Physica A 650 (2024) 129994

Contents lists available at ScienceDirect

Physica A

journal homepage: www.elsevier.com/locate/physa

Behavior recognition of non-motorized transport at intersections using dual-channel grid model based on disordered trajectory point data

Huanting Xu [a,c], Zhaocheng He [a,b,c], Yiyang Chen [a,c], Zhigang Wu [a,c], Yiting Zhu [a,b,c,*]

[a] School of Intelligent Systems Engineering, Sun Yat-sen University, Shenzhen, 518107, China
[b] Pengcheng Laboratory, Shenzhen, 518107, China
[c] Guangdong Provincial Key Laboratory of Intelligent Transportation System, Shenzhen, 518107, China

ARTICLE INFO

Keywords:
Non-motorized transport
Behavior recognition
Grid model
Disordered trajectory point

ABSTRACT

Accurately identifying the passing and waiting behavior of pedestrians and non-motorized vehicles at intersection is essential to planning management measures for non-motorized transport. Much previous study in this field has focused on methods based on continuous individual trajectory detection, which are almost ineffective under the poor detection condition in dense traffic scenarios. To address the task, this paper establish a workable framework for inferring the probabilities of passing or waiting behavior of pedestrians and non-motorized vehicles in arbitrary space using disordered trajectory point data. First, a two-channel model is proposed to perform a formalized grid-level representation of an intersection, in which the functional attributes and occupancy characteristics of grids are comprehensively defined and quantified. Then, the quantitative characteristics of the grids are used to detect the real-world occurrence space of passing and waiting behaviors, by the clustering and expanding operations on grids. Finally, through feature transfer along path, characteristics is decomposed into passing and waiting occurrence characteristics for behavior probability computation. Results indicate that the method achieves over 91% accuracy of behavior recognition, which is better than compared methods in various Multiple Object Tracking Accuracy (MOTA). Although the method is sensitive to spatial detection conditions, it obtains steady accuracy under various target detection settings.

1. Introduction

Statistics show that approximately 1.3 million people die in road traffic accidents each year in the world. Half of these deaths are among vulnerable road users such as pedestrians, cyclists and motorcyclists [1,2], which are called non-motorized transport. The abnormal behaviors of non-motorized transport, such as waiting outside safety islands and passing at red lights, are one of the main causes of accidents [3–5]. While there are numerous strategies to lessen the frequency of abnormal behaviors, they are all dependent on behavior recognition [6,7]. Therefore, accurately identifying the crossing behaviors (passing or waiting) of pedestrians and non-motorized vehicles at intersections is a necessary prerequisite for improving traffic safety at intersections.

Previous research on behavior recognition of pedestrians has relied on individual-based continuous tracking methods. These methods typically use video as the primary input to understand and classify spatio-temporal information of individuals. Depending

* Corresponding author at: School of Intelligent Systems Engineering, Sun Yat-sen University, Shenzhen, 518107, China.
E-mail address: zhuyt25@mail.sysu.edu.cn (Y. Zhu).

https://doi.org/10.1016/j.physa.2024.129994
Received 15 May 2024; Received in revised form 15 July 2024
Available online 25 July 2024

on the technical approach, they can be categorized as either feature-based or neural network-based methods. Feature-based approaches extract individual features from the video and achieve behavior recognition through a process of feature fusion and classification [8,9]. Neural network approaches construct an end-to-end route from the video to the behavior classification results [10–16]. These methods can produce satisfactory results in situations with light traffic and low pedestrian activity. However, their effectiveness depends on the accurate acquisition of time series data. In intersections with high traffic volume and dense population, tracking individual targets becomes challenging due to the high degree of similarity and occlusion between individuals. As a result, individual continuous trajectory would deteriorate into disordered trajectory point (DTP) data when MOTA is low, in which the above methods may not be effective.

In cases where continuous tracking of individuals is challenging, space-based methods offer an effective alternative way to recognize stopping and passing behaviors. Some methods delineate spatial behavior by leveraging information from spaces, and reconstruct overall spatial behavior by integrating spatial topology. In this approach, the grid model is typically employed to analyze their behavior. The earliest grid motion analysis model is the occupancy grid model for robot motion detection [17], which observes and calculates the grid occupancy at different times to determine the robot motion. Subsequently, space-based methods are mostly used for indoor map modeling and location tracking. The method employs a grid-based model to enhance the localization of the wifi-based system within indoor spaces [18]. This allows for the iterative determination of the traveler's trajectory by analyzing the data at each position. However, these studies mostly focus on individual motion recognition and analysis, and ignore address the analysis of multi-individual movement processes. Therefore, it is necessary to develop a method based on spatial group information to identify waiting and passing behavior of non-motorized transport, in order to address the challenges posed by inadequate target tracking in busy crossings.

Due to financial and technical limitations, the data collected from crowded intersections, typically consists of information about pedestrians and non-motorized vehicles, known as DTP data. DTP data comprises the point ID, time, latitude, and longitude, which give only the current spatial location of the object but not its trajectory. The absence of tracking information poses a significant challenge in recognizing multi-target behavior, as this task typically relies on information about the individual's state both before and after a given time, rather than just at that moment.

Confronting the challenge, we establish a behavior recognition method using a Dual-Channel Grid Model (DCGM) based on DTP data. This method utilizes feature analysis in grid-level to identify and calculates the probability of various behaviors and the occurrence of the behavior in each space. The primary contributions of this study are:

- We develop a DCGM framework to infer passing or waiting probabilities of pedestrians and non-vehicles in any given space. Using easily gathered unordered trajectory-point data, our approach can handle tasks at crossings with high pedestrian and non-vehicle, where approaches based on individual-based continuous tracking data are almost useless.
- Since many pedestrians and non-vehicles do not really use designated passing or waiting spaces, we employ characteristics of grids to recognize the real-world occurrence space of passing or waiting behaviors. Experiments show that our method yields 94% recognition accuracy.
- We conducted a series of experiments to validate the effectiveness of the current method. The experiments clearly illustrate the superior outcomes of our method across various detection conditions, showcasing its robust operational efficiency and stability.

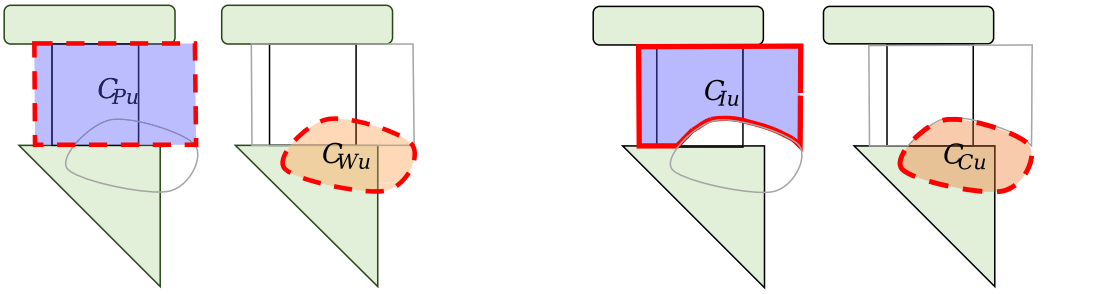
The remainder of this paper is organized as follows: Section 2 summarizes the related literature. Section 3 presents the framework and detail of proposed approach. Section 4 reports the experimental results. Finally, we conclude the paper in Section 5 and discuss the future research directions.

2. Related work

2.1. Individual behavioral recognition based on continuous tracking

Multi-individual behavior recognition methods based on continuous tracking are the current mainstream methods. There are many sources of individual time-series information, such as Red–Green–Blue (RGB), skeleton, depth, infrared sequences, point clouds, event streams, audio, acceleration signals, radar, and WiFi [19], among which video is used by most of the research institutes due to its low cost, wide deployment, easy acquisition, and rich data. Video-based behavior recognition methods can be subdivided into feature engineering-based and deep learning-based methods based on their methodology. Feature engineering needs to extract individual features from video, and then go through feature fusion and feature classification to achieve behavior recognition, and the Dense Trajectories (DT) algorithm proposed by Wang is a representative method of this type. DT is to use the optical flow field to obtain the trajectory in the video sequence, and then extract the trajectory shape features and Histogram of Oriented Gradients (HOF), Histogram of Oriented Gradients (HOG), and Modified Histogram of Back-projected Descriptors (MHB) features along the trajectory, and then features are encoded using the Bag of Features (BOF) method, and finally the Support Vector Machine (SVM) classifier [8] is trained based on the encoding results. Then an improved method based DT called Improved Dense Trajectories (iDT) appeared, and became the best traditional algorithm in behavior detection. iDT improved motion recognition accuracy by estimating camera motion to eliminate optical flow and trajectories on the background, which help it to get the good results and robustness, achieving 91.2% accuracy on the UCF50 dataset and 57.2% on HMDB51 [20].

Deep learning, on the other hand, uses video directly as input to achieve end-to-end behavior recognition, which can be classified into four main types. Donahue et al. [15] proposed a long-time recurrent convolutional network (LRCN) that incorporates long and



(a) the space mainly used for passing or waiting behavior (b) the space used for independent or combined behavior

Fig. 1. Illustrations of some notations.

