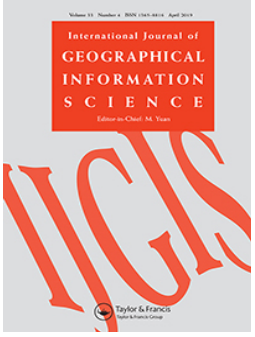


International Journal of Geographical Information Science

ISSN: 1365-8816 (Print) 1362-3087 (Online) Journal homepage: https://www.tandfonline.com/loi/tgis20

Inferring user tasks in pedestrian navigation from eye movement data in real-world environments

Hua Liao, Weihua Dong, Haosheng Huang, Georg Gartner & Huiping Liu

To cite this article: Hua Liao, Weihua Dong, Haosheng Huang, Georg Gartner & Huiping Liu (2019) Inferring user tasks in pedestrian navigation from eye movement data in real-world environments, International Journal of Geographical Information Science, 33:4, 739-763, DOI: 10.1080/13658816.2018.1482554

To link to this article: https://doi.org/10.1080/13658816.2018.1482554

Published online: 26 Jun 2018.

INTERNATIONAL JOURNAL OF GEOGRAPHICAL INFORMATION SCIENCE
2019, VOL. 33, NO. 4, 739–763
https://doi.org/10.1080/13658816.2018.1482554

Taylor & Francis
Taylor & Francis Group

RESEARCH ARTICLE

Check for updates

Inferring user tasks in pedestrian navigation from eye movement data in real-world environments

Hua Liao[a,b], Weihua Dong[a], Haosheng Huang[c], Georg Gartner[b] and Huiping Liu[a]

[a]State Key Laboratory of Remote Sensing Science, Beijing Key Laboratory for Remote Sensing of Environment and Digital Cities, and Faculty of Geographical Science, Beijing Normal University, Beijing, China; [b]Department of Geodesy and Geoinformation, Vienna University of Technology, Vienna, Austria; [c]GIScience Center, Department of Geography, University of Zurich, Zurich, Switzerland

ABSTRACT

Eye movement data convey a wealth of information that can be used to probe human behaviour and cognitive processes. To date, eye tracking studies have mainly focused on laboratory-based evaluations of cartographic interfaces; in contrast, little attention has been paid to eye movement data mining for real-world applications. In this study, we propose using machine-learning methods to infer user tasks from eye movement data in real-world pedestrian navigation scenarios. We conducted a real-world pedestrian navigation experiment in which we recorded eye movement data from 38 participants. We trained and cross-validated a random forest classifier for classifying five common navigation tasks using five types of eye movement features. The results show that the classifier can achieve an overall accuracy of 67%. We found that statistical eye movement features and saccade encoding features are more useful than the other investigated types of features for distinguishing user tasks. We also identified that the choice of classifier, the time window size and the eye movement features considered are all important factors that influence task inference performance. Results of the research open doors to some potential real-world innovative applications, such as navigation systems that can provide task-related information depending on the task a user is performing.

ARTICLE HISTORY
Received 20 August 2017
Accepted 27 May 2018

KEYWORDS
Wayfinding; random forests; task inference; eye tracking; machine learning

1. Introduction

Navigation or the goal-directed movement to a distal destination through a familiar or unfamiliar environment (Golledge 1999, Montello 2005), is a cognitively demanding process (Montello and Raubal 2012). Pedestrians must accomplish a series of tasks to reach their destinations successfully, such as finding where they are, determining the direction they are facing, planning the route to the destination, maintaining the correct orientation while moving and recognizing the destination (Downs and Stea 1977). Such cognitively demanding tasks drive the need to develop pedestrian navigation applications that are adaptive to user context and task-related information needs (Jiang and Yao 2006). Over the years, multiple solutions have been proposed for inferring high-level user context

CONTACT Weihua Dong ✉ dongweihua@bnu.edu.cn

and satisfying users' information needs. Examples include recommending routes based on user history data (Huang and Gartner 2012) and affective responses (Huang *et al*. 2014), inferring trip purposes by analysing GPS trajectories (Gong *et al*. 2015) and suggesting points-of-interest based on users' check-in behaviour (Liu *et al*. 2013).

Among various user behaviours, visual behaviour is considered to provide more direct and unambiguous cues that indicate a user's cognitive states and intentions than historical data (Bednarik *et al*. 2012, Anagnostopoulos *et al*. 2017). It is acknowledged that human visual attention in map reading and scene perception is guided by high-level (top-down) factors such as expert knowledge and user tasks (Henderson 2003, Wolfe *et al*. 2003). Different user tasks can produce different eye movement patterns (Yarbus *et al*. 1967). Thus, the question of whether or not it is possible to infer user tasks from eye movements arises. Multiple studies have explored using machine-learning methods to infer user tasks from eye movement data, and many of them have reported positive results (Bulling *et al*. 2011, Borji and Itti 2014, Boisvert and Bruce 2016). Such studies indicate that eye tracking has potential for use in navigation assistance. For example, Kiefer *et al*. (2013) explored the use of eye movement data to recognize activities on maps such as searching a point-of-interest and planning a route. They trained an SVM classifier to distinguish six map-reading tasks, which resulted in an overall accuracy of 77.7%. The results indicate that distinguishing among map-reading tasks is possible.

Extending previous studies, we explore the use of machine-learning methods to infer pedestrian navigation tasks in real-world scenarios. To this end, we collected eye movement data from 38 participants when they were conducting five pre-defined tasks related to pedestrian navigation. We trained a random forest classifier to distinguish these tasks (5-task classification) based on the features extracted from the participants' eye movements. The effectiveness of the classifier was cross-validated using the collected data. Our goal is to quantify how much of the users' eye movement data can be employed to infer user tasks in real-world pedestrian navigation, and additionally to explore which types of eye movement features are most important for such task inferences. Our methods and findings can contribute to our understanding of the relationship between visual cues and navigational tasks. It is also helpful for the development of applications that employ emerging technologies such as gaze-based wayfinding tools (Giannopoulos *et al*. 2015).

The remainder of this paper is structured as follows. Section 2 reviews prior work on task inference using eye movement data. Section 3 details a real-world pedestrian navigation experiment and describes the methods used for eye movement data processing, feature extraction, classification and cross-validation. In Section 4, we present the results and compare the performances achieved across different influencing factors. A discussion is presented in Section 5. We conclude the article with suggestions for future work in Section 6.

2. Background and related work

2.1. *Eye tracking for pedestrian navigation*

Eye movements are an external manifestation of cognitive activities; eye movement data convey a wealth of information regarding internal cognitive processes (Goldberg *et al*. 2002). Based on the eye-mind hypothesis (Just and Carpenter 1976), eye tracking is widely used to investigate overt attention and visual cognition in pedestrian navigation (Spiers and Maguire

2008, Wang *et al*. 2016, Schrom-Feiertag *et al*. 2017, Liao and Dong 2017a). Over the years, many studies have employed eye tracking to evaluate user performance and the usability of cartographic interfaces (Franke and Schweikart 2017, Ohm *et al*. 2017, Liao *et al*. 2017b). However, it is only recently that eye movement data have been used as a rich data source for mining contextual information and developing adaptive interfaces for pedestrian navigation (Ohm *et al*. 2014, Giannopoulos *et al*. 2015, Viaene *et al*. 2016, Lander *et al*. 2017). In this study, we further pursue this line of research by using machine learning methods to infer user tasks in real-world navigation from eye movement data. For a review of recent eye tracking studies for spatial research, see Kiefer *et al*. (2017).

2.2. *User tasks in pedestrian navigation*

Pedestrian navigation comprises of a series of user tasks. For example, at the beginning of navigation, a user must determine where he/she is located and in which direction he/she is facing (self-localization and orientation) by matching a map and the surrounding environment. We define a *user task* as *an activity performed by a navigator or wayfinder to reach a certain goal*. A goal may be to determine where one is located or to reach a building while avoiding obstacles (e.g., locomotion, or physical movement to a nearby location; cf. Montello 2005). When a task is accomplished, it causes a change that can be either physically sensed (e.g., the user has arrived at the destination, and his/her position has changed) or mentally sensed (e.g., the user has decided in which direction to go before moving, meaning that the user simply ‘knows’ the result of his/her decision). In addition, user tasks in pedestrian navigation have a geographic context. To accomplish a task, a navigator must acquire spatial knowledge from maps and/or the environment by interacting with them. The above characteristics differentiate the user tasks in our study from those in the previous studies, as summarized in Table 1.

