(3):287–312, 2011.

[37] X. Wang, X. Ma, and W. E. L. Grimson. Unsupervised activity perception in crowded and complicated scenes using hierarchical bayesian models. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 31(3):539–555, 2009.

[38] S. Yi and X. Wang. Profiling stationary crowd groups. In *Proc. ICME*. IEEE, 2014.

[39] S. Yi, X. Wang, C. Lu, and J. Jia. L0 regularized stationary time estimation for crowd group analysis. In *Proc. CVPR*. IEEE, 2014.

[40] C. Zhang, H. Li, X. Wang, and X. Yang. Cross-scene crowd counting via deep convolutional neural networks. In *Proc. CVPR*. IEEE, 2015.

[41] B. Zhou, X. Tang, and X. Wang. Measuring crowd collectiveness. In *Proc. CVPR*. IEEE, 2013.

[42] B. Zhou, X. Tang, and X. Wang. Learning collective crowd behaviors with dynamic pedestrian-agents. *International Journal of Computer Vision*, 111(1):50–68, 2015.

[43] B. Zhou, X. Tang, H. Zhang, and X. Wang. Measuring crowd collectiveness. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 36(8):1586–1599, 2014.

[44] B. Zhou, X. Wang, and X. Tang. Random field topic model for semantic region analysis in crowded scenes from tracklets. In *Proc. CVPR*. IEEE, 2011.

[45] B. Zhou, X. Wang, and X. Tang. Understanding collective crowd behaviors: Learning a mixture model of dynamic pedestrian-agents. In *Proc. CVPR*. IEEE, 2012.