Table 1
Notations and definitions.

Notation	Definition
c_i	Grid, the basic units that make up the intersection space.
$G(T, c_i)$	Occupancy characteristics, the occupancy duration of the grid c_i in the time interval T.
$G_P(T, c_i)$	Passing Occupancy characteristics, the occupancy duration by the passing behavior in the grid c_i in the time interval T.
$G_W(T, c_i)$	Waiting Occupancy characteristics, the occupancy duration by the waiting behavior in the grid c_i in the time interval T.
C_{Pd}	Designed space for passing behavior. A space originally designed for the passing behavior, such as a crosswalk.
C_{Wd}	Designed space for waiting behavior. A space originally designed for the waiting behavior, such as a island.
C_{Pu}	Actual occurrence space of passing behavior. A space mainly used for the passing behavior in reality, as seen in Fig. 1(a).
C_{Wu}	Actual occurrence space of waiting behavior. A space mainly used for the waiting behavior in reality, as seen in Fig. 1(a)
C_{Iu}	Independent occurrence space of passing behavior. A space exclusively used for the passing behavior, calculating by $C_{Pu} - C_{Wu}$, as seen in Fig. 1(b).
C_{Cu}	Combined occurrence space of behaviors. A space used for both waiting and passing behavior, equal to C_{Wu} in space, as seen in Fig. 1(b).
$P_P(T, c_i)$	The passing probability of pedestrians and non-vehicles in the grid c_i in the time interval T.
$P_W(T, c_i)$	The waiting probability of pedestrians and non-vehicles in the grid c_i in the time interval T.

short-term memory networks to capture motion information. The network considers the persistent episodic changes exhibited by previous frames in an impact simulation to enhance behavior recognition accuracy. Following the initial use of 3D convolutional grids for behavior recognition by Ji and Xu [12], Tran et al. [21] introduced C3D, a comprehensive framework for 3D convolution-based behavior recognition networks, which has significantly advanced the method. Wang et al. [11] introduced a Two-Stream Network (TSN) that effectively classified behaviors by sampling video frames using both RGB stream branching and optical stream branching. The TSN addresses the limitation of not considering the behavioral information of the entire video segment by fusing the classification scores from both branches. This year, the increased popularity of graph convolutional networks has resulted in the rapid development of several methods for behavior recognition that are based on graph convolutions. One notable example is the dual-stream graph Convolutional Neural Network (CNN) model [14].

2.2. Behavior analysis based on grid

Contrary to tracking individuals over time, spatial-based behavior recognition methods gather information from groups of people at various spatial locations. This information is then used to analyze behavior patterns based on spatial topology. The grid model is a classical and widely used model that can represent different scene spaces and accurately describe and recognize the movement of pedestrians [18]. Grid models were first used for robot motion detection by constructing a behavioral probabilistic lattice of the indoor space and integrating the possibility of integrating sensor readings over time through Bayesian networks to obtain the probability and direction of motion of the robot at each position [17,22]. In the field of pedestrian behavior analysis, the grid model was mainly applied to the simulation study of the motion of traffic objects, and the simulation of pedestrian motion is carried out by combining the pedestrian motion characteristics through metacellular automata [23–26]. Grid model-based pedestrian position estimation is another application. Some scholars used sensors to obtain the pedestrian's position in indoor grid space and estimate the direction of the pedestrian's movement based on the Alman filter to recognize the pedestrian's movement in the indoor area [27].

Grid models rely solely on spatial features to estimate the behavior of objects, thus requiring less temporal data compared to the trajectory analysis approach. Simultaneously, the data from multiple objects is consolidated into a grid for analysis, resulting in a reduction in the impact of data noise and outliers, and enhancing its robustness. Furthermore, the grid model exhibits superior performance in terms of processing speed, making it well-suited for tasks involving behavior recognition in scenarios where numerous objects need to be recognized.

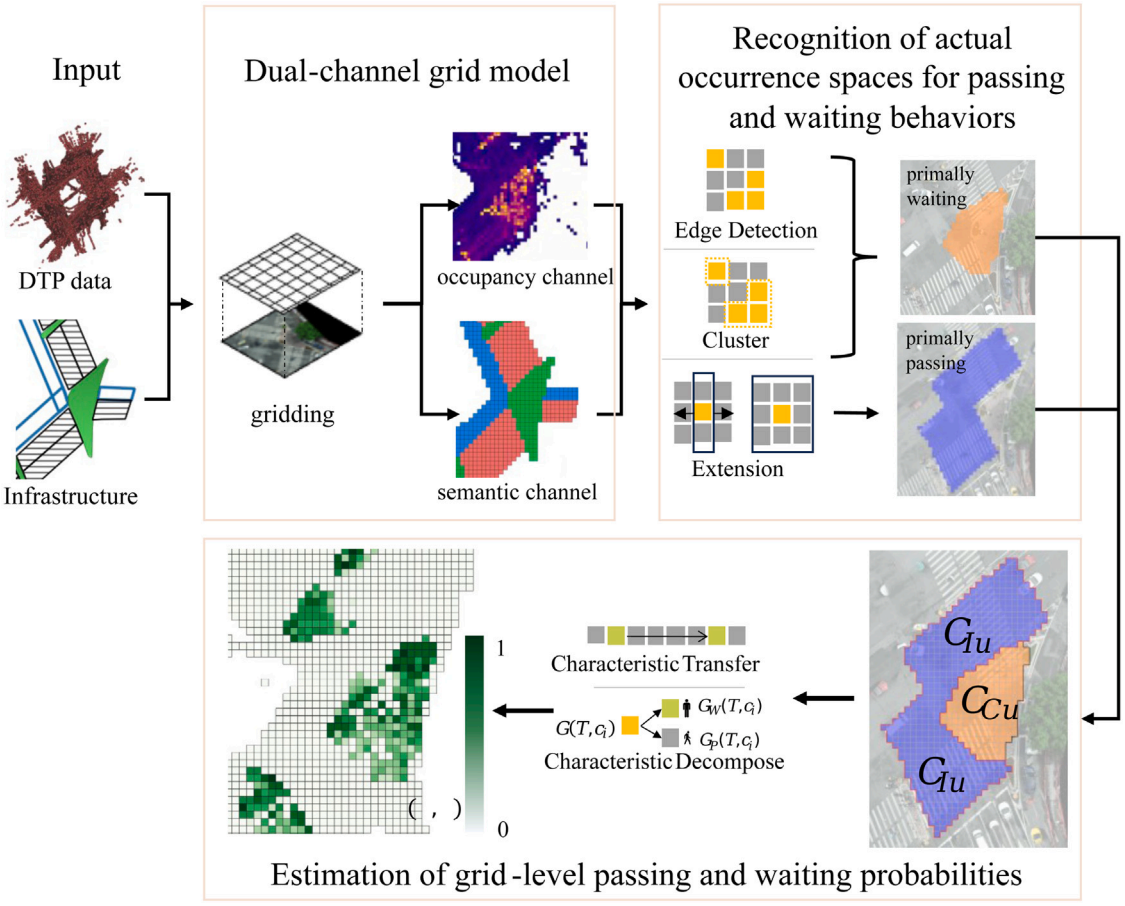


Fig. 2. Process of dual-channel grid model to infer probabilities of passing or waiting behaviors of pedestrians and non-vehicles in any space.

3. Method

3.1. Notations and definitions

Table 1 provides basic definitions and explanations of the symbols that appear in our paper.

3.2. Framework

We have created a framework that identifies the actual occurrence space where behaviors occur and their probability in each grid. These two components are essential for the task of behavior recognition. The structure of our approach is depicted in Fig. 2.

Firstly, by building a two-channel grid model, we discretize the intersection into a space made up of grids, which is then utilized for storing the occupancy and semantic characterizes. Then, based on the semantic characteristics, the actual pedestrian non-motorized occupancy characteristics are used to cluster and extend the space to recognize the main occurrence space of the two types of behaviors. Finally, the behavioral probability of the grid is calculated by decomposing the occupancy characteristics through the transfer of passing behavior characteristics.

3.3. Dual-channel grid model based on DTP data

3.3.1. Model description

A dual-channel grid model is designed to describe signal-controlled intersections that contain geometric, topological, semantic, and temporal occupancy information about the intersection space. The dual-channel grid model is expressed as $C = \{c_i\}$, where $c_i = (X, E, G)$.

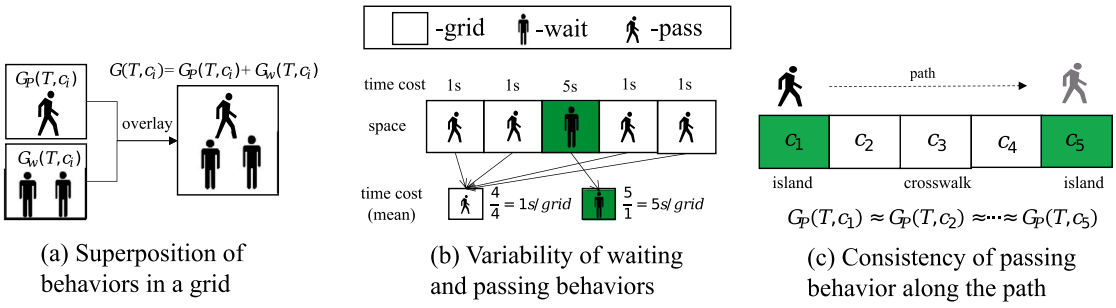


Fig. 3. Grid-level properties of street-crossing behaviors.

