Pedestrian Travel Time Estimation in Crowded Scenes

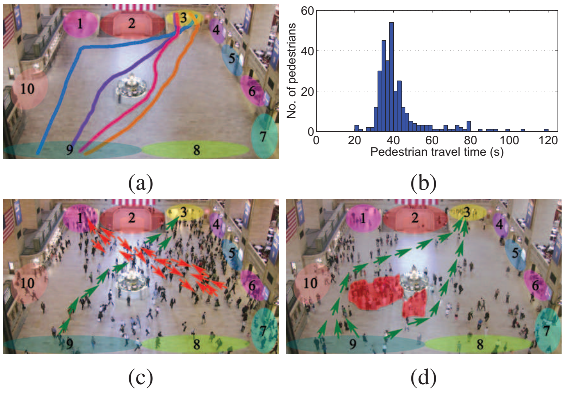
Shuai Yi [1] Hongsheng Li [1] Xiaogang Wang [1,2]
[1] Department of Electronic Engineering, The Chinese University of Hong Kong
[2] Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences
{syi,hsli,xgwang}@ee.cuhk.edu.hk

Abstract

In this paper, we target on the problem of estimating the statistic of pedestrian travel time within a period from an entrance to a destination in a crowded scene. Such estimation is based on the global distributions of crowd densities and velocities instead of complete trajectories of pedestrians, which cannot be obtained in crowded scenes. The proposed method is motivated by our statistical investigation into the correlations between travel time and global properties of crowded scenes. Active regions are created for each source-destination pair to model the probable walking regions over the corresponding source-destination traffic flow. Two sets of scene features are specially designed for modeling moving and stationary persons inside the active regions and their influences on pedestrian travel time. The estimation of pedestrian travel time provides valuable information for both crowd scene understanding and pedestrian behavior analysis, but was not sufficiently studied in literature. The effectiveness of the proposed pedestrian travel time estimation model is demonstrated through several surveillance applications, including dynamic scene monitoring, localization of regions blocking traffics, and detection of abnormal pedestrian behaviors. Many more valuable applications based on our method are to be explored in the future.[1]

1. Introduction

Crowd scene understanding [19, 20, 22, 29] and pedestrian behavior analysis [1, 3, 10, 17] are important for video surveillance. People would like to study scene properties and understand what is happening in the scene. In the meanwhile, they are also interested in revealing the rules governing individual behaviors. Scene information and pedestrian behaviors are correlated. Pedestrian travel time from an entrance to an exit is such a measurement that reflects information from both sides.



(a) (b)

(c) (d)

Figure 1. (a) Four examples of pedestrian walking routes from source 9 to destination 3. (b) Travel time distribution of all the 297 pedestrians from source 9 to destination 3 in a one-hour video. (c) The traffic flow from source 9 to destination 3 (shown in green) is intersected by the moving traffic (shown in red). (d) The traffic flow from source 9 to destination 3 (shown in green) is blocked by stationary persons (shown in red). The frames are extracted from the New York Grand Central Station dataset [30]. Ten source/destination regions are marked and numbered.

In crowd surveillance and traffic management systems, people care more about statistics of crowd population than individuals. Thus we estimate the average travel time of multiple pedestrians within a temporal window, such as one minute, instead of each individual. Such statistic reflects scene properties, and depends on scene structures, distributions of crowd densities and velocities. In crowded scenes, it is impossible to simply estimate the travel time based on the starting and ending points of pedestrian trajectories, because tracking fails frequently due to heavy occlusions. Instead, we will show that the travel time can be estimated from scene features that encode global scene properties, such as the spatial distributions of crowd densities and velocities. Features used by us can be computed from highly fragmented tracks and individual subjects are not required to be tracked over periods.

Pedestrian travel time between entrances and exits indicates traffic efficiency and travel cost of a scene, and thus attracts great attention in surveillance applications. When the travel time increases due to scene congestion, security ad-

[1] Project webpage can be found at http://www.ee.cuhk.edu.hk/∼syi/.

ministrators can take prompt actions, such as blocking some entrances until the congested crowds disperse, or opening extra exists, to control traffic. Travelers can also use such information to make plans. Travel time itself can be also regarded as an important feature to describe each individual's behavior and determine whether a pedestrian behaves normally or not.

However, estimating travel time is challenging, especially for scenes with crowds [7, 11]. Firstly, pedestrian travel time shows large inter-person variation. Even under the same situation and for the same source-destination pair, the walking paths and speed of individuals might be quite different, which leads to large variance of travel times. Figure 1(a) shows four examples of walking routes of pedestrians from source 9 to destination 3, with large diversity. The distribution of travel times of all the 297 pedestrians from source 9 to destination 3 in a one-hour video is shown in Figure 1(b). The mean of travel time of these pedestrians is 42.5 seconds, and the standard deviation is 13.5 seconds. Secondly, pedestrian decision making is complex and the travel time of individuals might be influenced by a variety of factors, such as the interactions with moving persons, stationary persons, and the scene layout. For example, as shown in Figure 1(c), the traffic flow from source 9 to destination 3 is intersected by the moving persons shown in red. Another example is shown in Figure 1(d), the traffic flow is blocked by the stationary persons shown in red. Both could increase the travel time. Lastly, the problem becomes much more challenging in crowded scenes, where existing computer vision techniques, such as pedestrian detection, tracking, and re-identification, cannot provide accurate results.

Some computer vision techniques, *e.g.* pedestrian simulation [10] and tracking [13, 26], can simulate or predict each individual's walking behavior, which can be used to obtain each individual's travel time. These methods have several limitations. Firstly, they rely on exact pedestrian information (*e.g.* speed, location, or appearance) to estimate the walking path, while obtaining these cues is difficult, especially in crowded situations. Secondly, they are local methods for individuals, *i.e.* pedestrians making decisions based on their local environment. However, the average travel time within a temporal window is a global property of many pedestrians and should be estimated with a global view of the whole scene.