There is no consensus regarding the user task division of pedestrian navigation. However, previous studies serve as the initial basis for user tasks in our experimental design, as described in Section 3.1. We identified five important user tasks in navigation: self-localization and orientation (Task 1), environment target search (Task 2), map target search (Task 3), route memorization (Task 4) and walking to the destination (Task 5). These five tasks cover many of the activities within map-aided pedestrian navigation; and the order of these tasks generally follows the steps of real-world navigation. As mentioned by Downs and Stea (1977) and Delikostidis *et al*. (2015), self-localization and orientation (i.e., Task 1) is often needed at the beginning of navigation. Local environment target search (Task 2) and map target search (Task 3) are often needed in destination identification, map-environment matching, and locomotion. When starting the actual locomotion, wayfinders often memorize the planned route (Task 4), although they can look at the map at any time during navigation. Task 5 (walking to the destination) represents the route control process (i.e., route confirmation and reorientation) through navigation. Most of Task 5 includes locomotion, or the physical movement to a nearby location (Montello 2005). These tasks involve user interactions with either the environment (Task 2), the map (Tasks 3 and 4), or both (Tasks 1 and 5). They are essential activities for navigation.

Table 1. Selected task inference experiments using eye movement data.

Reference	N	Time	Performance	Classifier	Experimental tasks
Greene *et al.* (2012)	16	10 s, 60 s	25.9% (Chance: 25.0%)	SVM	(1) Memorize the picture (2) Estimate the decade in which the picture was taken (3) Estimate how well the people in the picture know each other (4) Estimate the wealth of the people in the picture
Haji-Abolhassani and Clark (2014)	16	10 s	59.64% (Chance: 25.0%)	HMM	
Kanan *et al.* (2014)	16	60 s	37.9% (Chance: 25.0%)	SVM	
Borji and Itti (2014)	16	60 s	34.12% (Chance: 25.0%)	KNN, RUSBoost	
Borji and Itti (2014)	21	5 s	24.21% (Chance: 14.3%)	KNN, RUSBoost	The same as the 7 tasks of Yarbus *et al.* (1967): (1) Freely view the painting (2) Evaluate the wealth of the family in the painting (3) Determine the ages of the people in the painting (4) Determine what the people were doing before the arrival of the visitor (5) Memorize the clothes worn by the people (6) Memorize the positions of the people and objects in the painting (7) Evaluate how long the visitor had been away from the family
Boisvert and Bruce (2016)	19	Unknown	Aggregated: 69.6% (chance: 25%) Individual: 56.4% (binary)	Random forest	Binary classification between 4 tasks: (1) Free view (2) Object search (3) Saliency viewing (4) Selecting the most salient location
Kiefer *et al.* (2013)	17	20 s	77.7% (Chance: 16.7%)	SVM	(1) Freely explore the map (2) Global search: search for a labelled point on the map (3) Route planning: plan the shortest route from A to B (4) Local search: search for the three closest objects around A (5) Line following: follow a line from north to south and count the number of intersections (6) Polygon comparison: compare the areas of two lakes and name the larger one
Bulling *et al.* (2011)	8	5 min	Precision: 76.1% Recall: 70.5% (Chance: 16.7%)	SVM	(1) Copy a text (2) Read a printed sheet of paper (3) Write handwritten notes (4) Watch a video (5) Browse the Web (6) No specific activity (NULL)
Bulling *et al.* (2013)	4	unknown	Precision: 76.8%, recall: 85.5%	SVM	Binary classification: (1) Social (social interaction vs. no interaction) (2) Cognitive (concentrated vs. leisure) (3) Physical (physically active vs. resting) (4) Spatial (inside vs. outside)
Steil and Bulling (2015)	10	5 min	F1: 43% ~ 75%	LDA	Nine activities (not mutually exclusive): Outdoor, social interaction, focused work, travel, reading, computer work, watching media, eating and special

Notes: N: the number of participants; SVM: support vector machine; HMM: hidden Markov model; KNN: k-nearest neighbour; LDA: latent Dirichlet allocation.

2.3. *Task inference using eye movement data*

To date, very few studies have addressed real-world task inference (Bulling *et al*. 2013). Many task inference studies have been conducted in the laboratory using natural images as stimuli. A selection of these studies are summarized in Table 1. It is important to note that the concept of 'tasks' that is used in these studies is different from the one used in our study. Therefore, we use 'experimental tasks' to refer to these tasks (e.g., estimating the decade in which a picture was taken) and use 'user tasks' to refer to tasks in our study (e.g., self-localization and orientation).

There are two important factors in task inference. The first important factor is eye movement features. Greene *et al*. (2012) used summary eye movement features such as the number of fixations and the mean fixation duration to classify four tasks, but the work resulted in an accuracy of only 25.9% (chance = 25%). They concluded that eye movement information was insufficient for inferring experimental tasks. However, in a subsequent study, Borji and Itti (2014) replicated the Greene *et al*. (2012) experiment and achieved an overall accuracy of 34.12%, which was significantly above the chance level. Borji and Itti (2014) also repeated the Yarbus *et al*. (1967) experiment with an overall accuracy of 24.21% (chance = 14.3%). They additionally demonstrated that spatial fixation patterns contain diagnostic information for task inference. The importance of spatial information for task inference has been confirmed by other studies (Haji-Abolhassani and Clark 2014, Kanan *et al*. 2014, Boisvert and Bruce 2016).

Aside from the above aggregated eye movement features, saccade direction and saccade encoding features have been explored (Bulling *et al*. 2011, Kiefer *et al*. 2013). For example, in order to distinguish six office activities (namely, copying a text, reading a printed paper, taking handwritten notes, watching a video, browsing the Web or no specific task), Bulling *et al*. (2011) encoded saccades into sequential characters based on a 8-cardinal-direction scheme. The characters were then scanned by a sliding window and converted into 'micro-patterns'. Spatial and temporal changes of eye movements were represented in these encodings. Such dynamic information was considered to be more distinctive for inferring tasks in dynamic environments. However, there is little evidence indicating the relative importance of different eye movement features for task inference.

The second important factor is the choice of classifier. As shown in Table 1, many studies have employed support vector machines (SVMs), partly because SVMs are able to determine a hyperplane to maximize the margin between two classes of given samples; it is insensitive to the increase of dimensions (Wu *et al*. 2008). However, evidence shows that the performance of task inference depends on the classifier that is used. Borji and Itti (2014) found that using a Boosting classifier could achieve higher accuracy than k-nearest neighbour (KNN) and the linear SVM of Greene *et al*. (2012). In a real-life task inference study, Steil et al. (Steil and Bulling 2015) proposed using latent Dirichlet allocation (LDA) to discover users' everyday activities such as eating, reading, and travel from long-term eye movement recordings. They demonstrated the LDA outperformed SVM and naive Bayes.

Unlike natural images, maps are abstract representations of geographic areas. Map interpretation can lead to different eye movement patterns from those witnessed in relation to natural images. However, the use of eye movement data with regard to map-based pedestrian navigation in real-world scenarios has not been tested.

3. Methods

3.1. *Eye movement data collection*

3.1.1. *Participants*

A total of 44 volunteers (20 female and 24 male) with a mean age of 23 (SD = 2.5) participated in the experiment. They were unaware of the purpose of the experiment. They included bachelor, master and PhD students from various backgrounds (e.g., geography, psychology, engineering, arts, management) at the university to which the authors belong. All participants had normal or corrected-to-normal vision. None of them reported eye diseases. Each participant received ¥100 (*yuan*) as compensation for their participation. The experiment was reviewed and approved by the local institutional review board (IRB). All participants provided their written informed consent.

3.1.2. *Apparatus*

We used SMI eye tracking glasses (ETG) (Apple, USA; https://www.smivision.com) to record binocular eye movement data. The ETG had a sampling rate of 60 Hz. According to the ETG user manual, the theoretically best tracking accuracy of the glasses was 0.5°. Their horizontal and vertical tracking ranges were 80° and 60°, respectively. During the experiment, synchronous scene video and sound were also automatically recorded by the ETG. The resolution of the scene video was 1280 × 960 at 24 fps. All of the data were stored on a Thinkpad laptop that was connected to the ETG.

3.1.3. *Stimuli and procedure*

The experiment was conducted on sunny or partly cloudy days with little wind in November and December 2016. The participants were first welcomed and given a brief introduction to the experiment. They then donned the ETG, with appropriate myopic lenses if needed. Their eyes were calibrated using a 3-point calibration process. After calibration, the participants were guided to the starting point of the experiment. They were given a folder containing a set of sheets of A4 paper with instructions and maps printed on them (Figure 1).