$X=(x_1,x_2)$ denotes the row and column numbers of the grid, describing the relative position of the grid in the scene. Through X it is possible to know the spatial location of the mesh and its spatial topological relationship with other meshes. E denotes the semantic channel of the mesh, describing the semantic information represented by the mesh space in the scene, which depends on the type of facility the space is in. We focus on the slow crossing facilities, such as safety islands, pedestrian crossings, non-motorized crossing lanes, which are semantically represented by the grid as $E(c_i)$. G denotes the occupancy channel of the grid, describing the occupancy characterizes of non-motorized transport in the grid space, which is expressed using the proportion of time occupancy. As in Eq. (1), we compute the time occupancy ratio for each grid by dividing the number of occupancy frames to the total number of frames, where x_t denotes if frame t contains pedestrians or non-motorized vehicles in time interval T, 1 indicates yes, 0 indicates no; n_T denotes the total number of frames in time interval T.

$$G(T,c_i)=\frac{\sum x_t}{n_T} \tag{1}$$

3.3.2. Grid-level properties of street-crossing behaviors

Superposition of behaviors in a grid. Over a period of time, the grid can be used for waiting and passing sequentially, and this superposition of behaviors leads to a decomposition of the grid occupancy characteristics into two parts: occupancy characteristics due to passing behaviors and occupancy characteristics due to waiting behaviors (Fig. 3a).

Variability of waiting and passing behaviors. The waiting and passing behavior of non-motorized transport exhibit distinct spatio-temporal characteristics. On one hand, the duration of the green light for a specific direction at an intersection is usually shorter than the duration of the red light. Consequently, the time spent waiting (t_P) is longer than the time spent passing (t_W), which can be expressed as $t_P \geq t_W$. On the other hand, occupancy during waiting behavior is confined to a limited number of grid cells, whereas passing behavior covers a larger array of grids. This difference is captured by the relationship between the set of grids for passing (C_P) and those for waiting (C_W), where $|C_P| > |C_W|$.

Essentially, waiting behavior concentrates time usage within a restricted spatial area, while passing behavior distributes it across multiple pathways (Fig. 3b). This difference in spatial and temporal allocation results in varying occupancy characteristics between the two behaviors, as demonstrated by the ratio $\frac{t_P}{|C_P|} < \frac{t_W}{|C_W|}$. Thus, the grid occupancy impact of waiting behaviors is more pronounced than that of passing behaviors.

Consistency of passing behavior along the path. It is assumed that pedestrians and non-motorized vehicles maintain a steady speed and direction when crossing a roadway. Under the assumption, the behavior of passing through multiple grids in a forward direction can be estimated. As in Fig. 3c, this behavior manifests as similar occupancy characteristics across these grids. In this context, w^T represents the slope coefficient of the direction of passage, and k denotes the linear intercepts. The mathematical model describing this behavior is expressed as:

$$G_P(T,c_i)\approx G_P(T,c_j), \quad \text{if} \quad w^T c_i \cdot X = k \quad \text{and} \quad w^T c_j \cdot X = k \tag{2}$$

3.4. Recognition of the actual occurrence spaces

3.4.1. Actual occurrence space of waiting behaviors

The recognition algorithm is illustrated in Fig. 4. It consists of three main steps: edge detection, clustering, and convex hull construction. During the edge identification process, the grid occupancy characteristics are transformed into pixel values. The Canny operator is then employed to detect boundaries of space where the waiting behavior primarily occurs. Clustering is performed using the classical k-means algorithm, modified to incorporate the center of the designated residency space as the cluster center. This adjustment guarantees that the clustering results precisely mirror the true distribution of spaces. Convex hulls are used to enclose the scattered grids, identifying the actual occurrence spaces for waiting behavior. These spaces are indicated as $C_{Wui}, i=1,2.....k$, where k is the total number of identified spaces.

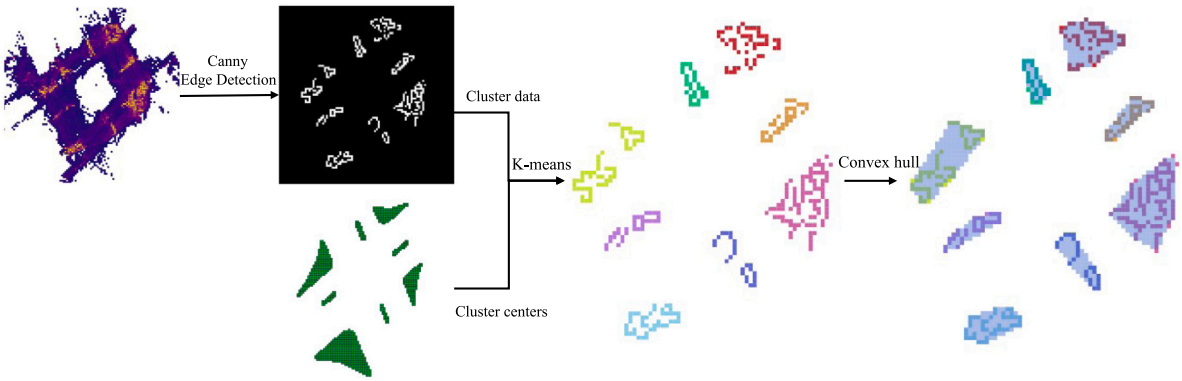
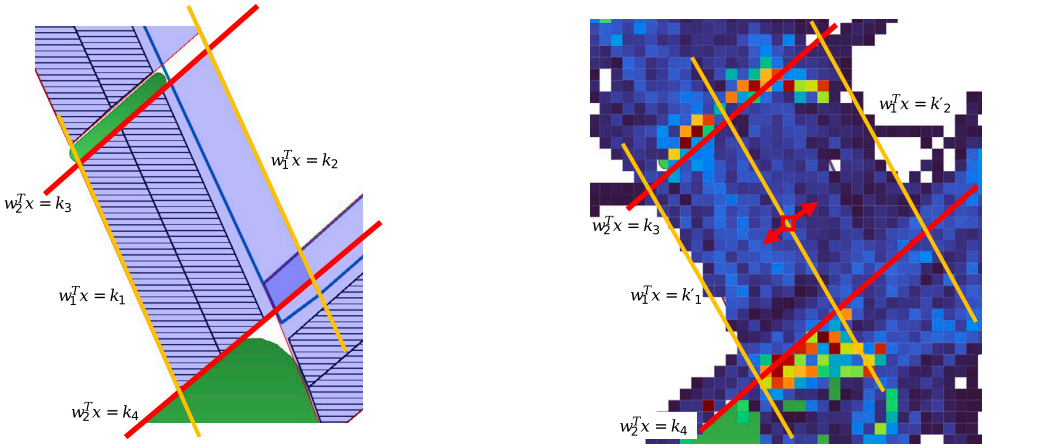


Fig. 4. Recognition of actual occurrence spaces of waiting behaviors.



(a) Space designed for passing behavior (b) Schematic diagram of the regional growth algorithm

Fig. 5. Recognition of the actual occurrence space of passing behaviors.

3.4.2. *Actual occurrence space of passing behaviors*

A grid c_i is considered part of the passing behavior space if $G(T, c_i) > 0 \quad and \quad c_i \notin C_{wd}$. Our study specifically examines regular spaces that serve as connecting pathways between two designed waiting space for waiting, which are used as actual spaces for people to pass. These spaces function as a platform for enlarging the area according to the grid's occupancy, thereby establishing the actual occurrence space.

As shown in Fig. 5(a), we consider the designed space of passing behavior as a regular quadrilateral. Its border is defined by two sets of parallel lines (shown by the red lines and yellow lines in Fig. 5(a)). The yellow lines in Fig. 5(b) represents the actual boundaries of the occurrence space. To determine the actual occurrence spaces for passing behaviors, denoted as $C_{Pui}, i = 1, 2.....k$, the region growth algorithm (Algorithm 1) is used to extend the left and right boundaries (the yellow line in Fig. 5(b)). The borderlines are depicted by normal vector w^T and intercept k, and the space decided by these borderlines is denoted as B. In Algorithm 1, $Getgrids()$ is the function designed for archiving a set of grids which are located between the borders, defined in Eq. (3). Additionally, the function $Coveragerate()$ as defined in Eq. (4) calculates the coverage rate for C_{Pu}.