Our method is specially designed for travel time estimation and it extracts features from the whole scene instead of complete trajectories of individuals. The estimation of travel time over source-destination traffic flows is regarded as a global scene modeling problem. Our method only requires fragmented trajectories (tracklets) and detection of stationary groups, both of which can be obtained even from crowded scenes. An active region is first computed for each source-destination pair to model the possible walking regions over the traffic flow. Features are then extracted from moving and stationary persons inside the active regions to describe scene properties and their influences on the travel times. Second-order polynomial regression is then adopted to map feature vectors to the values of travel time.

The contribution of this work is summarized as three-fold. (1) We propose a new research problem of estimating the statistic of travel time between entrances and exists without complete trajectories of pedestrians in crowded scenes. It provides useful information for both scene understanding and pedestrian behavior analysis. (2) A novel method is proposed by designing two new sets of scene features to describe the influences of moving and stationary persons on the travel time. (3) Several surveillance applications based on travel time estimation are introduced, *i.e.* monitoring scene dynamics, localizing of regions blocking traffics and detecting abnormal behaviors.

2. Related work

Crowd scene understanding is important in video surveillance. Multiple scene properties have been widely studied and modeled from different perspectives. There are a large number of works [2, 20, 24, 25, 29, 34, 35] proposed to discover major motion patterns of a scene and model scene dynamics. Some works [15, 27, 34, 38] focus on segmenting scene semantic regions and learning scene structures. Zhou *et al.* [37] and Shao *et al.* [23] studied scene-independent generic properties of crowds such as collectiveness, uniformity, purity and conflict. Studies on other important properties of crowd scenes include crowd counting [5, 8, 33], crowd density estimation [6, 28], and crossing-line flow rate estimation [14]. While most of these works focus on the properties of mobile pedestrians, Yi *et al.* [31, 32] recently showed that stationary groups play an equally important role in crowd scene understanding. In this paper, we propose to study the correlations between scene properties and pedestrian travel time, which can be viewed as a general description of the scene and can provide valuable information for scene understanding.

Many works have been done on analyzing pedestrian behaviors and predicting pedestrian walking patterns. However, existing pedestrian behavior models mainly consider where pedestrians are likely to walk instead of how long it takes to pass through the scene. Agent-based models [4] are a major category of behavior modeling techniques. The social force model [10] has been used for pedestrian simulation [9], pedestrian tracking [18], pedestrian interaction analysis [21], and abnormal behavior detection [16]. Statistic on pedestrian travel time is a global property of the whole scene while all these methods focus on the decision making process at each individual time step. Although these pedestrian behavior models can be used to estimate travel time by simulating pedestrian behaviors, our experimental results show that they generate inferior results compared with directly predicting the average travel time from other crowd scene properties.

Some existing computer vision techniques such as tracking [26] and person re-identification [12, 36] can be used for travel time estimation if they can provide the complete trajectories of pedestrians or can accurately match pedestrians at source and destination regions. However, they fail frequently in crowded scenes. A tracking failure at a single frame or a wrong matching of person re-identification may lead to completely wrong estimation of the travel time, which is also verified by our experiments.

3. Method

In this section, we introduce a method for estimating the travel time of pedestrians. For each time point t, we estimate an average travel time $T^t_{S,D}$ within a short period $[t-\tau, t+\tau]$, for pedestrians coming from a source S and going to a destination D. $T^t_{S,D}$ can be considered as a property of the scene. From the travel times between sources and destinations, one can better understand the current status of the scene. For clarity, we introduce our method by estimating the travel time between one source and one destination, and T is used to denote $T^t_{S,D}$ in the rest of this section.

In Section 3.1, we first investigate the factors influencing the travel time between a source and a destination through statistical analysis. Such statistics on real data reveals valuable facts on travel time and crowds. Based on the statistical results, an active region representing the areas corresponding to the source-destination traffic flow is first estimated (Section 3.2). Two sets of features are then computed inside the active region, and they are used to describe the influences of moving persons (Section 3.3) and stationary persons (Section 3.4) respectively. Our studies show that moving persons and stationary groups influence traffic flows and travel time in quite different ways. Finally, these features are concatenated and a regression method is adopted to map features to travel time values (Section 3.5).

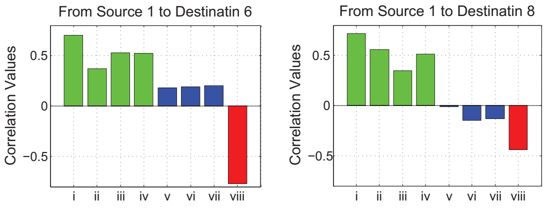
3.1. Statistical study

Our method is motivated by the statistical study on which factors have large influence on the travel time between sources and destinations. The large-scale pedestrian walking path dataset proposed by Yi *et al*. [30] is utilized for the statistical analysis. It provides manually annotated and complete trajectories of all the pedestrians. The scene contains ten source/destination regions as shown in Figure 1.

Table 1 lists eight statistics (i)-(viii) to be investigated. The correlations between travel time and these statistics are shown in Figure 2. One natural finding is that the travel time has strong positive correlation with (i) the travel distance and has strong negative correlation with (iv) the initial walking speed in the source region.

Index	Description
(i)	Travel distance
(ii)	No. of stationary persons along the walking route
(iii)	No. of moving persons along the walking route
(iv)	Total No. of persons along the walking route
(v)	No. of stationary persons in the whole scene
(vi)	No. of moving persons in the whole scene
(vii)	Total No. of persons in the whole scene
(viii)	Pedestrian initial speed in the source region

Table 1. Eight statistics for correlation analysis.