The participants were required to complete two routes in sequence. The first route (Route 1) was located on the campus of the authors' university, and the second (Route 2) was in a residential area near the campus (Figure 2). Each route was approximately 500 m long. Route 1 was considered to be familiar to the participants, whereas Route 2 was assumed to be unfamiliar. However, their familiarity was further self-rated in a later questionnaire. Before starting each route, the participants' eyes were checked and recalibrated if necessary. On each route, they were required to perform five tasks in sequence (described below). Each task consisted of two phases: the *instruction reading phase* and the *task execution phase*. During the *instruction reading phase*, the participants read the instructions for the upcoming task and could ask questions if anything was unclear. When they were ready, they were required to say 'Begin' and turn the page to the map to enter the *task execution phase*. During this phase, the experimenter remained as silent as possible. The participants were required to say 'I have found it' immediately upon completing the task. These two phases were defined to provide clear indications of

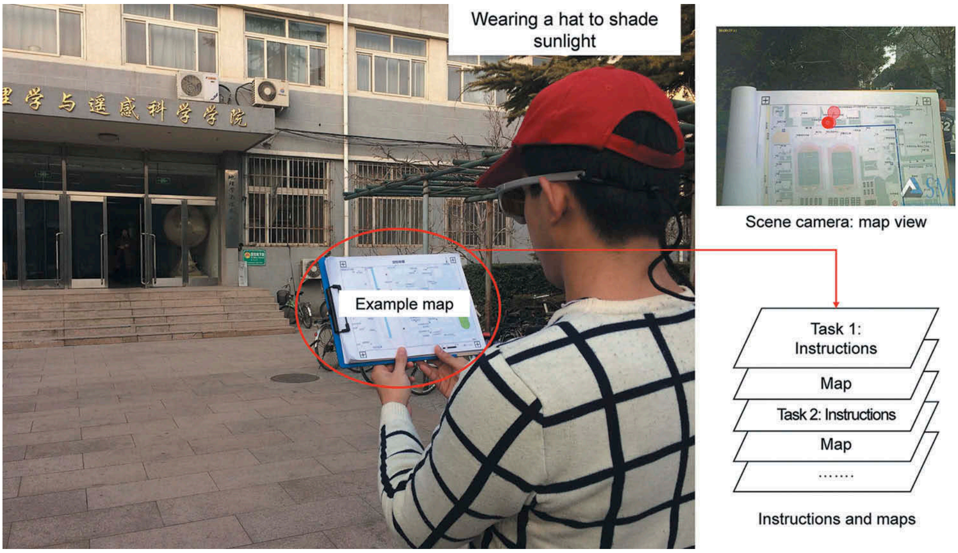


Figure 1. Illustration of the experimental set-up.

the starting and ending time points of each task for later video segmentation. The participants were told that there was no minimum or maximum time limit for any task unless otherwise stated in the instructions. The instructions for these tasks were as follows.

- **Task 1 (self-localization and orientation)**. *Please compare the map on the next page to the surrounding environment, find where you are on the map and determine which direction is north in the environment.*
- **Task 2 (local environment target search)**. *Please find object X within 50 m in the nearby environment and walk to it*. Object *X* on Route 1 was a sculpture, and on Route 2, it was a labelled sign (see Figure 2).
- **Task 3 (map target search)**. *Please search for a labelled target X on the map on the next page*. This task was performed with three targets (i.e., three trials were performed). The first target *X* was the destination of the route. The other two trials of this task were performed after Task 5.
- **Task 4 (route memorization)**. *Please memorize the route on the next page for 30 s. When the time is up, you will be informed. This is the route that you are about to navigate.*
- **Task 5 (walking to the destination)**. *Please walk from the origin to the destination along the route you have just memorized. You can look at the map if necessary during navigation.*

After finishing these tasks for both routes, the participants were required to complete a questionnaire to evaluate their sense of direction and the task load of the experiment. For these evaluations, we used the Santa Barbara Sense of Direction Scale (SBSODS) (Hegarty *et al*. 2002) and the NASA Task Load Index (TLX) (Hart and Staveland 1988), respectively. In addition, the participants were required to rate their familiarity with the

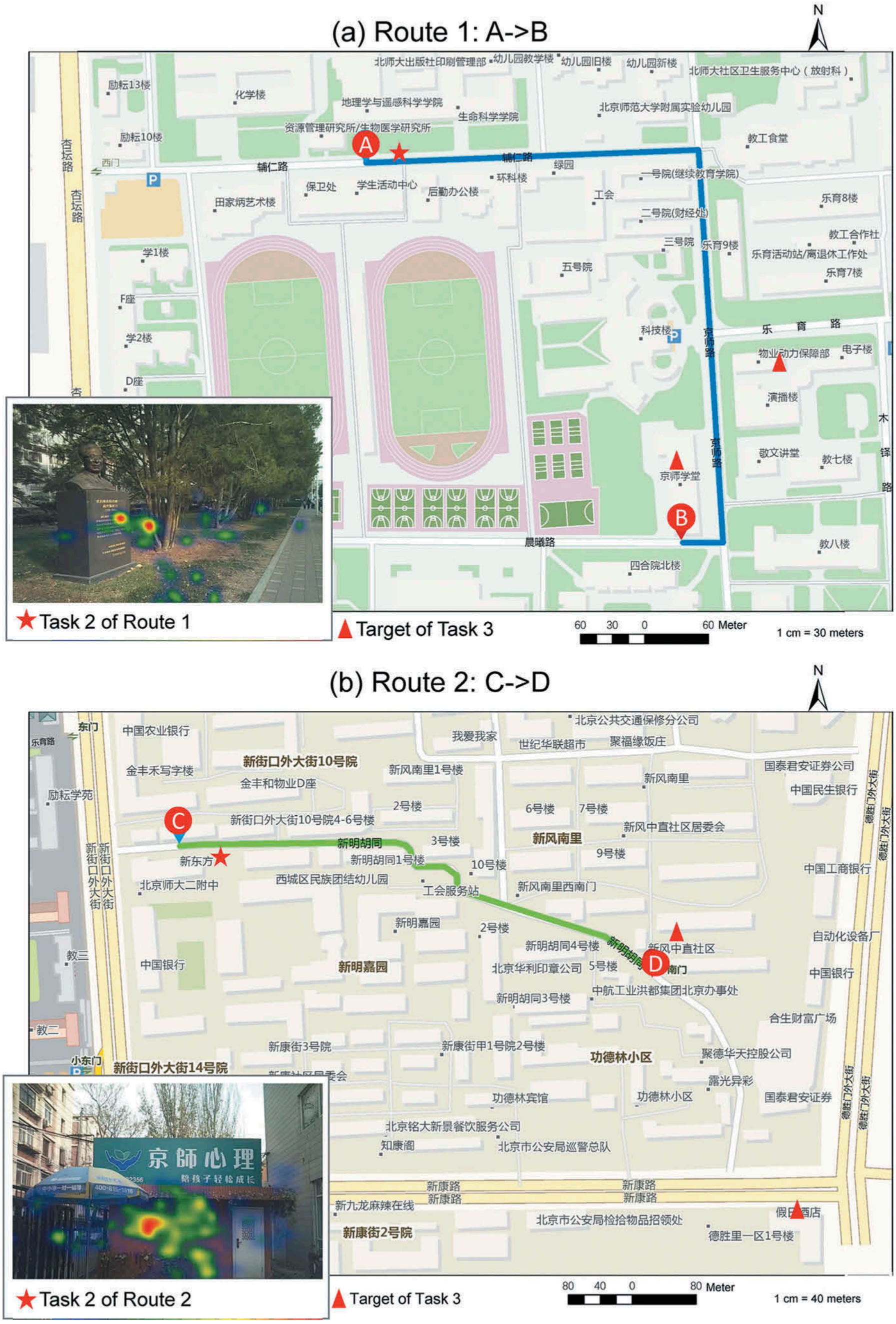


Figure 2. The two routes of the experiment.

routes on a 7-point scale (1: not familiar at all, 7: very familiar). Each participant took approximately one hour to complete the experiment.

3.2. *Data processing*

We excluded six participants due to calibration failure or recording failure. For the data of the remaining 38 participants (a total length of 13 h of data), the following four processing steps were conducted:

(1) **Fixation identification**. We used the SMI Event Detection algorithm, the default fixation filter of the SMI BeGaze v3.7 software, to identify fixations from raw gaze points. The algorithm considers the head motions of participants and thus is considered to be superior to traditional methods, such as the I-DT algorithm (Salvucci and Goldberg 2000).

(2) **Video segmentation**. We manually segmented the synchronously recorded video, sound, and eye movement data for each task. The participants' utterances 'Begin' and 'I have found it' were used as indicators of the starting and ending time points, respectively. Therefore, each resulting segment contained only the data from the participant's task execution phase. We excluded segments with tracking ratio < 80%, resulting in a mean tracking ratio of 97.4% (SD = 2.86%) for the remaining data. The tracking ratio was calculated as the number of correctly identified raw gaze points divided by the total number of attempts (i.e., 60 attempts per second in this study). The ratio was mainly affected by the lighting conditions, in which strong sunlight would decrease the tracking ratio. Example scenes are shown in Figure 3. The mean duration of each task is shown in Table 2.

(3) **Segment resampling**. We used a sliding window with a time window size (T_{win}, in [1 s, 20 s]) and a step (T_{step}, in [1 s, 10 s]) to scan each segment. We conducted an exhaustive search for combinations of T_{win} and T_{step} and found T_{step} = 4 s achieved the best performance; results across different time window sizes are presented in Section 4. Hereafter, the step size was set to 4 s. The sliding window method was also adopted by other task inference studies (Bulling *et al*. 2011, Bednarik *et al*. 2013, Steil and Bulling 2015).