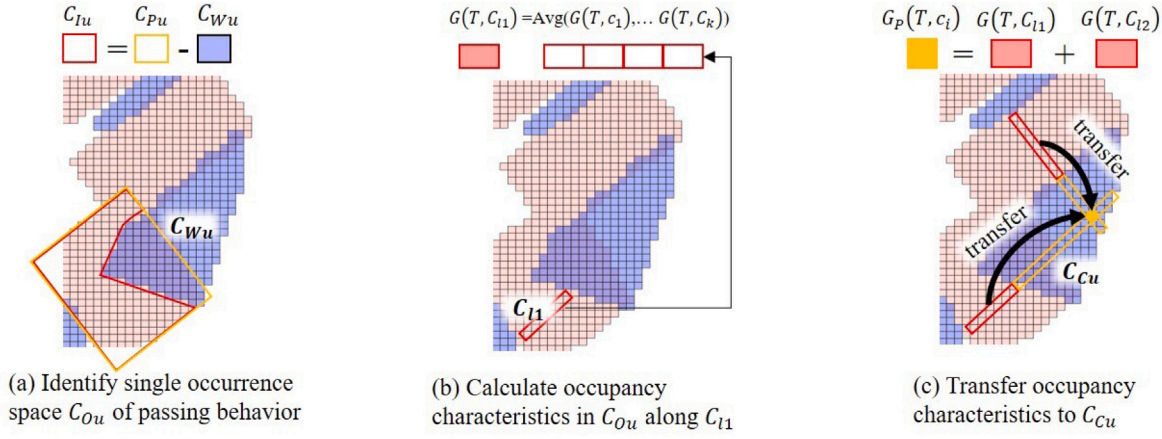
$$GetGrids(B) = \{ID(c_i) \mid \forall c_i \in C, c_i \quad is \quad within \quad B\} \tag{3}$$

$$Coveragerate(C_{Pu}) = \frac{|C_1|}{|C_{Pu}|}, \; C_1 = \{c_i | c_i \in C_{Pu} \quad and \quad G(T, c_i) > 0\} \tag{4}$$

Algorithm 1: Regional Growth Algorithm

Data: C: the whole intersection space, C_{Pd}: design passing space, $w_1^T \cdot X = k_i$: left and right boundaries of the design passing space, F: coverage rate threshold, S: growth step
Result: C_{Pu}: main occurrence space of passing behaviors

```
begin
    c_i ← max(G(T, C_Pd));
    k' ← Findintercept(w^T, c_i);
    k_1 ← k';
    k_2 ← k';
    f ← 0;
    while f ≤ F + (1 − F)/2 do
        k_1 ← k_1 − S;
        C_Pu ← Getgrids(w^T, k_1, k_2);
        f ← Coveragerate(C_Pu);
    end
    while f ≤ F do
        k_2 ← k_2 + S;
        C_Pu ← Getgrids(w^T, k_1, k_2);
        f ← Coveragerate(C_Pu);
    end
    return C_Pu
end
```



(a) Identify single occurrence space C_{Ou} of passing behavior

(b) Calculate occupancy characteristics in C_{Ou} along C_{l1}

(c) Transfer occupancy characteristics to C_{Cu}

Fig. 6. Process of characteristics transfer.

3.5. Estimation of grid-level passing or waiting probabilities

The waiting probability $P_W(T, c_i)$ of pedestrians and non-motorized vehicles at time interval T is estimated using Eq. (5) in grid c_i. The passing probability $P_P(T, c_i)$ is determined using Eq. (6).

$$P_W(T, c_i) = 1 - \frac{G_P(T, c_i)}{G(T, c_i)} \tag{5}$$

$$P_P(T, c_i) = \frac{G_P(T, c_i)}{G(T, c_i)} \tag{6}$$

In independent occurrence space of passing behavior C_{Iu} (shown in Fig. 6a), grids are exclusively utilized for passing, and thus, the occupancy characteristics of a grid are solely determined by the passing behaviors, i.e. $G\left(T, c_i\right) = G_P(T, c_i)$. Consequently, the probability of passing in this space is unequivocally 1.

The grids within the combined occurrence space exhibit dual behaviors where $G_P(T, c_i) > 0$ and $G_W(T, c_i) > 0$. $G_P(T, c_i)$ is achieved by utilizing the behavioral transmissibility inferred from the occupancy characteristics of the independent occurrence spaces. The computation entails the subsequent key procedures:

Step 1. **Identify independent occurrence space of passing behavior**. As shown in Fig. 6a, determine the grids from independent occurrence space connected with waiting behavior occurrence space C_{Wu}. Define these spaces as $C_{Iu} = \{c_i | c_i \in C_{Pu} \ and \ c_i \notin C_{Wu}\}$.

Step 2. **Calculate occupancy characteristics of passing behavior**. As shown in Fig. 6b, calculate the average occupancy characteristics of passing behavior on each pass path in C_{Iu}. These pass paths are straight lines that generate by the start grids and direction of C_{Iu}. As an example, for path $l_1 : w_1^T \cdot X = k_1$, the average is calculated using Eq. (7).

$$G_P\left(T, C_{l_1}\right) = \frac{\sum_{c_i}^{C_{l_1}} G\left(T, c_i\right)}{|C_{Iu}|} \tag{7}$$

Step 3. **Transfer occupancy characteristics of passing behavior**. As shown in Fig. 6c, for each grid c_i in the waiting behavior occurrence space that falls on a passing path, ascertain the occupancy characteristics of passing behavior by transferring the features of all pertinent paths. The function $G_P\left(T, c_i\right)$ represents the sum of the values of $G_P\left(T, C_l\right)$ for all paths in the set L that include the element c_i.

Step 4. **Determine behavioral probability in combined occurrence space**. After calculating the occupancy characteristics for both passing and waiting behaviors, determine the behavioral probability of each grid. The probability of waiting and passing behaviors at grid c_i is determined by Eqs. (5) and (6).

4. Experiment

Section 4.1 provides a brief introduction to the experimental setup, while Section 4.2 discusses the recognition accuracy using various methods under different MOTA metrics. Following that, the paper presents an analysis of sensitivity, runtime, and an ablation research. The practical significance of our approach will be examined in Section 4.6.

4.1. Experiment setup

Experiment datasets. We adopt infrastructure data and trajectory data from the real-world as our experiment datasets, where is collected using a DJI Mavic 2 ZOOM model drone. The detailed descriptions of these datasets are as follows:

- **infrastructure data:** The data comes from the intersection of Jiangnan and Changgang Road within the city of Guangzhou, China. There are infrastructure of Slow moving transportation, seen in Fig. 1(b).
- **trajectory data:** This dataset is collected using a DJI Mavic 2 ZOOM model drone. The road traffic was recorded in 2k resolution at 50 FPS. The yolov7 model is used for target identification, based on which the tracking Bot-sort model is used for target tracking to extract the individual labels and behavioral trajectories of the travelers' non-motorized vehicles. The DTP data comes from the trajectory, after processed with the 0% MOTA.

Comparison methods. In this paper, we compare our DCGM framework with three baseline methods. The details of the compared methods are as follows:

- **CB-SMoT** [28]**:** The CB-SMoT method is a velocity-based spatio-temporal clustering method for processing single trajectory data. It not only finds the user's expected stops and moves, but also discovers interesting locations that the user did not expect.
- **Hwang** [29]**:** The method is an algorithm for segmenting GPS trajectory data into wait and pass segments. By filling time gaps, state-dependent path-based interpolation and spatio-temporal clustering-based segmentation methods, it effectively handles signal loss and noise, extracts meaningful trajectory segments, and is able to accurately identify and estimate wait and movement segments.
- **SVBDSA** [30]**:** This method proposes an algorithm for discovering stay regions in indoor dynamic trajectory data. The new algorithm, SVBDSA, determines the stabilization value by calculating the distance and velocity between two points, which leads to the discovery of stay regions.

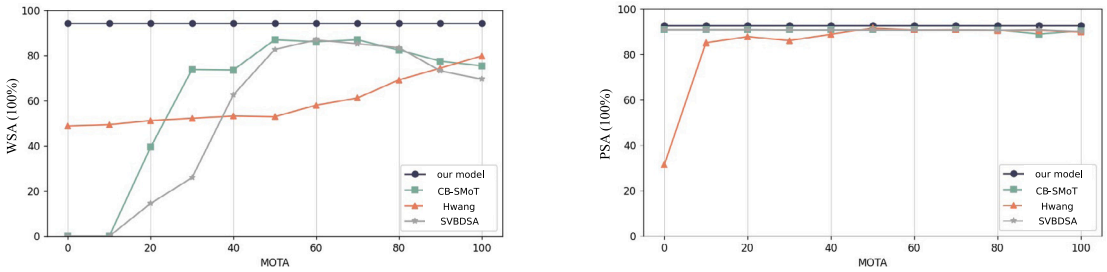
The canny algorithm model employs the default parameters (100, 150), and the growth threshold F is set at 0.9 for the growth algorithm.