(a) (b)

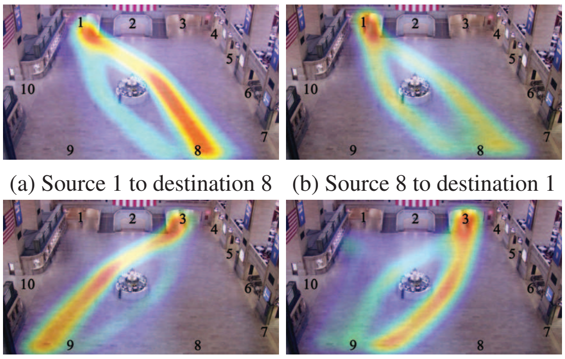
Figure 2. Correlations between travel time and statistics (i)-(viii). (a) For pedestrians from source 1 to destination 6, the correlations are 0.70, 0.37, 0.53, 0.52, 0.18, 0.19, 0.20, and −0.77. (b) For pedestrians from source 1 to destination 8, the correlations are 0.72, 0.56, 0.35, 0.51, −0.00, −0.15, −0.13, and −0.44. Strong positive correlations are shown in green, strong negative correlations are shown in red, and weak correlations are shown in blue.

Compared with the strong positive correlations between the travel time and statistics (ii)-(iv), *i.e*. the numbers of moving and stationary persons along the walking routes, the correlations between travel time and statistics (v)-(vii), *i.e*. the numbers of moving and stationary persons in the whole scene, are weak. It means that the traffic flow between a source and a destination is mainly influenced by activities near the walking routes. The influence of activities far away from the corresponding traffic flow is negligible.

Moreover, moving and stationary pedestrians influence travel time in different ways. When a person $\mathbf{B}$ is approaching the walking route of person $\mathbf{A}$, $\mathbf{A}$ would predict potential collision with $\mathbf{B}$ based on the location, moving direction and speed of $\mathbf{B}$. If there is a potential collision, in most cases, $\mathbf{A}$ and $\mathbf{B}$ would adjust their speed instead of moving directions to avoid collision. Differently, if there is a stationary group on the way of person $\mathbf{A}$, $\mathbf{A}$ is enforced to detour or passes through the group (if the group density is low). Therefore, walking directions are to be considered when modeling the relation between travel time and moving pedestrians, while the influence of stationary groups on travel time is related to the group size and the group density.

3.2. Active region

Since complete walking routes of pedestrians are not observed when estimating travel time, we estimate an *active region* $\mathcal{R}$ which covers the areas occupied by traffic flows of most pedestrians traveling between the source and the



(a) Source 1 to destination 8 (b) Source 8 to destination 1

(c) Source 3 to destination 9 (d) Source 9 to destination 3

Figure 3. Examples of active region maps. Regions shown in warmer color indicate higher weights of $\mathcal{R}_{S,D}$.

destination. We only need to take the activities happening inside $\mathcal{R}$ into account and extract features from these activities to estimate the travel time.

Different source-destination traffic flows have different active regions, which may have overlap. The active regions can be used to discover the underlying relationships between scene locations and source-destination traffic flows. From the abnormal increase of travel time of some source-destination flows, we can locate probable blocking areas in the active region. On the other hand, if certain activities happen at some locations inside the active region, we can predict the probable increase of travel time of some source-destination traffic.

The active region of one source-destination traffic flow should be a weighted map, instead of a binary one. This is because different locations inside the active region should have different importance. For example, activities happening at some important locations on the main roads should have greater influence on the travel time, and higher influence weights should be assigned to these locations.

Let $\mathcal{C}_{S,D}$ be the collection of locations covered by walking routes of all pedestrians coming from the source S and going to the destination D. The influence weight of the active region map at location l is calculated as

$$\mathcal{R}_{S,D}(l) = \frac{1}{\#\mathcal{C}_{S,D}} \sum_{l' \in \mathcal{C}_{S,D}} k(l, l'), \tag{1}$$

where $k(\cdot, \cdot)$ is the Gaussian kernel and the kernel bandwidth is defined as the size of one pedestrian. $\#\mathcal{C}_{S,D}$ counts the number of elements in $\mathcal{C}_{S,D}$ and is used as the normalization term. $\mathcal{C}_{S,D}$ can be obtained by clustering fragmented tracklets between source and destination regions with the dynamic agent-based model proposed in [39].

Several examples of active regions are shown in Figure 3. We observe that the active region of a source-destination pair may contain multiple potential walking routes. Moreover, narrow areas (*e.g.* entrance regions and exit regions) tend to have higher influence weights, which indicates more attention should be paid to these areas in traffic management. Note that the active region is not symmetric *i.e.* $\mathcal{R}_{1,8} \neq \mathcal{R}_{8,1}$, $\mathcal{R}_{3,9} \neq \mathcal{R}_{9,3}$, which means traffic flows in different directions might occupy different scene regions. This is common in transportation systems where mixing of traffic flows in opposite directions is avoided in order to increase traffic efficiency and ensure safety.

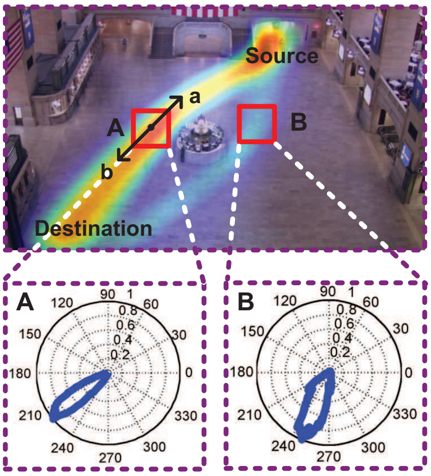


Figure 4. Illustration of moving pedestrian features. An active region map of a source-destination flow is shown on the top. Warmer colors indicate higher active weights. Two locations (**A**, **B**) with different active weights are marked by red rectangles. The walking direction distributions at **A** and **B** are shown in the bottom.

3.3. Features on moving pedestrians

For the traffic flow from source S to destination D at the current time point t, features are extracted from all the moving pedestrians inside the active region. The locations of these moving pedestrians are denoted as $l_i^{mp}, i = 1, ..., N^{mp}$, and $\mathcal{R}_{S,D}(l_i^{mp}) > 0$. For each of these moving pedestrians, two features are computed.