(4) **Task number balancing**. As shown in Table 2, the durations of Task 3/Task 5 are much shorter/longer than the other tasks, leading to imbalanced samples (e.g., Task 5 has much more segment samples than the other tasks). We thus employed a simple method to control the imbalance: for a given time window size, we first find the task with minimum number of segments (*N*); for the remaining tasks, we select the first *N* segments only. The reason is twofold. First, earlier eye movements contain the most critical information regarding users' intentions (Borji and Itti 2014, Boisvert and Bruce 2016). Second, from a practical point of view, it is better to infer tasks as early as possible. By using this method, the number of samples for each task became equal.

These steps resulted in 1750 data samples when T_{win} = 1 s and 345 data samples when T_{win} = 20 s.

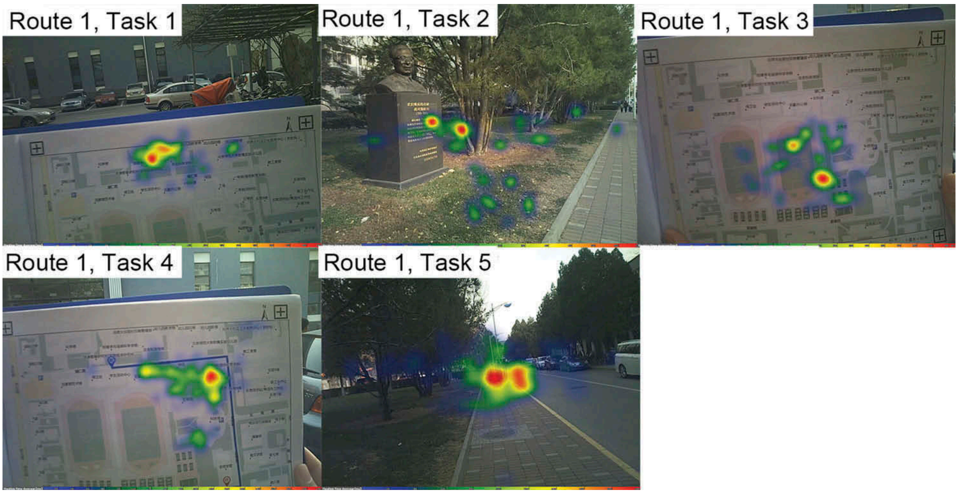


Figure 3. Example scenes from the five tasks. Note that the fixation heat maps overlaid on the scenes were created from short time durations (≈ 2 s) of data.

Table 2. Mean durations of each task.

	Route 1			Route 2		
Task	Total duration (seconds)	*M*	SD	Total duration (seconds)	*M*	SD
T1 (self-localization and orientation)	718.41	19.42	11.12	1916.21	50.43	23.81
T2 (local environment target search)	873.81	22.41	5.45	1177.33	31.82	14.18
T3 (map target search)	1209.47	11.52	6.48	1560.10	14.86	9.55
T4 (route memorization)	1304.03	33.44	7.42	1237.53	33.45	7.59
T5 (walking to destination)	14,097.33	361.47	48.15	14,659.76	396.21	59.51

3.3. *Feature extraction*

For each eye movement segment, we computed the following five types of features to quantify the statistical, spatial and temporal characteristics of the eye movements. This resulted in 568 features when T_{win} = 1 s and 649 features when T_{win} = 20 s.

3.3.1. *Basic statistical features*

Features of this type are based on basic eye movements (fixations and saccades), blinks and pupil diameters. They are widely adopted eye movement indicators in eye tracking studies (Goldberg and Kotval 1999, Ooms *et al*. 2012, Dong *et al*. 2014, Liao *et al*. 2018). For each indicator except frequency, the mean, maximum (max), minimum (min) and skewness (skew) were calculated, resulting in a total of 35 features (Table 3).

3.3.2. *Fixation density features*

To capture the spatial distributions of the fixations, we computed fixation density images using Gaussian kernel density estimation (Silverman 1986). The resulting 2D density images were then down-sampled to 20 × 20 or 1 × 400 feature vectors.

Table 3. Basic statistical features.

Type	Indicator	Unit	Statistic	N
Fixation	Fixation frequency	Count/second		1
	Fixation duration	Millisecond (ms)	Mean, max, min, skew	8
	Fixation dispersion	pixel		
Saccade	Saccade frequency	Count/second		1
	Saccade duration	Millisecond (ms)	Mean, max, min, skew	16
	Saccade amplitude	Degree (°)		
	Saccade velocity	Degree/second		
	Saccade latency	Millisecond (ms)		
Blink	Blink frequency	Count/second		1
	Blink duration	Millisecond (ms)	Mean, max, min, skew	4
Pupil	Pupil diameter	Millimetre (mm)	Mean, max, min, skew	4

3.3.3. *Saccade direction features*

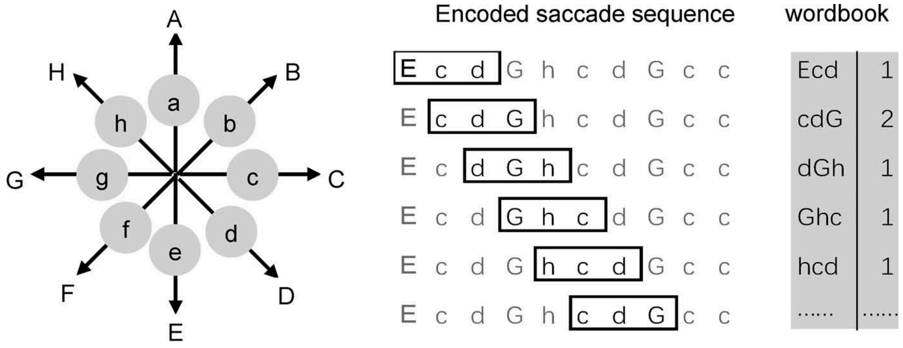
Features of this type are used to quantify the directional characteristics of eye movements, as inspired by Kiefer *et al*. (2013). They are an extension of the basic statistical features. We first divided the saccade directions using 4- and 8-cardinal-direction schemes. For each direction, we computed the number of saccades (N = 12) as well as the mean, max and min values of saccade amplitude and saccade duration (N = 72), which resulted in 84 features in total.

3.3.4. *Time slicing statistical features*

Features of this type are used to characterize temporal variations of eye movements. They are also an extension of the basic statistical features. We first sliced each data sample into smaller segments using a time bin size of two seconds (T_{bin} = 2 s). We then computed the mean, max and min values of fixation duration, fixation dispersion, and pupil diameter for each time bin. For a given time window size T_{win}, the number of features calculated was $N = 9 \times (T_{win}/T_{bin})$.

3.3.5. *Saccade encoding features*

The method used for encoding saccade sequences was adopted from Bulling *et al*. (2011) with slight modifications. We first divided the saccades into those with large amplitudes and those with small amplitudes using a threshold of 1.1 (Kiefer *et al*. 2013). Based on the 8-cardinal-direction scheme, each saccade sequence was encoded into a string of characters using 16 distinct characters (Figure 4). Upper-case characters represent saccades with large amplitudes (≥ 1.1), and lower case characters represent saccades with small amplitudes (< 1.1). The encoded string was then scanned with a sliding window of a given length (e.g., length = 3 characters) and a step size of one character. The consecutive characters within the sliding window were defined as a micro-pattern. When the sliding window reached the end of the string, all micro-patterns were identified, counted, and added to a wordbook. The wordbook size (len), the max, min and mean counts of the micro-patterns and the difference (diff) between max and min were calculated as the eye movement features for the wordbook. By varying the sliding window length from 1 to 4 characters for both the 4- and 8-cardinal-direction schemes, we calculated a total of 40 features of this type. The naming scheme for these features and the above four types of features is shown in Table 4.



(a) the 8-cardinal-direction scheme

(b) eye movement pattern analysis

Figure 4. The 8-cardinal-direction coding scheme and an example of eye movement pattern analysis, modified from Bulling *et al.* (2011).

Table 4. The naming scheme for the five feature types.

Feature type (abbr.)	Naming method and example
Basic statistical features (BS)	**Naming method**: BS—Indicator—statistic **Example**: *BS-FixationDuration-mean* (the mean value of fixation duration)
Fixation density features (FD)	**Naming method**: FD—Cell ID **Example**: *FD-Cell100* (Cell100 means the 100th cell of a fixation density image)
Saccade direction features (SD)	**Naming method**: SD—Cardinal direction scheme—Direction—Indicator—statistic **Example**: *SD-Dir4A-FixationDuration-mean* (Dir4A means the Direction A, i.e., the up direction, in a 4-cardinal-direction scheme)
Time slicing statistical features (TS)	**Naming method**: TS—time bin ID—Indicator—statistic **Example**: *TS-Bin1-FixationDuration*: Bin1 mean the first time bin, i.e., the first two seconds
Saccade encoding features (SE)	**Naming method**: SE—Cardinal direction scheme—Sliding window size—statistic **Example**: *SE-Dir8Win4-len*: Dir8Win4 means a sliding window with the length of 4 characters applied to a saccade string based on 8-cardinal-direction scheme

3.4. *Classification and cross-validation*

We adopted the random forest approach (Breiman 2001) for learning and classification. The random forest method is an ensemble machine-learning method for classification. It is based on the concept of combining many weak learners (decision trees) into a single stronger learner (Friedman *et al*. 2001). In the random forest method, a subset of data is randomly selected to grow a decision tree (Figure 5). The grown tree then serves as an independent classifier voting for a target class. This procedure is repeated a number of times. Finally, the class that earns the most votes is selected as the final prediction by the random forest classifier. The introduction of randomness allows to avoid overfitting of the training data.