4.2. Recognition results with different methods in various MOTA

4.2.1. Recognition results of actual occurrence spaces of behaviors

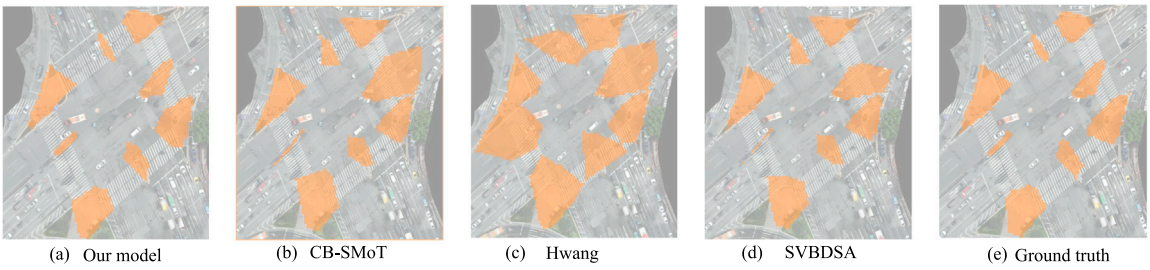
We aim to evaluate the effectiveness of space recognition from three dimensions: shape, size, and spatial arrangement. The shape is quantified by the contour index, while the size is measured by the number of grid cells. Mean Squared Error (MSE) is used to assess errors related to shape and size, whereas spatial discrepancies are calculated using the non-overlap rate. The recognition accuracy of the actual occurrence spaces for waiting and passing behaviors are denoted as WSA and PSA, respectively.

Fig. 7 illustrates the fluctuation in recognition accuracy across different methods for behavior occupancy space at various levels of MOTA. The graph demonstrates a positive correlation between the rise in MOTA for all compared methods. Nevertheless, it is clear that our procedures regularly attain the utmost precision.



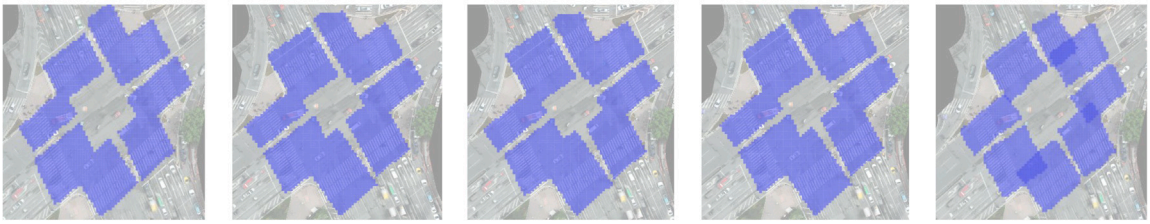
(a) WSA (b) PSA

Fig. 7. Performance of different models for actual occurrence spaces of behaviors under various MOTA.



(a) Our model (b) CB-SMoT (c) Hwang (d) SVBDSA (e) Ground truth

(a) actual occurrence space of waiting behavior



(a) Our model (b) CB-SMoT (c) Hwang (d) SVBDSA (e) Ground truth

(b) actual occurrence space of passing behavior

Fig. 8. Illustrations of actual occurrence spaces from different models (MOTA = 80%).

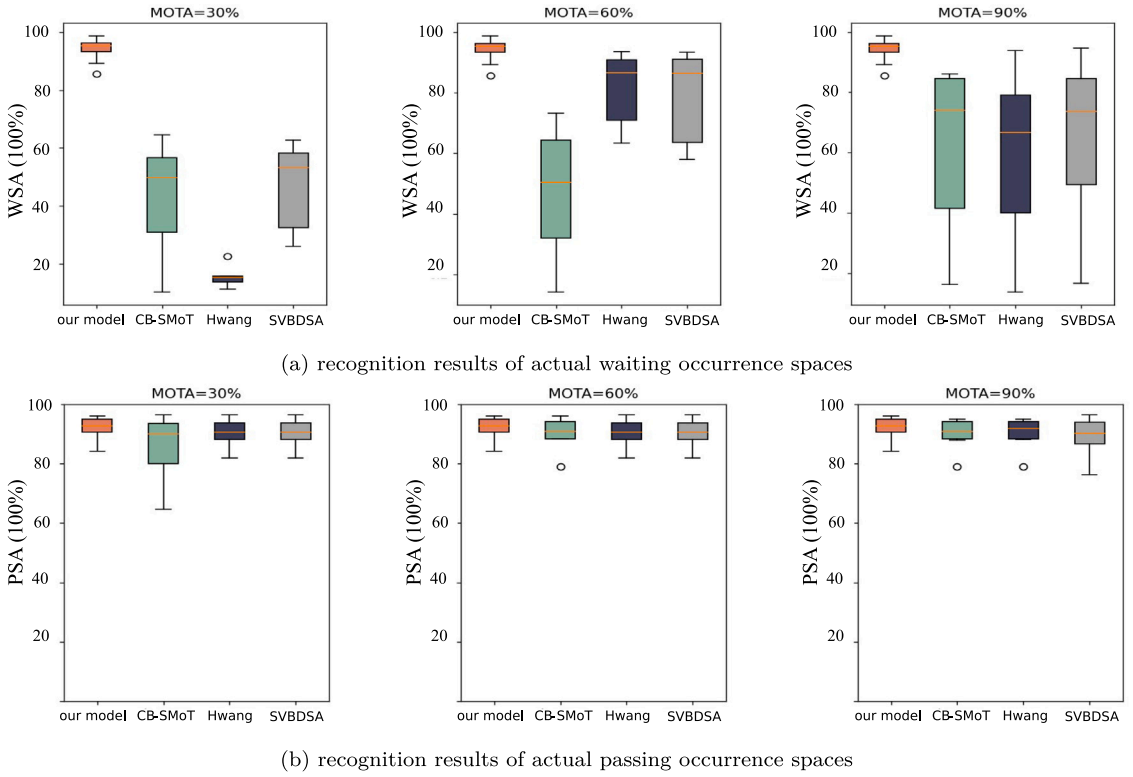
Surprisingly, our method is not influenced by MOTA, maintaining an accuracy of over 92% in recognizing both actual occurrence spaces of waiting and passing behaviors. In contrast, the competing methods demonstrate significantly lower performance at reduced MOTA levels, particularly in identifying the occupancy space of waiting behaviors. This decline in performance is attributed to inadequate tracking of non-motorized pedestrians at low MOTA levels, where only short movements can be detected, therefore posing a challenge in precisely identifying behaviors.

As depicted in Fig. 8, the occupancy space identified by our method is strategically distributed around several safety islands, aligning closely with actual conditions. On the other hand, the space that is recognized by the comparison approaches is significantly bigger. This is because our methodology focuses exclusively on significant actions within the group, whereas comparison methods are susceptible to incorrectly recognizing individual behaviors.

Fig. 9 displays the box plots that illustrate the performance of each approach in different spaces recognition at MOTA levels of 30%, 60%, and 90%. From Fig. 9(b), it is clear that the different methods consistently achieve high recognition accuracy for spaces occupied by passing behaviors. This consistency is observed not only in the average values but also across each individual passing space. In contrast, Fig. 9(a) highlights the inefficiency of the comparison methods in identifying actual occurrence spaces of waiting behavior. The results of the comparative approaches exhibit substantial variety, with certain areas being highly recognized while others are poorly recognized. As depicted in Fig. 8, the comparative method achieves high recognition accuracy in space 5 but performs poorly in space 4, displaying significant error in terms of shape, size, and spatial attributes.

4.2.2. Recognition results of grid-level probability of behavior

To evaluate the accuracy of the model's predictions, we used the mean square error (MSE) between the predicted values and the actual data in grid-level. In Eq. (8), $p'(c_i)$ and $p(c_i)$ represent the actual and estimated probabilities of waiting behavior, respectively.



(a) recognition results of actual waiting occurrence spaces

(b) recognition results of actual passing occurrence spaces

Fig. 9. Box plots of recognition results of actual occurrence spaces.

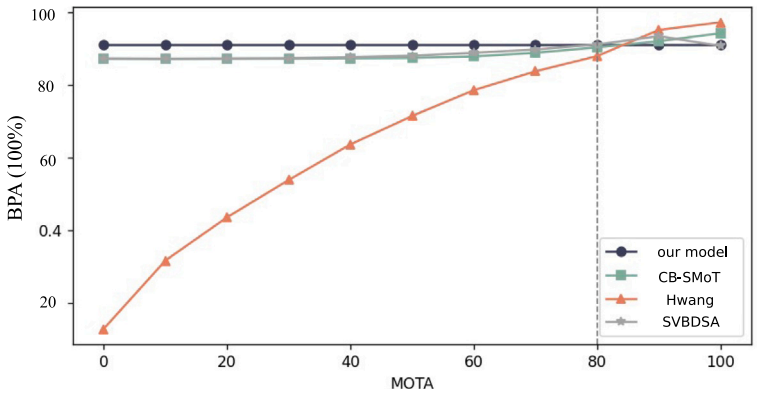


Fig. 10. Accuracies of grid-level probabilities for waiting behaviors under different MOTA.

The Behavioral Probability Accuracy (BPA) is calculated using the formula $1 - E_r(c_i)$.

$$E_r(c_i) = \frac{1}{|C|} \sum_{c_i}^{C} (p'(c_i) - p(c_i))^2 \tag{8}$$

We evaluated the accuracy of several behavior identification methods at different MOTA values, as depicted in Fig. 10. Our method demonstrated stable accuracy above 91% across all MOTA levels. In contrast, the accuracy of comparison methods, particularly the Hwang method, was highly sensitive to changes in MOTA. Although in the case of full tracking (MOTA = 100%), the behavior recognition accuracy of the comparison method exceeds that of our method. These approaches demonstrated enhanced precision with larger MOTA values, but performed inadequately compared to our approach when the tracking rate dropped below 80%. This trend of decreasing accuracy with declining tracking rates was consistent among the comparative methods.