The active region weights at the locations of the N^{mp} moving persons are used as the first feature. For the ith moving pedestrian, the *location* feature is calculated as

$$\mathcal{M}_i^1 = \mathcal{R}_{S,D}(l_i^{mp}). \tag{2}$$

For example, an active region map and two locations (**A** and **B**) are shown in Figure 4. Location **A** is in the main traffic flow so the active weight at **A** is greater than that at **B**. A moving pedestrian appearing at **A** plays a more important role than that appearing at **B**.

The influence to travel time delay is also related to pedestrian's walking direction $\theta_i, i = 1, ..., N^{mp}$. As shown in Figure 4, if a moving pedestrian appears at location **A** and walks along the direction $\mathbf{a}$, *i.e.* being opposite to the source-destination traffic flow, the influence of this pedestrian should be significant. In contrast, if he/she walks along the direction $\mathbf{b}$, *i.e.* similar to the source-destination flow, the influence should be small.

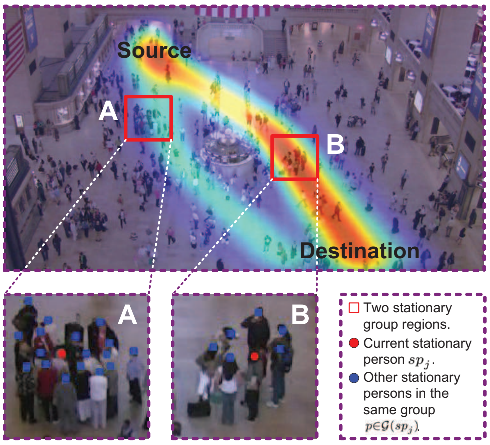

Source
A
B
Destination
A
B
□ Two stationary group regions.
● Current stationary person sp_j.
● Other stationary persons in the same group $p\in\mathcal{G}(sp_j)$

Figure 5. Illustration of stationary pedestrian features. An active region map of a source-destination flow is shown on the top. Warmer colors indicate higher active weights. Two stationary groups (**A**, **B**) with different active weights are marked by the red rectangles. The current stationary person sp_j is marked in red, and the other stationary persons within the same group $\mathcal{G}(sp_j)$ are marked in blue.

For all the locations inside the active region, the distributions of walking directions of the source-destination flow are first computed. Examples of such distributions computed at location **A** and **B** are shown in Figure 4. The second feature is proposed to describe the deviation of the moving pedestrian's walking direction from the speed directional distribution of the source-destination traffic flow. For the ith moving pedestrian, the *influence* feature is calculated as

$$\mathcal{M}_i^2 = \int_{\theta^*} \phi(\theta^*; l_i^{mp})[1 - \cos(\theta_i - \theta^*)], \tag{3}$$

where $\phi(\theta; l_i^{mp})$ is the directional distribution of the traffic flow at the location l_i^{mp}.

3.4. Features on stationary pedestrians

Features are extracted from all the stationary pedestrians $sp_j, j = 1, ..., N^{sp}$ inside the active region, whose locations are denoted by l_j^{sp} with $\mathcal{R}_{S,D}(l_j^{sp}) > 0$. For each stationary pedestrian, three features are computed.

Stationary pedestrians at different locations have different importance, which can be described by the map weights of $\mathcal{R}_{S,D}$. An active region map and two stationary groups are shown in Figure 5. Stationary pedestrians of group **B** block the main source-destination traffic flow, which leads to larger influence on the travel time than the stationary pedestrians of group **A**. For the jth stationary pedestrian, the *location* feature is calculated as

$$\mathcal{S}_j^1 = \mathcal{R}_{S,D}(l_j^{sp}). \tag{4}$$

There are two other *influence* features, $\mathcal{S}_j^2$ and $\mathcal{S}_j^3$. $\mathcal{S}_j^2$ is related to the size of the stationary crowd group. It is more likely for a pedestrian to change the route and detour a longer way when facing a larger stationary group. As shown in Figure 5, stationary pedestrians of group **A** should have larger blocking effect on the traffic flow than those of group **B**. The second stationary pedestrian feature describes the size of a stationary group,

$$\mathcal{S}_j^2 = \#\mathcal{G}(sp_j), \tag{5}$$

where $\mathcal{G}(sp_j)$ is the collection of all the stationary persons that form the same stationary crowd group with sp_j, and $\#$ is the element counting operation.

The influence feature $\mathcal{S}_j^3$ is related to the density of the stationary crowd group. Stationary pedestrians of denser groups should have larger blocking effect. If a stationary group is small or sparse, some aggressive pedestrians may choose to go through it instead of changing their routes. The third stationary pedestrian feature therefore describes the density of a stationary group,

$$\mathcal{S}_j^3 = \frac{1}{\#\mathcal{G}(sp_j)} \sum_{p\in\mathcal{G}(sp_j)} ||l_p - l_j^{sp}||_2^2, \tag{6}$$

where p is a stationary pedestrian from the same stationary group with sp_i, l_p is the location of stationary pedestrian p, and $||l_p - l_{sp_j}||_2^2$ measures the distance between p and sp_j.

3.5. Travel time estimation

Features extracted from moving and stationary pedestrians are concatenated into feature vectors and a regression pipeline is adopted to map these feature vectors to travel times. For a source-destination traffic flow at time point t, moving features $\mathcal{M}_i, i = 1, ..., N^{mp}$ are extracted from N^{mp} moving pedestrians inside the active region, and stationary features $\mathcal{S}_j, j = 1, ..., N^{sp}$ are extracted from N^{sp} stationary pedestrians inside the active region, *i.e.*

$$\begin{aligned} \mathcal{M}_i &= [\mathcal{M}_i^1, \mathcal{M}_i^2, 1]^{tr}, & (7)\\ \mathcal{S}_j &= [\mathcal{S}_j^1, \mathcal{S}_j^2, \mathcal{S}_j^3, 1]^{tr}, & (8) \end{aligned}$$

where tr denotes matrix transpose. In order to keep a fixed number of features at each time point, only N^{m*} moving pedestrians with top $\mathcal{M}_i^1$ values and N^{s*} stationary pedestrians with top $\mathcal{S}_j^1$ values are selected. These selected pedestrians are expected to have the most influences on travel time as they have the largest active region weights, *i.e.* they are at locations that significantly influences travel time. Features of the selected persons are then used for regression. Zeros will be padded if $N^{m*} > N^{mp}$ or $N^{s*} > N^{sp}$.