The main reason for choosing the random forest method in this study is its ability to evaluate feature importance for classification. We used the open-source Python implementation from the Scikit-Learn library (https://github.com/scikit-learn/scikit-learn) (Garreta and Moncecchi 2013). The key parameter of the random forest classifier, the number of trees (*n_estimators*), was set to 500. The optimal value was found using randomized search cross-validation. The maximum number of features (*max_features*) for each tree was set to *sqrt (n_features)*, that is, the square root of the number of features.

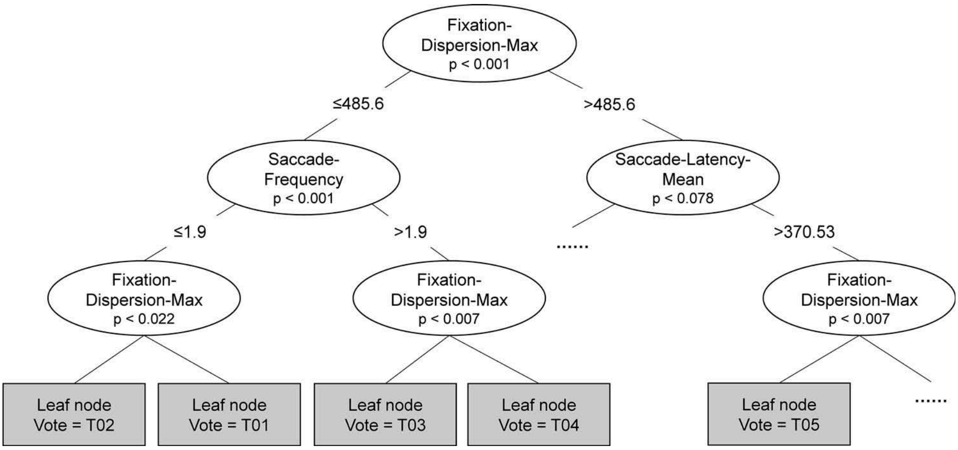


Figure 5. A sample decision tree created from a random subset of the data.

We adopted the leave-one-person-out method (Bulling *et al.* 2011) for task classification and cross-validation. A single run of the leave-one-person-out method is as follows: use the data set of all participants but one for training; then, use the data set of the excluded participant for testing; repeat this process until all participants were processed. We used accuracy (the number of predicted classes that exactly match the corresponding true classes by the total number of samples) to measure the classification performance.

We explored the influence of the following two variables (time window size and feature type) on task inference performance:

- Time window size (20 levels): we tested time window sizes varying from 1 s to 20 s.
- Feature type (6 levels): we used each of the five types of features individually for classification and then used the combination of all features (thereafter referred to as *combined features*) for classification.

For each combination of the levels of these variables, we performed a single run of classification and cross-validation, leading to 120 (=20 × 6) runs in total.

4. Results

In this section, we first examine the task inference performance across different feature types and time window sizes. We then compare the classification accuracy among tasks and explore the importance of the various features for task inference. Finally, we compare the results with cross-route validation results and with other classifiers.

4.1. *Performance across feature types and time window sizes*

Figure 6 shows the mean accuracies using different feature types (T_{win} varies from 1 s to 20 s). When combined features are used and T_{win} = 17 s, the random forest classifier achieves the highest accuracy of 67% in distinguishing among the five tasks (chance level = 1/5 = 20%). When the time window size varies from 9 s to 16 s, the accuracy

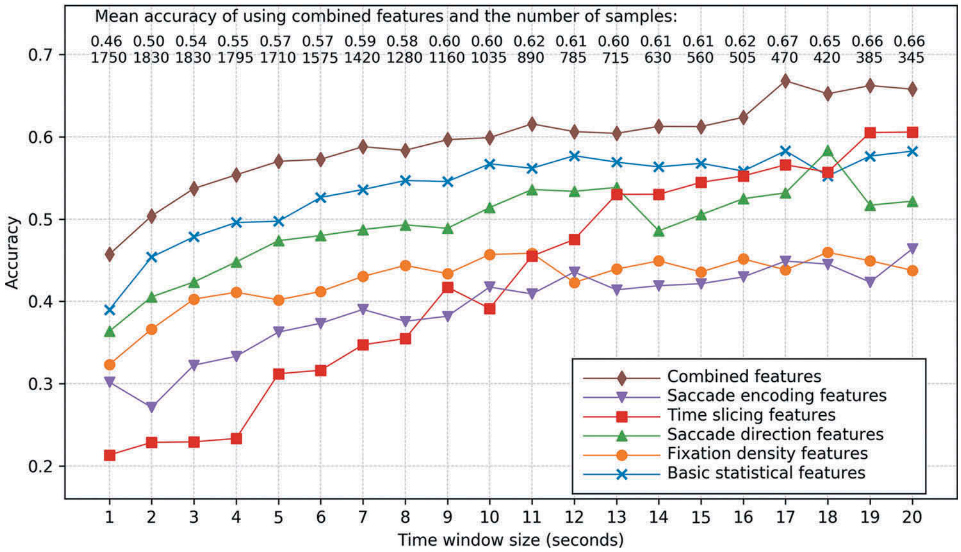


Figure 6. Classification accuracy across feature types.

ranges from 60% to 62%. Regardless of the type of features used, the classification accuracy continuously increases for the first several seconds (approximately 7–8 s) and then remains generally stable until 17 s. This finding confirms the previous evidence that the first several seconds of eye movement reveal the most important information of user tasks (Borji and Itti 2014, Boisvert and Bruce 2016).

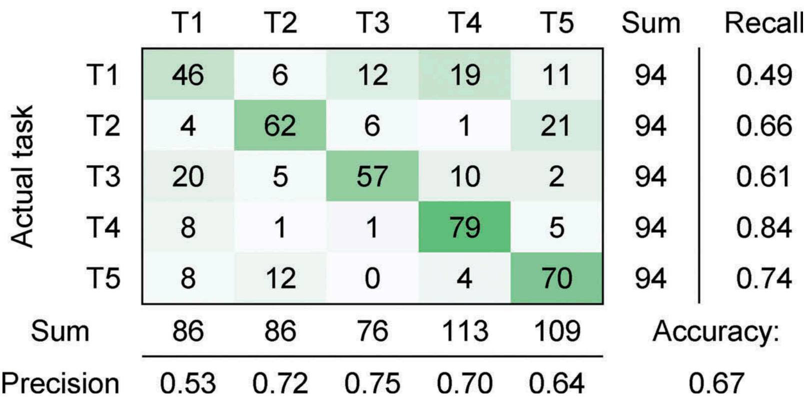
Unsurprisingly, using combined features can increase the accuracy compared to the use of any of the different feature types alone. However, this enhancement in accuracy is not simply equivalent to the sum of the accuracies achieved with the other five feature types. The accuracy achieved using basic statistical features alone, although lower than that achieved using the combined features, is higher than that for any of the other feature types. This result differs from the previous evidence that indicates the insufficiency of summary eye movement features (Greene *et al*. 2012). The importance of the statistical features is further discussed in Section 4.3.

It is interesting to note that the classification accuracy of time slicing statistical features continuously increases with the growing time window size. The accuracy of saccade direction features is lower than that of basic statistical features, but higher than that of fixation density features. The lowest accuracy is observed when the saccade encoding features are used alone.

The accuracy achieved using the fixation density features is approximately 45%, probably because the spatial distribution for a given task mainly depends on the spatial structure of the environment and the map. Another important reason is that in real-world environments, users can move their heads freely to centre their visual targets in their visual fields. Such free head motion makes the spatial distribution information less distinct.

4.2. *Performance across tasks*

We examine the confusion matrix to assess the classifier performance for each task individually (Figure 7). The results show that with a time window size of 17 s, correct



		T1	T2	T3	T4	T5	Sum	Recall
Actual task	T1	46	6	12	19	11	94	0.49
	T2	4	62	6	1	21	94	0.66
	T3	20	5	57	10	2	94	0.61
	T4	8	1	1	79	5	94	0.84
	T5	8	12	0	4	70	94	0.74
Sum		86	86	76	113	109	Accuracy:	
Precision		0.53	0.72	0.75	0.70	0.64	0.67	

Figure 7. Summed confusion matrix when using combined features (Time window size = 17 s).

classification is achieved for 314 out of 470 samples, corresponding to an overall accuracy of 67%. The recall is between 49% and 84%, whereas the precision is between 53% and 75%. Task 4 (route memorization) shows the highest recall (84%) and a relatively high precision (70%), followed by Task 5 (walking to the destination), with 74% recall and 64% precision.