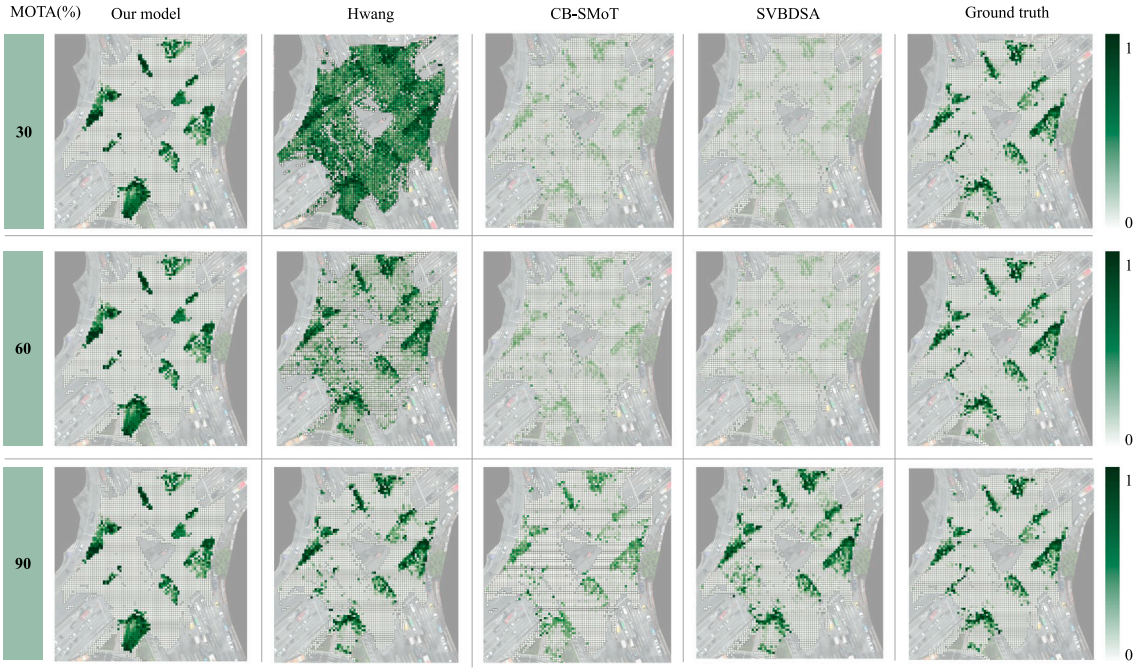


Fig. 11. Waiting probability under various methods.

Fig. 11 illustrates the actual and estimated distribution of grid-level probabilities in the intersection at MOTA levels of 30%, 60% and 90% . The green resident grids in the junction are primarily located around the eight safety islands, functioning independently of one another. The grids with high waiting probability are mainly spread along the edges. Comparing the results of our method, it is evident that there is a significant resemblance in the overall arrangement and coloration of the residency grid, notwithstanding a few mistakes observed at the southernmost safety island.

At a MOTA of 90%, the comparison methods successfully identify the actual distribution of the resident grids. However, they also mistakenly recognize extra resident grids, which negatively affects the overall accuracy of recognition. This issue arises since the majority of comparison methods generally concentrate on examining behaviors that take place on a wider magnitude, rather than in local areas like junctions.

At a MOTA of 60%, as the tracking rate declines, the effectiveness of the comparison methods decreases. The Hwang method has a tendency to classify a greater range of actions as waiting, leading to a larger grid of standing behaviors. The diminished precision shown in the CB-SMoT and SVBDSA methods can be attributed to their intrinsic tendency to underestimate waiting behaviors, frequently misidentifying them as passing behaviors. This leads to a reduced probability of waiting presence within the grids, manifesting as a greener overall hue in Fig. 11.

At a MOTA of 30%, it is clear that Hwang method mistakes all behavior in a large number of grids for resident behavior, due to the large number of short trajectory effects from low tracking rates. At the same time, Fig. 11 shown in the CB-SMoT and SVBDSA method are all density-based methods, which are still able to obtain results with a certain level of accuracy for the density of short trajectories at low tracking rates.

Our method emphasizes the main areas where residents behave, in order to reduce the influence of scattered resident grids. In addition, it consistently achieves accurate recognition results across all levels of MOTA, even when operating at reduced tracking rates.

4.3. Sensitivity study

This section assesses the influence of different data input conditions on the effectiveness of our method. Fig. 12a examines the influence of extending the original sampling interval from 20 to 6000 ms on the accuracy of our technique. The line graph indicates that as the sampling interval increases, the accuracy drops, albeit at a gradual pace.

Fig. 12b illustrates the results under different object detection loss. The accuracy of our method maintains high accuracy up to 80% leakage detection rate, and then the minor losses in precision occur. That is because the dense intersections consist of a large number of items, which makes the overall population resistant to partial losses.

Fig. 12c presents the effects resulting from the implementation of loss detection processes on specific grids. Due to the significant dependence on spatial factors in our approach, higher loss rates result in a noticeable decrease in accuracy, highlighting the method's susceptibility to spatial data reliability.

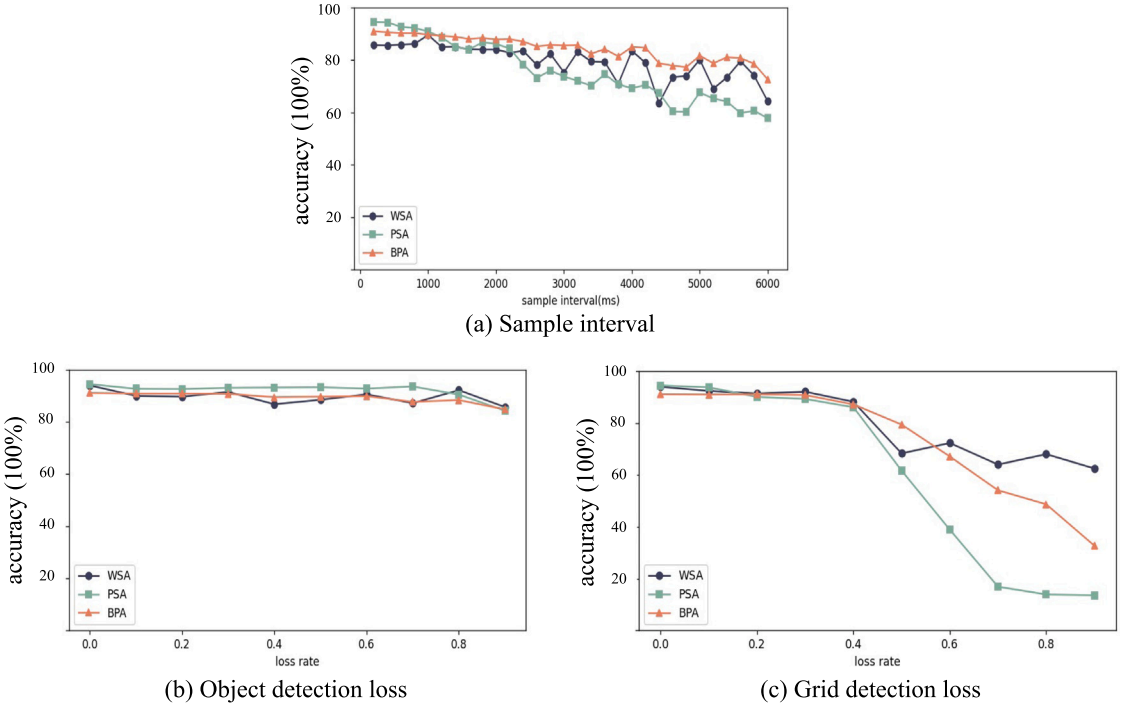


Fig. 12. Recognition results in various detection condition.