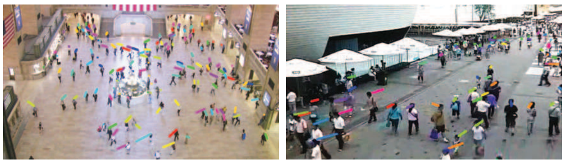
We assume the mapping function from moving features to travel time should be the same for all the moving pedestrians (denoted as f_m). Similarly, the mapping function from stationary features to travel time is denoted as f_s. The travel time can then be estimated as

$$\widehat{T} = \sum_{i\in\mathcal{B}_{mp}} f_m(\mathcal{M}_i) + \sum_{j\in\mathcal{B}_{sp}} f_s(\mathcal{S}_j), \tag{9}$$

where $\mathcal{B}_{mp}$ is the index set of moving pedestrians with top $\mathcal{M}_i^1$ values, and $\mathcal{B}_{sp}$ is the index set of stationary pedestrians with top $\mathcal{S}_j^1$ values.

	Dataset I [30]	Dataset II [33]
Scene type	Indoor square	Outdoor aisle
Resolution (pixel)	$1,920 \times 1,080$	720×576
No. of sources/destinations	10	3
Video duration (s)	$4,000$	900
Frame rate (fps)	25	50
No. of annotated pedestrians	$12,684$	$1,842$
Average travel time (s)	27.1	15.6

Table 2. Details of the datasets used for evaluation.



(a) Dataset I [30] (b) Dataset II [33]

Figure 6. Example frames of the two datasets. Annotated walking routes are shown in different colors.

In our method, the second-order polynomial regression is used, and $f_m(\mathcal{M}_i)$ and $f_s(\mathcal{S}_j)$ can be written as,

$$f_m(\mathcal{M}_i) = \mathcal{M}_i^{tr}\mathbf{W}_m\mathcal{M}_i, \tag{10}$$

$$f_s(\mathcal{S}_j) = \mathcal{S}_j^{tr}\mathbf{W}_s\mathcal{S}_j, \tag{11}$$

where $\mathbf{W}_m$ and $\mathbf{W}_s$ are systemic matrices to be learned.

4. Experiments

4.1. Datasets

Two datasets are used to evaluate the proposed travel time estimation method. One is the pedestrian walking path dataset (Dataset I) proposed by Yi *et al.* [30], which contains 12,684 ground truth walking paths. The other one is a video clip of the Shanghai Expo Dataset (Dataset II) proposed by Zhang *et al.* [33]. Pedestrian walking routes of Dataset II are manually annotated by us because the walking routes are not provided by the original dataset. The details of the two datasets are listed in Table 2 and sample frames of the annotated videos are shown in Figure 6.

4.2. Experimental setup

For each time point t, a temporal window $[t-\tau, t+\tau]$ is adopted. τ is set as 30 seconds for both datasets. All pedestrians from source S to destination D within the temporal window will be considered. The average travel time of these pedestrians $\mu(T_{S,D}^t)$ is used as ground truth to train the regression model. During testing stage, pedestrian travel time $\widehat{T_{S,D}^t}$ would be estimated for each time point based on the extracted features of the current frame. Cross validation is adopted for evaluation. The video frames are randomly divided into ten folds for each source-destination flow.

4.3. Evaluation metrics

The average travel time $\mu(T_{S,D}^t)$ and standard deviation of travel time $\sigma(T_{S,D}^t)$ are used for evaluation. Three evaluation metrics are used to measure the performance of the proposed method.

ET is calculated as the average estimation Error of travel Time. ER_1 and ER_2 are calculated as the average estimation Error Ratios of travel time with ground truth value $\mu(T_{S,D}^t)$ and ground truth variance $\sigma(T_{S,D}^t)$, respectively.

$$\mathrm{ET} = \mathbb{E}_{t,S,D}\left[\left|\widehat{T_{S,D}^t} - \mu(T_{S,D}^t)\right|\right], \tag{12}$$

$$\mathrm{ER}_1 = \mathbb{E}_{t,S,D}\left[\frac{\left|\widehat{T_{S,D}^t} - \mu(T_{S,D}^t)\right|}{\mu(T_{S,D}^t)}\right], \tag{13}$$

$$\mathrm{ER}_2 = \mathbb{E}_{t,S,D}\left[\frac{\left|\widehat{T_{S,D}^t} - \mu(T_{S,D}^t)\right|}{\sigma(T_{S,D}^t)}\right], \tag{14}$$

where $\widehat{T_{S,D}^t}$ is the estimation result, $\mu(T_{S,D}^t)$ is the ground truth travel time, $\sigma(T_{S,D}^t)$ is the standard deviation of the travel time of the observed pedestrians within the temporal window, and $\mathbb{E}_{t,S,D}[\cdot]$ denotes the average among all the time points and all the source-destination pairs.

4.4. The compared methods

In Section 3.1, we observe that statistic (ii) the number of stationary persons along the walking route, and statistic (iii) the number of moving persons along the walking route have strong correlations with the pedestrian travel time. A baseline feature set containing only these two statistic numbers is designed for comparison.

In order to evaluate the effectiveness of each of the proposed features, several feature subsets are used for comparison, *i.e.* (a) only using the *location* features $\mathcal{M}_1$, $\mathcal{S}_1$; (b) only using the *influence* features $\mathcal{M}_2$, $\mathcal{S}_2$, $\mathcal{S}_3$; (c) only using the *moving* person features $\mathcal{M}_1$, $\mathcal{M}_2$; and (d) only using the *stationary* person features $\mathcal{S}_1$, $\mathcal{S}_2$, $\mathcal{S}_3$. The estimation results of these comparisons on Dataset I and Dataset II are listed in Table 3.