The worst performance is observed for Task 1 (self-localization and orientation), with 49% recall and 53% precision. For this task, some samples are misclassified as map-related tasks. Specifically, 19 of the 94 samples (20.2%) are misclassified as Task 4 (route memorization), and 12 of the 94 samples (12.8%) are misclassified as Task 3 (map target search). We speculate that this result is partly because locating and orientating oneself requires searching for targets on a map.

4.3. *Feature importance*

Examining the relative importance of different features can reveal deeper insights into task relatedness and provide explanations for the observed task inference performance. Figure 8 shows the top 15 most important features as identified by the classifier. Statistical eye movement features and saccade encoding features are observed to be crucial for task inference. We applied one-way ANOVA and pairwise tests to six of the statistical features (fixation- and saccade-based) to examine the significance of their differences between different tasks. The results are shown in Figure 9.

Task 2 (local environment target search) is associated with significantly higher fixation dispersion. Fixation dispersion is the spatial spread of the fixation points (Salvucci and Goldberg 2000). It is calculated using the formula $D = [\max(X) - \min(X)] + [\max(Y) - \min(Y)]$, where X and Y are the x and y coordinates, respectively, of the gaze points (Komogortsev *et al.* 2010). Since dispersion is usually used as a parameter in fixation identification algorithms such as the I-DT algorithm (Salvucci and Goldberg 2000), its cognitive meaning has rarely been discussed.

Task 2 is also associated with significantly higher saccade durations than the other tasks. The saccade duration, namely, the amount of time required to move the eyes from one fixation to another, depends on the distance moved (Rayner 2009). Thus, it is

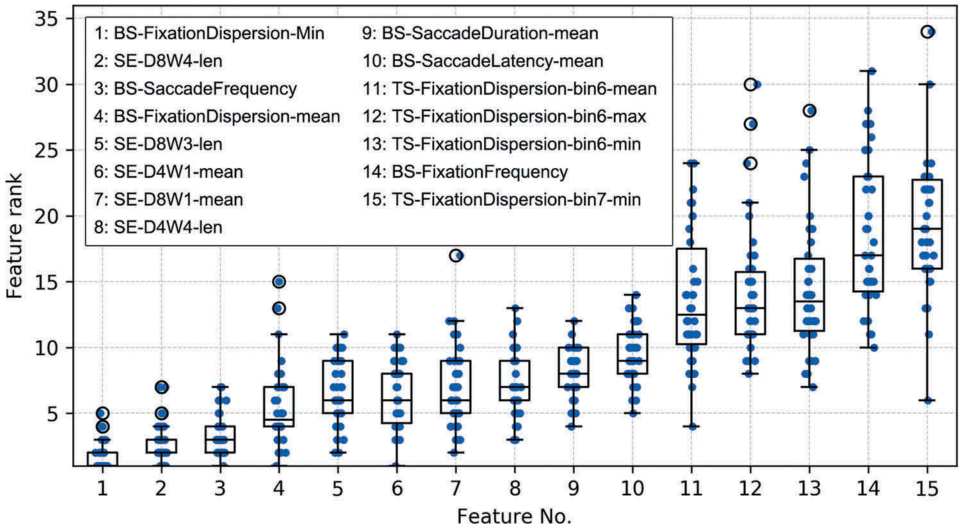


Figure 8. The top 15 most important features as identified by the classifier (Time window size = 17 s). The X-axis shows the feature number; the corresponding feature names are shown in the box in the upper left. The Y-axis shows the ranks of the features across participants. For each feature, each dot represents its rank for one participant. Refer to Table 4 for the names of the features.

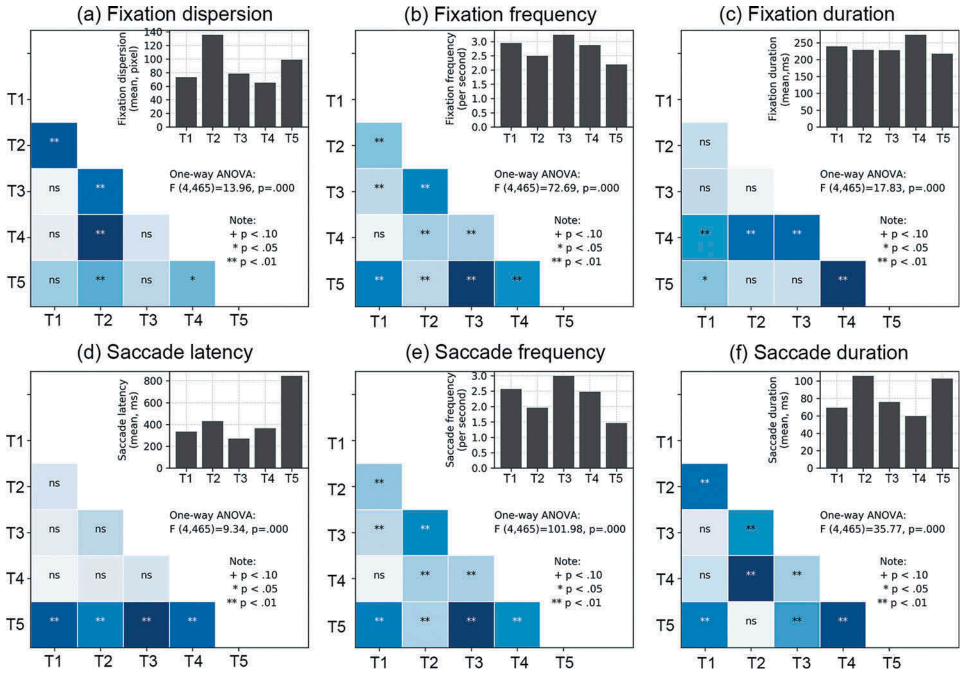


Figure 9. Results of one-way ANOVA and pairwise tests for six fixation- and saccade-based features (Time window size = 17 s).

reasonable that moving the eyes to fixate on different points in the environment (Tasks 2 and 5) takes longer than moving the eyes around a map.

Task 3 (map target search) is associated with significantly higher fixation frequencies and higher saccade frequencies than all other tasks; this finding indicates a high information processing efficiency and a high visual search efficiency when searching for labelled points on a map. Compared with searching for an object in the local environment (Task 2), which requires large body and head motions, searching for a labelled target on a map requires users to move their eyes only within a small area.

Task 4 (route memorization) caused the participants to focus their attention on the detailed information of a route, leading to significantly higher fixation durations than in the other tasks. Although previous studies have suggested that longer fixation durations indicate either higher difficulty of the stimuli or a higher level of the participants' own interest in the stimuli (Goldberg and Kotval 1999); the case reported here does not seem to belong to either of these scenarios. These results indicate that memorization intentions can also lead to long fixation durations.

Task 5 (walking to the destination) is associated with significantly higher saccade latency than any of the other tasks. Saccade latency is the amount of time between the end of one saccade and the start of the next saccade. It is the time that is needed to encode the information of the current visual target and to determine the next location on which to fixate (Rayner 2009). We find that a higher saccade latency is associated with environment-related tasks (Tasks 2 and 5), probably because the constantly changing environment presents the viewer with a larger amount of information to be interpreted. More time is required to initiate an eye movement from the current visual target to the next one, which is consistent with the observation that Task 5 has a significantly lower fixation frequency.

4.4. *Comparison with cross-route validation results*

There are two equally important types of generalizability for task inference studies such as the work reported here: the generalizability for inferring tasks from new people and from new environments. In this study, we recruited 38 participants and adopted the leave-one-person-out method for cross-validation. This method allowed us to test the robustness of the classification over different participants. However, the number of scenes for each participant was small because the participants were required to complete only two routes. It is more difficult to increase the number of routes in a real-world experiment than in a lab-based environment. We thus conducted a cross-route validation (i.e., using one route to train and the other route to test and vice versa) to test the generalizability of the proposed method over different environments. Hereafter, we use 'Route 1 -> Route 2′ to represent the process of using data of Route 1 to train the classifier and using data of Route 2 to test; 'Route 2 -> Route 1′ represents the opposite process. The results are shown in Figure 10. We observed two findings that merit further discussion.