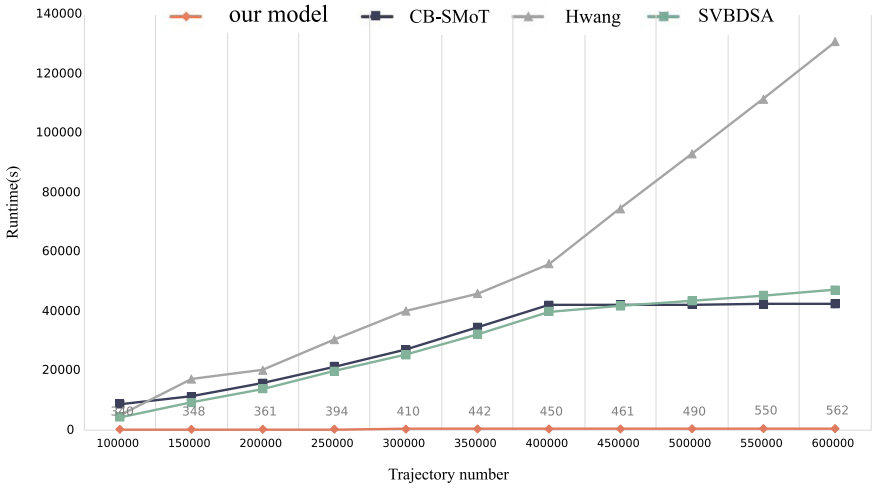
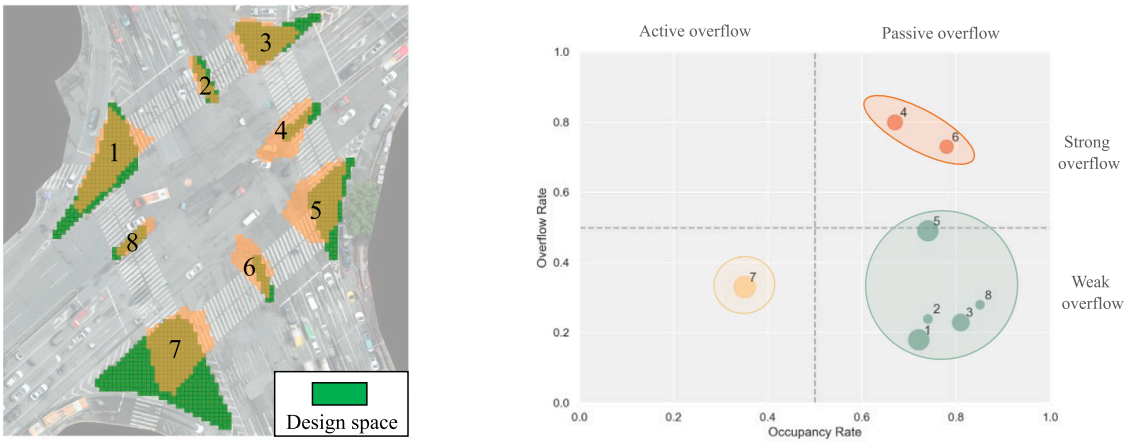


Fig. 13. Runtime of different methods.

4.4. Runtime study

Our approach analyzes the trajectory points by matching them to a grid based on the occupancy characteristics of the grid, and its running time depends on the number of grids. The comparison methods determine the behavior of the object by analyzing the spatio-temporal characteristics of each trajectory, and their running time depends on the number of trajectories. We have designed experiments to compare the runtime performance of our method with that of the comparison methods. All algorithms are run on the same hardware configuration to ensure fair and comparable results. As shown in Fig. 13 The runtime of the continuous trajectory based method is significantly higher than that of our method. Furthermore, as the number of trajectories increases, the runtime of the baseline methods increases dramatically, which is different from the stability of the time consumption of our method.



(a) space overflow (b) overflow rate

Fig. 14. Waiting space analysis.

Table 2
The model accuracy in ablation study.

Method	WSA	PSA	BPA
Baseline(A)	0.537	0.857	0.855
Model B(A+B)	0.943	0.927	0.883
Model A(A+C)	0.537	0.857	0.897
Model C(A+B+C)	0.943	0.927	0.911

4.5. Ablation study

The methodology outlined in this paper is segmented into three primary components: grid generation(A), space recognition(B), and characteristic transfer(C). Space recognition refers to the utilization of the original design space rather than the actual used space. In the absence of characteristic transfer, the probability of behaviors in combined occurrence space is substituted with a default value of 0.5. As in Table 2, we used the model that simply includes operation A as baseline, and successively added different key components to examine their contributions on performance.

Space recognition. The use of the space recognition procedure significantly enhanced precision, particularly in regions predominantly linked to waiting behavior, by a 40% margin. The operation successfully detected spatial utilization beyond the intended design parameters and uncovered pedestrian inclinations towards space utilization at junctions. The operation also contributes to enhancing the accuracy of behavioral recognition, resulting in a 2.7% improvement.

Characteristic transfer. We assess the influence of the characteristic transfer procedure on the outcomes. The operation can effectively distinguish between waiting and passing behaviors in combined occurrence space, rather than assuming an equal occurrence of the two types of behaviors. There is a 4.2% increase in accuracy in recognizing spatial patterns compared to the baseline.

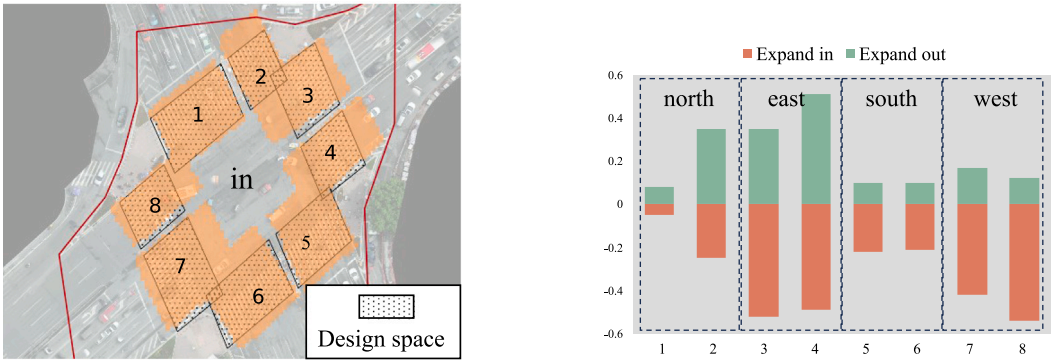
When comparing Model A with Model B, it is evident that while the characteristic transfer does not enhance the accuracy of spatial recognition results, it does demonstrate improved performance in recognizing behaviors. The remarkable precision of Model C in comparison to Models A and B suggests that the combined impact of spatial and behavioral operations synergistically improves overall recognition abilities.

4.6. Application value discussion

In our final analysis, we draw conclusions from the identification results to elucidate the practical application of our method.

Despite the conventional inclusion of safety islands and crosswalks in intersection design to accommodate slow-moving vehicles, our research has uncovered a different way in which non-motorized transport use the space. As depicted in Fig. 14(a), there is a distinct discrepancy between the spaces occupied by waiting behaviors and those designated as safety islands.

We measured this difference by utilizing indicators such as the proportion of overflow area and the rate of safety island usage. The outcomes are presented in Fig. 14(b). Spaces with large overflow areas and strong use of safety islands, such as areas 4 and 6, demonstrate a passive overflow pattern. Overflow commonly happens beyond the safety island, particularly in roadway safety islands that are limited in space. These places are considered to effectively utilize the safety islands despite the overflow. On the other hand, spaces 1, 2, 3, 5, and 8 are considered to be efficiently utilizing the safety islands. These spaces have high usage rates but limited overflow, which suggests that they are being used optimally within their intended capacity.



(a) space overflow (b) overflow rate

Fig. 15. Passing space analysis.

Space 7 depicts a distinct scenario in which overflow continues to happen despite the presence of abundant safety island space. This suggests a current overflow pattern, most likely caused by the pressing need for walkers and non-motorized vehicles to cross.

These observations emphasize the subtle connections between urban infrastructure and user actions, emphasizing areas where intersection design may be enhanced to effectively meet the changing requirements of all users.

Fig. 15(b) depicts the proportion of non-motorized transport that exceeds the capacity of the crosswalk. The green color indicates overflow toward the outer side of the safety island, while the yellow color indicates overflow toward the inner side of the safety island. The study results suggest that pedestrians and non-motorized vehicles often cross the roadway outside the specified crosswalk and have a clear tendency to enter the intersection.

This observation reveals the exact pathways used by non-motorized transport and provides vital data for improving intersection traffic control. Expanding the crosswalk into the intersection can improve the ability to handle traffic patterns, resulting in increased safety and efficiency.

5. Conclusions

This study presents a method for recognizing the behavior of pedestrians and non-motorized vehicles using a two-channel grid model. The system accepts DTP data as input and calculates the actual occurrence space and grid-level probability by analyzing the occupancy and semantic characteristics. Experimental results demonstrate that our method outperforms the comparison method in terms of recognition accuracy and processing speed, particularly in scenarios with low MOTA. The proposed method can facilitate the space identification of pedestrians and non-motorized vehicle and their behavioral probabilities. This information can be utilized to enhance the management strategy for chronic traffic crossing and enhance overall traffic safety.

The method possesses two primary benefits: (1) The method can accurately detect the behaviors of pedestrians and non-motorized vehicles, even in high-density intersection scenarios where individual tracking is challenging. (2) This approach utilizes grid space to discretize and compute the entire intersection space. It combines the semantic characteristics and the actual usage characteristics of the intersection to provide a comprehensive understanding of the behaviors occurring in each space of the intersection.

The model presented in this paper still has certain limitations. However, there are two factors that can be employed to improve this study. This paper's methodology focuses solely on analyzing the fundamental pedestrian non-motorized crossing behaviors of stopping and passing. It acknowledges the existence of more intricate behaviors, which will be the subject of future research. Currently, we analyze pedestrian and non-motorized vehicle groups as a unified entity, but there are actually differences in the behavior of these objects at intersections. Effectively distinguishing between the two types of objects is an area for improvement in this paper.