The problem of pedestrian travel time estimation is introduced in this paper for the first time and there is no existing work specially designed for or can be directly applied to this problem. We select some existing computer vision techniques, and evaluate them as baselines for the problem. (a) The social force model [10] can simulate pedestrian behaviors and the simulated travel time can be estimated as pedestrian travel time.[2] (b) Based on person re-identification [12], the time interval of matched pedestrians in two frames can be computed as travel time. (c) The pedestrian travel time can also be calculated from the starting and ending points of trajectories extracted from pedestrian tracking [26]. We

[2] Parameters to be set in SFM: Reaction time is manually set as 0.8s, which is the same as the time interval of our annotations. The other model parameters are adaptively specified by adjusting simulation results as close as real observations. 1% of the annotated ground truths are randomly selected as the training samples.

Feature	Description	Results on Dataset I			Results on Dataset II		
		ET	ER_1	ER_2	ET	ER_1	ER_2
Statistics (ii), (iii)	Baseline features	6.58s	22.92%	282.52%	2.59s	17.29%	125.11%
$\mathcal{M}_1, \mathcal{S}_1$	(a) *Location* features	2.69s	9.39%	98.97%	1.52s	10.21%	77.01%
$\mathcal{M}_2, \mathcal{S}_2, \mathcal{S}_3$	(b) *Influence* features	2.76s	9.70%	105.03%	1.49s	10.22%	77.72%
$\mathcal{M}_1, \mathcal{M}_2$	(c) *Moving* features	2.76s	9.32%	101.61%	1.49s	10.02%	75.07%
$\mathcal{S}_1, \mathcal{S}_2, \mathcal{S}_3$	(d) *Stationary* features	2.75s	9.75%	105.95%	1.53s	11.22%	83.47%
$\mathcal{M}_1, \mathcal{M}_2, \mathcal{S}_1, \mathcal{S}_2, \mathcal{S}_3$	Proposed features	**2.32**s	**8.00**%	**88.45**%	**1.40**s	**9.56**%	**73.45**%

Table 3. Travel time estimation results on Dataset I and Dataset II by using different sets of features.

Method	Results on Dataset I			Results on Dataset II		
	ET	ER_1	ER_2	ET	ER_1	ER_2
(a) Pedestrian simulation [10]	7.14s	21.60%	232.68%	4.13s	32.03%	206.31%
(b) Person re-identification [12]	11.60s	37.29%	388.87%	4.65s	37.82%	252.55%
(c) Pedestrian tracking [26]	9.45s	29.27%	332.28%	4.41s	35.17%	253.05%
(d) Motion pattern features [2]	7.33s	19.58%	123.41%	2.50s	28.32%	118.23%
(e) Stability features [24]	7.73s	21.45%	126.47%	2.62s	33.38%	158.09%
Proposed method	**2.32**s	**8.00**%	**88.45**%	**1.40**s	**9.56**%	**73.45**%

Table 4. Travel time estimation results on Dataset I and Dataset II by the proposed method and the comparisons.

also try some crowd related features, *i.e.* (d) motion pattern features [2] and (e) stability features [24]. The features are used to regress travel time. The estimation results of these comparisons on both Datasets are shown in Table 4.

4.5. Experimental results

From the experimental results shown in Tables 3, we observe that the proposed feature set achieves much better performance than the baseline features. This is because the baseline features simply count the number of persons along the walking routes, but the different roles and influences of these persons are not modeled. The estimation errors increase when using the subsets of the proposed features, . It demonstrates the effectiveness of each of the proposed features, including the *location* features, *influence* features, *moving* features, and *stationary* features.

From Table 4, we can see better performance can be achieved by the proposed method compared with (a) - (e). It is partially because these methods are not specifically designed for the travel time estimation task. Moreover, for (a) and (c), they mainly focus on the pedestrians' reactions to local environments. However, global modeling of the whole scene is needed when solving the travel time estimation problem. For (b), it is quite challenging for person re-identification algorithms when pedestrians are occluded by each other frequently, which is common in crowded scenes. Moreover, as the appearance of many pedestrians are quite similar, person re-identification fails frequently.

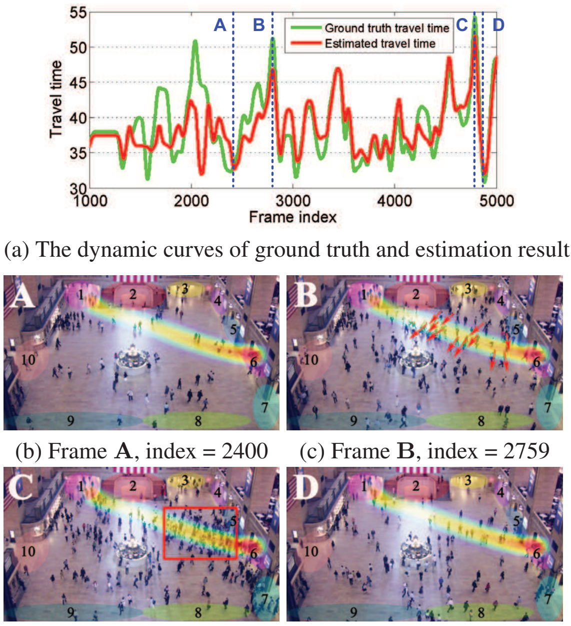
Directly using existing motion features, the ones in (d) and (e), cannot provide good results for a few reasons. Firstly, without modeling active regions between source-sink pairs, the influences of different locations on travel time are not modeled. However, we showed that crowd densities and velocities outside the active region have much less influence on the travel time. Secondly, the conflict between instantaneous velocities of individuals and main flow patterns at different locations are not modeled ,while our explicit modeling makes estimation easier. Lastly, the information of stationary crowds is not used. However, stationary crowds' interaction with moving pedestrians, is quite different from the interaction among moving pedestrians. Such interaction is affected by the size and density of stationary groups, which was never considered. Overall, our features are specially designed for travel time estimation, well motivated by statistical analysis, and more effective than generic crowd features.