The first finding is that the accuracy of Route 2 -> Route 1 (≈ 55%) was higher than that of Route 1 -> Route 2 (≈ 50%). We speculate that the higher accuracy of Route 2 -> Route 1 was due to the influence of familiarity. In unfamiliar environments, the users' eye movement behaviour is more explicit than that in familiar environments. For example, in familiar environments, the users did not need to match the environment to the map to

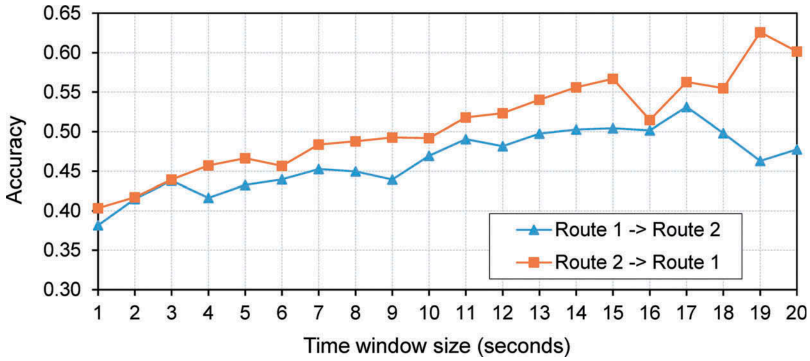


Figure 10. Accuracy of cross-route validation.

locate and orient themselves, making eye movements less distinctive than in unfamiliar environments. In our experiment, the mean score of self-report familiarity of Route 1 was 6.42 (SD = 1.25), whereas the mean score of Route 2 was 1.55 (SD = 1.21). Therefore, eye movements of Route 2 were more informative than those of Route 1, making Route 2 -> Route 1 more effective than Route 1 -> Route 2.

The second finding is that the accuracy of cross-route validation (50% to 63% when $T_{win} > 9$ s; Figure 10) is lower than the accuracy of using the leave-one-person-out method (60% to 67% when $T_{win} > 9$ s; Figure 6). This finding implies that the classifier might have learned the characteristics of the two routes when the leave-one-person-out method is used. More routes are needed to further confirm the environment influence.

4.5. *Comparison with other classifiers*

We compared random forest with seven other commonly used classifiers such as Gradient Boosting, SVM and Naïve Bayes (Figure 11). The classification was performed using the leave-one-person-out method. Optimal parameter values for Gradient Boosting and SVM were found using randomized search cross-validation. As seen from the figure, Gradient Boosting could achieve competitive performance compared to the random forest classifier; both could achieve a significantly higher accuracy across all time window sizes than the others. The accuracies of Gradient Boosting and the random forest were approximately 60%, whereas the accuracies of the other classifiers were below 50%.

5. Discussion

5.1. *Comparison with previous studies*

As presented in the results, the best overall accuracy achieved in this study is 67%, which is comparable to accuracies that have been reported in previous studies (as shown in Table 1). The performance of map-reading tasks are similar to those of Kiefer *et al.* (2013). For example, the performance of Task 3 (map target search; precision: 75%, recall: 61%) is competitive with the performance of Kiefer *et al.* (2013)'s focused search task (recall: 63.5%, precision: 76%), and the performance of Task 4 (route memorization; precision: 70%, recall: 84%) is similar to their line-following task (precision: 76%, recall: 75.3%).

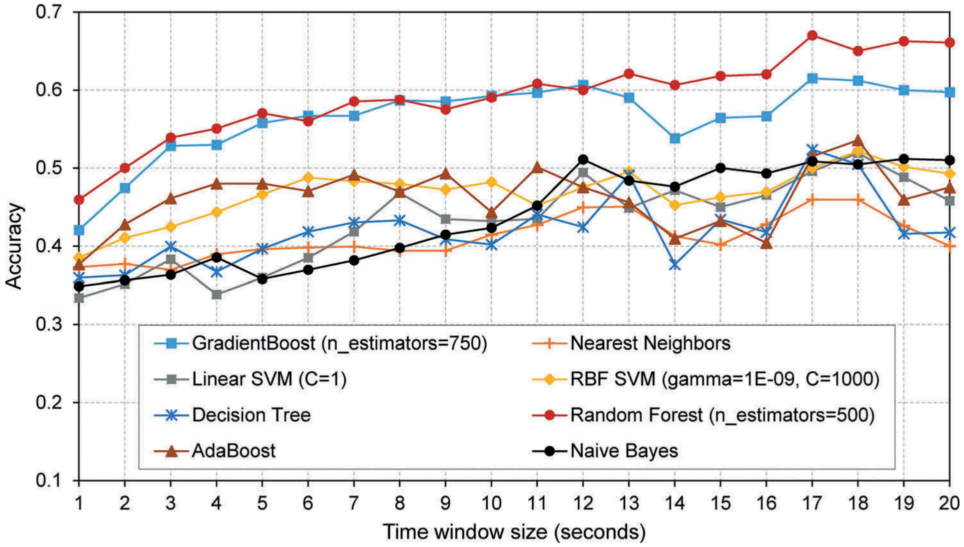


Figure 11. Comparison of classification accuracy across eight commonly used classifiers. The combined features and leave-one-person-out methods were used. Key parameters of the classifiers are indicated in the brackets.

With regard to tasks in real-world scenarios, our overall accuracy (67%) is slightly lower than that of Bulling *et al.* (2011)'s office activity detection (precision: 76.1%, recall: 70.5%) and is lower than Bulling *et al.* (2013)'s results of high-level contextual cue recognition (precision: 76.8%, recall: 85.5%). Note that Bulling *et al.* (2013) is a binary classification, such as social interaction versus no interaction.

5.2. *Implications of the current results*

Prior research has explored utilizing users' visual inputs to mine user context information and modulate human-computer interactions for navigation (Giannopoulos *et al.* 2012, Gkonos *et al.* 2017). For instance, Giannopoulos *et al.* (2015) developed a mobile navigation system called 'GazeNav' that combines eye tracking and vibration. When a user walks to a street junction and looks at a left or right street to decide which direction to go, the system will start vibrating if the direction that the user is gazing at is correct. An advantage of the system is that the user does not need to switch attention between the real-world environment and the mobile device. In addition, the system can make the user's hands free. The authors demonstrated that the GazeNav system could improve user experience compared to traditional map-based turn-by-turn instructions.

Extending prior research, this work provides empirical results that indicate the relationship between eye movements and user tasks in real-world pedestrian navigation. Our results open some potential avenues for the development of navigation assistants that can provide task-related information depending on the task a user is performing. Such gaze-aware assistants can arguably provide a more tailored and active experience for pedestrians than traditional navigation tools such as Google Maps. For example, when using a traditional navigation tool, users can only passively follow the steps and functions that are pre-defined by the tools. In contrast, a gaze-aware navigation assistant can actively detect a user's task (e.g., memorizing a route) and provide related information (e.g., highlighting the route and related landmarks to facilitate the process of memorizing). Moreover, it is possible to combine mobile eye tracking and other sensing technologies such as GPS and head movements to improve the task inference accuracy (Doshi and Trivedi 2009).

5.3. *Important influencing factors*

Aside from the above-mentioned implications for navigation applications, this work also provides empirical results related to the following three important influencing factors, which are helpful for future investigations regarding real-world task inference.

The first factor is the choice of classifier. As indicated by the results in Section 5.2, the classifier significantly affects the task inference performance. The results show that ensemble classifiers such as Gradient Boosting and random forest tend to achieve higher accuracy than other types of classifiers for the task inference in our case. It is important to note that temporal dependencies of eye movements (e.g., a fixation is affected by the previous fixated objects and locations) are important in visual attention modelling. Hidden Markov models and conditional random fields are commonly used to model such temporal dependencies and can improve the task inference performance (Haji-Abolhassani and Clark 2014). However, this study assumes that eye movements can be modelled independently over time, as many other task inference studies do.

The second factor is the time window size. Previous studies have used a fixed time duration (e.g., 20 s; see Table 1) during which participants might repeat their visual behaviour. In a real-world scenario, it is meaningless if a navigation assistant infers a user's task after the user has accomplished it. Our results show that the classification accuracy continuously increases during the first several seconds but stops increasing at later times. The minimum suitable length of the time window is 7 to 8 s. Given that a shorter time window size can save considerable computation capacity without sacrificing much accuracy, a shorter time window size is preferable in practice.

The third factor is the types of eye movement features that are considered. The results reveal differences in the power of feature types to differentiate user tasks. Statistical eye movement features and saccade encoding features (micro-patterns) are more useful than the other feature types investigated here. The results presented in Section 4.3 also indicate that the relative importance of different features depends on the specific task. It should be noted that we considered only pure eye movement features in this study. We ignored visual cues from the map and the environment, which not only provide useful information for task inference but also may be beneficial for understanding users' cognitive processes in pedestrian navigation. In future studies, computer vision methods can be adopted to identify the contents of fixations from the map and the environment (Anagnostopoulos *et al*. 2017).

5.4. *Learning effects*

It is worth noting that learning effects existed in the experiment. For example, two out of three trials of map search tasks (Task 4) were conducted after the navigation task, during which participants became familiar with the environment and the map. This approach is deliberate because our purpose is to address real-world task inference. It is possible that a user needs to search other map objects during navigation. We did not constrain participants and tried to recreate real navigation conditions.