CRediT authorship contribution statement

Huanting Xu: Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Software, Resources, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization. **Zhaocheng He:** Writing – review & editing, Validation, Supervision, Resources, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Conceptualization. **Yiyang Chen:** Visualization, Validation, Data curation. **Zhigang Wu:** Writing – review & editing, Writing – original draft, Methodology. **Yiting Zhu:** Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Resources, Project administration, Data curation, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgments

Thanks for the reviewers for their valuable comments on the paper. This work is supported by the National Key Research and Development Program of China (2023YFB4301900), National Natural Science Foundation of China (52202406 and U21B2090), China Postdoctoral Science Foundation, China (2023M744002), Postdoctoral Fellowship Program of CPSF (GZC20233244), the Science and Technology Program of Shenzhen (grant No. SGG20220831094604008).

References

[1] W.H. Organization, World Health Statistics 2023: Monitoring Health for the Sdgs, Sustainable Development Goals, Tech. Rep., World Health Organization (WHO), 2023, Licence: CC BY-NC-SA 3.0 IGO.
[2] A. Kumar, I. Ghosh, Non-compliance behaviour of pedestrians and the associated conflicts at signalized intersections in India, Saf. Sci. 147 (2022) 105604, http://dx.doi.org/10.1016/j.ssci.2021.105604.
[3] R. Raoniar, S. Maqbool, A. Pathak, M. Chugh, A.K. Maurya, Hazard-based duration approach for understanding pedestrian crossing risk exposure at signalised intersection crosswalks – a case study of kolkata, India, Transp. Res. F 85 (2022) 47–68, http://dx.doi.org/10.1016/j.trf.2021.12.015.
[4] J.O. Malenje, J. Zhao, P. Li, Y. Han, An extended car-following model with the consideration of the illegal pedestrian crossing, Phys. A 508 (2018) 650–661, http://dx.doi.org/10.1016/j.physa.2018.05.074.
[5] Q. Sun, C. He, Y. Wang, H. Liu, F. Ma, X. Wei, Reducing violation behaviors of pedestrians considering group interests of travelers at signalized crosswalk, Phys. A 594 (2022) 127023, http://dx.doi.org/10.1016/j.physa.2022.127023.
[6] Q. Sun, H. Liu, Y. Wang, Q. Li, W. Chen, P. Bai, C. Xue, Cooperation in the jaywalking dilemma of a road public good due to points guidance, Chaos Solitons Fractals 160 (2022) 112277, http://dx.doi.org/10.1016/j.chaos.2022.112277.
[7] Q. Sun, C. He, Y. Wang, H. Liu, F. Ma, X. Wei, Reducing violation behaviors of pedestrians considering group interests of travelers at signalized crosswalk, Phys. A 594 (2022) 127023, http://dx.doi.org/10.1016/j.physa.2022.127023.
[8] H. Wang, A. Kläser, C. Schmid, C.-L. Liu, Dense trajectories and motion boundary descriptors for action recognition, Int. J. Comput. Vis. 103 (1) (2013) 60–79, http://dx.doi.org/10.1007/s11263-012-0594-8.
[9] Z. Wang, P. Peng, K. Geng, X. Cheng, X. Zhu, J. Chen, G. Yin, Analysis of pedestrian crossing behavior based on centralized unscented Kalman filter and pedestrian awareness based social force model, Phys. A 632 (2023) 129350, http://dx.doi.org/10.1016/j.physa.2023.129350.
[10] R. Yi, M. Du, W. Song, J. Zhang, Fast trajectory extraction and pedestrian dynamics analysis using deep neural network, Phys. A 638 (2024) 129611, http://dx.doi.org/10.1016/j.physa.2024.129611.
[11] L. Wang, Y. Xiong, Z. Wang, Y. Qiao, D. Lin, X. Tang, L. Van Gool, Temporal segment networks: Towards good practices for deep action recognition, in: B. Leibe, J. Matas, N. Sebe, M. Welling (Eds.), Computer Vision – ECCV 2016, Springer International Publishing, Cham, 2016, pp. 20–36.
[12] S. Ji, W. Xu, M. Yang, K. Yu, 3D convolutional neural networks for human action recognition, IEEE Trans. Pattern Anal. Mach. Intell. 35 (1) (2013) 221–231, http://dx.doi.org/10.1109/TPAMI.2012.59.
[13] W. Xue, M. Yang, R. Liu, Y. Takuma, Y. Takahiro, K. Takeshi, Spatial-temporal graph convolutional network for skeleton-based gait recognition, in: 2022 3rd International Conference on Pattern Recognition and Machine Learning, PRML, 2022, pp. 77–82, http://dx.doi.org/10.1109/PRML56267.2022.9882242.
[14] L. Shi, Y. Zhang, J. Cheng, H. Lu, Two-stream adaptive graph convolutional networks for skeleton-based action recognition, 2019, pp. 12018–12027, http://dx.doi.org/10.1109/CVPR.2019.01230.
[15] J. Donahue, L.A. Hendricks, M. Rohrbach, S. Venugopalan, S. Guadarrama, K. Saenko, T. Darrell, Long-term recurrent convolutional networks for visual recognition and description, 2017, http://dx.doi.org/10.1109/TPAMI.2016.2599174.
[16] K. Chen, X. Song, D. Han, J. Sun, Y. Cui, X. Ren, Pedestrian behavior prediction model with a convolutional LSTM encoder–decoder, Phys. A 560 (2020) 125132, http://dx.doi.org/10.1016/j.physa.2020.125132.
[17] A. Elfes, Using occupancy grids for mobile robot perception and navigation, Computer 22 (6) (1989) 46–57, http://dx.doi.org/10.1109/2.30720.
[18] W. Xu, L. Liu, S. Zlatanova, W. Penard, Q. Xiong, A pedestrian tracking algorithm using grid-based indoor model, Autom. Constr. 92 (2018) 173–187, http://dx.doi.org/10.1016/j.autcon.2018.03.031.
[19] Z. Sun, Q. Ke, H. Rahmani, M. Bennamoun, G. Wang, J. Liu, Human action recognition from various data modalities: A review, IEEE Trans. Pattern Anal. Mach. Intell. 45 (3) (2023) 3200–3225, http://dx.doi.org/10.1109/TPAMI.2022.3183112.
[20] H. Wang, C. Schmid, Action recognition with improved trajectories, in: 2013 IEEE International Conference on Computer Vision, 2013, pp. 3551–3558, http://dx.doi.org/10.1109/ICCV.2013.441.
[21] D. Tran, L. Bourdev, R. Fergus, L. Torresani, M. Paluri, Learning spatiotemporal features with 3D convolutional networks, in: 2015 IEEE International Conference on Computer Vision, ICCV, 2015, pp. 4489–4497, http://dx.doi.org/10.1109/ICCV.2015.510.
[22] W. Burgard, D. Fox, D. Hennig, T. Schmidt, Estimating the absolute position of a mobile robot using position probability grids, in: Proceedings of the Thirteenth National Conference on Artificial Intelligence - Volume 2, AAAI '96, AAAI Press, 1996, pp. 896–901.
[23] S. Wolfram, Statistical mechanics of cellular automata, Rev. Modern Phys. 55 (1983) 601–644.
[24] Z. Fu, Q. Jia, J. Chen, J. Ma, K. Han, L. Luo, A fine discrete field cellular automaton for pedestrian dynamics integrating pedestrian heterogeneity, anisotropy, and time-dependent characteristics, Transp. Res. C 91 (2018) 37–61, http://dx.doi.org/10.1016/j.trc.2018.03.022.
[25] C.-Z. Xie, T.-Q. Tang, P.-C. Hu, L. Chen, Observation and cellular-automaton based modeling of pedestrian behavior on an escalator, Phys. A 605 (2022) 128032, http://dx.doi.org/10.1016/j.physa.2022.128032.
[26] R.J.B.J. Shuqi Xue, X. Zhang, Pedestrian counter flow in discrete space and time: experiment and its implication for CA modelling, Transp. B 7 (1) (2019) 169–184, http://dx.doi.org/10.1080/21680566.2017.1365662.
[27] J. Geng, L. Xia, J. Xia, Q. Li, H. Zhu, Y. Cai, Smartphone-based pedestrian dead reckoning for 3D indoor positioning, Sensors 21 (24) (2021) http://dx.doi.org/10.3390/s21248180.
[28] A.T. Palma, V. Bogorny, B. Kuijpers, L.O. Alvares, A clustering-based approach for discovering interesting places in trajectories, in: Proceedings of the 2008 ACM Symposium on Applied Computing, SAC '08, Association for Computing Machinery, New York, NY, USA, 2008, pp. 863–868, http://dx.doi.org/10.1145/1363686.1363886.
[29] C.V.N.D. Sungsoon Hwang, R.T. Crews, Segmenting human trajectory data by movement states while addressing signal loss and signal noise, Int. J. Geogr. Inf. Sci. 32 (7) (2018) 1391–1412, http://dx.doi.org/10.1080/13658816.2018.1423685.
[30] Y. Zhou, Y. Chen, D. Pi, Discovery of stay area in indoor trajectories of moving objects, Expert Syst. Appl. 170 (2021) 114501, http://dx.doi.org/10.1016/j.eswa.2020.114501.