5. Applications

As introduced in Section 1, the estimated travel time can provide rich information for video surveillance, and various applications can be implemented based on the proposed travel time estimation pipeline, including crowd scene understanding and pedestrian behavior analysis.

5.1. Scene dynamic monitoring

The estimated travel time is an important indicator of scene status and can be used to monitor whether the scene is unobstructed or not. For pedestrians from source 1 to destination 6, the dynamic curves of the estimated pedestrian travel time $\widehat{T^t_{S,D}}$, together with the ground truth travel time $\mu(T^t_{S,D})$, are plotted in Figure 7(a). Four example frames are shown in Figure 7(b)-(e). At frame $\mathbf{A}$, the scene is unobstructed and the estimated travel time is short. However, at frame $\mathbf{B}$, the source-destination flow is intersected by multiple moving pedestrians (red arrows), which leads to the increase of travel time. At frame $\mathbf{C}$, region 6 is blocked by the large dense stationary crowds (red rectangle), thus the



(a) The dynamic curves of ground truth and estimation result.

(b) Frame $\mathbf{A}$, index = 2400 (c) Frame $\mathbf{B}$, index = 2759

(d) Frame $\mathbf{C}$, index = 4770 (e) Frame $\mathbf{D}$, index = 4849

Figure 7. (a) The dynamic updates of ground truth travel time (green curve) and the estimated travel time (red curve) for pedestrians from source 1 to destination 6. (b)-(e) Four example frames marked in (a) as $\mathbf{A}$-$\mathbf{D}$.

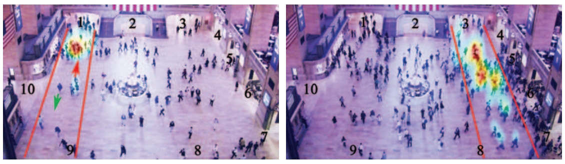
estimated travel time increases significantly. When the stationary crowds disperse at frame $\mathbf{D}$, the estimation of travel time returns to normal value.

5.2. Blocking region localization

From Eq. (9), we can estimate the travel time by summing all the influences of moving and stationary pedestrians inside the active region together. On the other way around, we can also factorize the total travel time into different influence factors, and the time delay caused by each individual inside the active region can be inferred. In this way, the regions that blocking traffics can be located by identifying the moving/stationary persons that contribute most to the estimation result of the travel time. Two examples are shown in Figure 8.

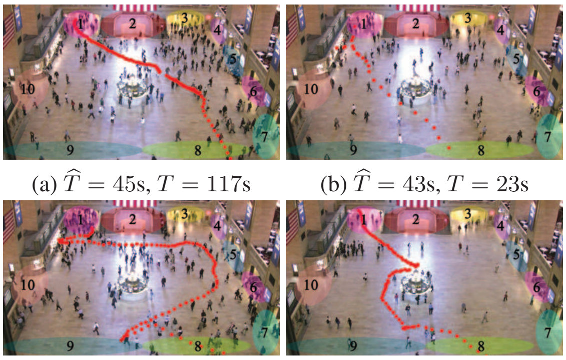
5.3. Abnormal behavior detection

If we focus on each individual, we can determine whether the pedestrian is walking in a normal way, *i.e*. the travel time is close to the estimation result. Four pedestrians with abnormal travel times are shown in Figure 9. Pedestrian (a) is walking too slowly (dense dots), so that the actual travel time is significantly greater than the estimation. Pedestrian (b) is running fast (sparse dots), which leads to a much shorter travel time. Pedestrians (c) and (d) are walking along tortuous routes, so the travel times are greater than expected.



(a) Source 1 to destination 9 (b) Source 3 to destination 8

Figure 8. Examples of blocking region localization. Different individuals' contribution to the travel time delay are represented by colors. Blocking regions can be located as the regions with warm colors. Active region boundaries are roughly outlined by red lines. In (a), we can also observe that different walking directions (as denoted by the arrows) have different influences on the travel time delay. The red and green arrows point against and along the traffic flow respectively and therefore have different influences.



(a) $\widehat{T} = 45\text{s}, T = 117\text{s}$ (b) $\widehat{T} = 43\text{s}, T = 23\text{s}$

(c) $\widehat{T} = 47\text{s}, T = 94\text{s}$ (d) $\widehat{T} = 40\text{s}, T = 83\text{s}$

Figure 9. Examples of four pedestrians with abnormal travel times. The estimated time $\widehat{T}$ and actual travel time T are shown. One red dot is plotted at the pedestrian location for each second.

6. Conclusion

The problem of pedestrian travel time estimation is introduced for the first time in this paper. A novel travel time estimation method is proposed for the challenging task, which consists of active region generation, moving/stationary feature design, and regression. The estimated travel time provides rich information for crowd scene understanding and pedestrian behavior analysis. Three applications based on the proposed technique are introduced in this paper and more interesting applications about travel time information are yet to be discovered.

Acknowledgment

This work is partially supported by the General Research Fund sponsored by the Research Grants Council of Hong Kong (Project Nos. CUHK 419412 and CUHK 147011), Hong Kong Innovation and Technology Support Programme (Project ITS/221/13FP), Shenzhen Basic Research Program (JCYJ20130402113127496), NSFC (No. 61301269), and Sichuan High Tech R&D Program (No. 2014GZX0009).

References

[1] G. Antonini, M. Bierlaire, and M. Weber. Discrete choice models of pedestrian walking behavior. *Transportation Research Part B: Methodological*, 40(8):667–687, 2006. 1

[2] A. Basharat, A. Gritai, and M. Shah. Learning object motion patterns for anomaly detection and improved object detection. In *Proc. CVPR*, 2008. 2, 7

[3] B. Benfold and I. Reid. Stable multi-target tracking in real-time surveillance video. In *Proc. CVPR*, 2011. 1

[4] E. Bonabeau. Agent-based modeling: Methods and techniques for simulating human systems. *PNAS*, 99(Suppl 3):7280–7287, 2002. 2

[5] A. B. Chan and N. Vasconcelos. Counting people with low-level features and bayesian regression. *IEEE Trans. Image Processing*, 21(4):2160–2177, 2012. 2

[6] K. Chen, S. Gong, T. Xiang, and C. C. Loy. Cumulative attribute space for age and crowd density estimation. In *Proc. CVPR*, 2013. 2

[7] D. Forsyth. *Group dynamics*. Cengage Learning, 2009. 2

[8] W. Ge and R. T. Collins. Marked point processes for crowd counting. In *Proc. CVPR*, 2009. 2

[9] D. Helbing, I. Farkas, and T. Vicsek. Simulating dynamical features of escape panic. *Nature*, 407(6803):487–490, 2000. 2

[10] D. Helbing and P. Molnar. Social force model for pedestrian dynamics. *Physical review E*, 51(5):4282, 1995. 1, 2, 6, 7

[11] G. Le Bon. *The crowd: A study of the popular mind*. Macmillian, 1897. 2

[12] W. Li, R. Zhao, T. Xiao, and X. Wang. Deepreid: Deep filter pairing neural network for person re-identification. In *Proc. CVPR*, 2014. 3, 6, 7

[13] W. Liu, A. B. Chan, R. W. Lau, and D. Manocha. Leveraging long-term predictions and online learning in agent-based multiple person tracking. *IEEE Trans. CSVT*, 25(3):399–410, 2015. 2

[14] Z. Ma and A. B. Chan. Crossing the line: Crowd counting by integer programming with local features. In *Proc. CVPR*, 2013. 2

[15] D. Makris and T. Ellis. Learning semantic scene models from observing activity in visual surveillance. *IEEE Trans. Systems, Man, and Cybernetics, Part B: Cybernetics*, 35(3):397–408, 2005. 2

[16] R. Mehran, A. Oyama, and M. Shah. Abnormal crowd behavior detection using social force model. In *Proc. CVPR*, 2009. 2

[17] M. Moussaïd, D. Helbing, and G. Theraulaz. How simple rules determine pedestrian behavior and crowd disasters. *PNAS*, 108(17):6884–6888, 2011. 1

[18] 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*, 2009. 2

[19] I. Saleemi, L. Hartung, and M. Shah. Scene understanding by statistical modeling of motion patterns. In *Proc. CVPR*, 2010. 1

[20] I. Saleemi, K. Shafique, and M. Shah. Probabilistic modeling of scene dynamics for applications in visual surveillance. *IEEE Trans. PAMI*, 31(8):1472–1485, 2009. 1, 2

[21] P. Scovanner and M. F. Tappen. Learning pedestrian dynamics from the real world. In *Proc. ICCV*, 2009. 2

[22] J. Shao, K. Kang, C. C. Loy, and X. Wang. Deeply learned attributes for crowded scene understanding. In *Proc. CVPR*, 2015. 1

[23] J. Shao, C. C. Loy, and X. Wang. Scene independent group profiling in crowd. In *Proc. CVPR*, 2014. 2

[24] B. Solmaz, B. E. Moore, and M. Shah. Identifying behaviors in crowd scenes using stability analysis for dynamical systems. *IEEE Trans. PAMI*, 34(10):2064–2070, 2012. 2, 7

[25] L. Song, F. Jiang, Z. Shi, R. Molina, and A. K. Katsaggelos. Toward dynamic scene understanding by hierarchical motion pattern mining. *IEEE Trans. Intelligent Transportation Systems*, 15(3):1273, 2014. 2

[26] C. Tomasi and T. Kanade. *Detection and tracking of point features*. School of Computer Science, Carnegie Mellon Univ. Pittsburgh, 1991. 2, 3, 6, 7

[27] X. Wang, X. Ma, and W. E. L. Grimson. Unsupervised activity perception in crowded and complicated scenes using hierarchical bayesian models. *IEEE Trans. PAMI*, 31(3):539–555, 2009. 2

[28] J. Yang, J. Li, and Y. He. Crowd density and counting estimation based on image textural feature. *Journal of Multimedia*, 9(10):1152–1159, 2014. 2

[29] Y. Yang, J. Liu, and M. Shah. Video scene understanding using multi-scale analysis. In *Proc. ICCV*, 2009. 1, 2

[30] S. Yi, H. Li, and X. Wang. Understanding pedestrian behaviors from stationary crowd groups. In *Proc. CVPR*, 2015. 1, 3, 6

[31] S. Yi and X. Wang. Profiling stationary crowd groups. In *Proc. ICME*, 2014. 2

[32] S. Yi, X. Wang, C. Lu, and J. Jia. L0 regularized stationary time estimation for crowd group analysis. In *Proc. CVPR*, 2014. 2

[33] C. Zhang, H. Li, X. Wang, and X. Yang. Cross-scene crowd counting via deep convolutional neural networks. In *Proc. CVPR*, 2015. 2, 6

[34] T. Zhang, H. Lu, and S. Z. Li. Learning semantic scene models by object classification and trajectory clustering. In *Proc. CVPR*, 2009. 2

[35] X. Zhao and G. Medioni. Robust unsupervised motion pattern inference from video and applications. In *Proc. ICCV*, 2011. 2

[36] L. Zheng, S. Wang, L. Tian, F. He, Z. Liu, and Q. Tian. Query-adaptive late fusion for image search and person re-identification. In *Proc. CVPR*, 2015. 3

[37] B. Zhou, X. Tang, and X. Wang. Learning collective crowd behaviors with dynamic pedestrian-agents. *IJCV*, 111(1):50–68, 2015. 2

[38] B. Zhou, X. Wang, and X. Tang. Random field topic model for semantic region analysis in crowded scenes from tracklets. In *Proc. CVPR*, 2011. 2

[39] B. Zhou, X. Wang, and X. Tang. Understanding collective crowd behaviors: Learning a mixture model of dynamic pedestrian-agents. In *Proc. CVPR*, 2012. 4