To further examine if learning effects influenced the participants' performance significantly, we performed a t-test on the trial duration of Task 4 between the first trial and the last two trials. The results are shown in Figure 12. We assumed that the participants could perform map search tasks more quickly after the navigation task than they did at the beginning of the experiment. We chose trial duration (i.e., response time) because response time is a widely used indicator to measure participants' performance before and after map learning (Fabrikant *et al*. 2010). It is seen that on Route 1, the mean duration of the first trial (M = 9.73 s, SD = 3.89 s) was similar to that of the last two trials (M = 10.06 s, SD = 4.91 s); the difference was not statistically significant, $t = -0.327$, p = 0.745 > 0.05. On Route 2, the mean duration of the first trial (M = 16.57 s, SD = 12.30 s) was longer than that of the last two trials (M = 13.03 s, SD = 5.71 s); the difference was not statistically significant, $t = 1.609$, $p = 0.111 > 0.05$. These results indicate that although learning effects existed in the experiment, its influence on participants' efficiency of map search tasks was not statistically significant. We argue that it is unnecessary to fully exclude learning effects in such a real-world experiment.

5.5. *Limitations*

Two additional issues should be mentioned. The first issue is navigational task division. In this study, we selected the five tasks that were based mainly on Delikostidis *et al.* (2015). Such task division is not comprehensive. We did not subdivide Task 5 (walking to the destination) into smaller tasks (subtasks) such as locomotion, route confirmation, reorientation, decision-making at turning points, and destination recognition. Given that the maximum time window size is 20 s, which is much shorter than the actual length of Task 5, the task tested here mainly contains the early part of locomotion; other subtasks in the latter part of the navigation have been ignored.

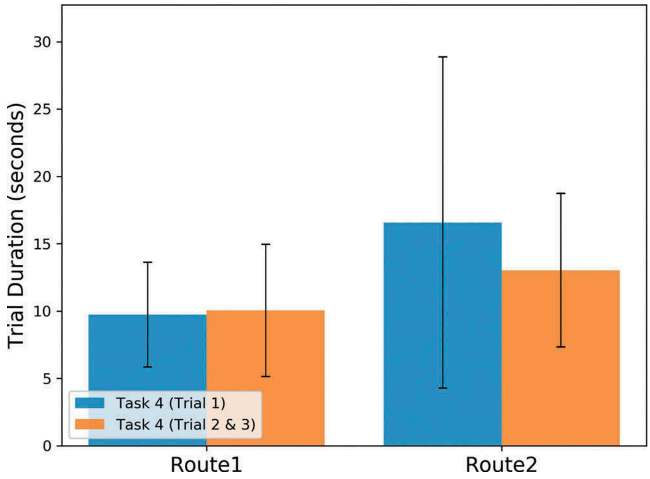


Figure 12. Comparison of mean durations between the first trial and the last two trials in Task 4.

There are at least two challenges in task division. The first challenge is how to determine the exact starting and ending time points of such subtasks without interfering with the participants during navigation because pedestrian navigation is an integrated and continuous process. One possible solution is to use concurrent thinking aloud, such that they verbally report what they are doing during navigation, but the risk of distracting the participants' visual attention should be carefully considered. The second challenge is how to establish a task hierarchy. We speculate that the performance of task inference is strongly related to the multi-level nature of the user's tasks of interest. For example, a self-localization task (Task 1) is composed of map searching, environment searching, and map-environment matching subtasks. Therefore, many trials of Task 1 were misclassified as Task 3 (map target search). It is important to note that tasks may overlap with each other, which is particularly common in real-world scenarios (Bulling *et al*. 2013, Steil and Bulling 2015). On one hand, such real-world task inference requires a sophisticated hierarchical task model; on the other hand, current task inference results, in return, can provide insights into task relatedness for task modelling. Future work could address new methods of modelling user tasks at different hierarchical levels.

The second issue is that in real-world environments, participants can be disturbed by environmental dynamics such as changes in traffic and other pedestrians on the street. The presence of such distractors in real-world environments cannot be avoided. Moreover, such environmental dynamics produce noise in eye movements, partially contributing to the lower performance in this real-world study than previous laboratory-based studies.

6. Conclusions and future work

This study explored the use of eye movement data to infer user tasks in real-world pedestrian navigation scenarios. We established a random forest-based model to classify five navigational tasks using five types of eye movement features. The results show that the overall accuracy exceeded 60% when the time window was greater than 8 s. The classifier achieved the best overall accuracy of 67% at 17 s. We found that statistical eye movement features and saccade encoding features are more important than the other investigated types of features for distinguishing tasks. We also found that there are significant differences in summary eye movement statistics among the five tested tasks. Based on these results, we conclude that it is possible to infer user tasks in real-world pedestrian navigation scenarios using pure eye movement data. Such task inference potentially enables real-world applications such as navigation assistance that can provide relevant information and guidance according to the task a user is currently performing.

As discussed in the previous section, future studies can combine gaze behaviour with other sources of information including multi-sensory information (e.g., GPS and head movements) and real-world objects identified from the map and the environment. Moreover, future studies can employ attention-modelling methods such as HMM to model temporal dependencies of eye movements into task inference.

Acknowledgments

The authors would like to thank all the reviewers for their helpful comments and suggestions.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This research is supported by National Key Research and Development Program of China [Grant No. 2017YFB0503602], National Natural Science Foundation of China [NSFC, Grant No. 41471382], Beijing Normal University Teaching Construction and Reform Project [Grant No. 14-07-01], and China Scholarship Council [Grant No. 201606040097].

ORCID

Hua Liao http://orcid.org/0000-0002-6304-329X
Weihua Dong http://orcid.org/0000-0001-6097-7946
Haosheng Huang http://orcid.org/0000-0001-8399-3607

References

Anagnostopoulos, V., *et al.*, 2017. Gaze-Informed location-based services. *International Journal of Geographical Information Science*, 1–28. doi:10.1080/13658816.2017.1334896

Bednarik, R., *et al.*, 2012. What do you want to do next: a novel approach for intent prediction in gaze-based interaction. *In*: Stephen N. Spencer, ed. *Proceedings of the symposium on eye tracking research and applications*. Santa Barbara, CA, 83–90.

Bednarik, R., *et al.*, 2013. A computational approach for prediction of problem-solving behavior using support vector machines and eye-tracking data. *In*: Y.I. Nakano, *et al.*, eds. *Eye Gaze in Intelligent User Interfaces: gaze-based Analyses, Models and Applications*. London: Springer-Verlag London, 111–134.

Boisvert, J.F.G. and Bruce, N.D.B., 2016. Predicting task from eye movements: on the importance of spatial distribution, dynamics, and image features. *Neurocomputing*, 207, 653–668. doi:10.1016/j.neucom.2016.05.047

Borji, A. and Itti, L., 2014. Defending Yarbus: eye movements reveal observers' task. *Journal of Vision*, 14 (3), 1–22. doi:10.1167/14.3.29

Breiman, L., 2001. Random forests. *Machine Learning*, 45 (1), 5–32. doi:10.1023/A:1010933404324

Bulling, A., *et al.*, 2011. Eye movement analysis for activity recognition using electrooculography. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 33 (4), 741–753. doi:10.1109/TPAMI.2010.86

Bulling, A., *et al.*, 2013. Eyecontext: recognition of high-level contextual cues from human visual behaviour. *In*: Susanne Bødker, *et al.*, eds. *Proceedings of the sigchi conference on human factors in computing systems*. Paris, France, 305–308.

Delikostidis, I., *et al.*, 2015. Overcoming challenges in developing more usable pedestrian navigation systems. *Cartography and Geographic Information Science*, 43 (3), 189–207. doi:10.1080/15230406.2015.1031180

Dong, W., *et al.*, 2014. Eye tracking to explore the potential of enhanced imagery basemaps in web mapping. *The Cartographic Journal*, 51 (4), 313–329. doi:10.1179/1743277413Y.0000000071

Doshi, A. and Trivedi, M.M., 2009. On the roles of eye gaze and head dynamics in predicting driver's intent to change lanes. *IEEE Transactions on Intelligent Transportation Systems*, 10 (3), 453–462. doi:10.1109/TITS.2009.2026675

Downs, R.M. and Stea, D., 1977. *Maps in minds: reflections on cognitive mapping*. New York, NY: HarperCollins Publishers.

Fabrikant, S.I., *et al.*, 2010. Cognitively inspired and perceptually salient graphic displays for efficient spatial inference making. *Annals of the Association of American Geographers*, 100 (1), 13–29. doi:10.1080/00045600903362378

Franke, C. and Schweikart, J., 2017. Mental representation of landmarks on maps – investigating cartographic visualization methods with eye tracking technology. *Spatial Cognition & Computation*, 17 (1–2), 20–38. doi:10.1080/13875868.2016.1219912

Friedman, J., *et al*., 2001. *The elements of statistical learning*. New York, NY: Springer series in statistics.

Garreta, R. and Moncecchi, G., 2013. *Learning scikit-learn: machine learning in python*. Birmingham, UK.: Packt Publishing Ltd.

Giannopoulos, I., *et al*., 2012. GeoGazemarks: providing gaze history for the orientation on small display maps. *In:Proceedings of the 14th ACM international conference on Multimodal interaction*. SantaMonica, California., 165–172.

Giannopoulos, I., *et al*., 2015. GazeNav: gaze-based pedestrian navigation. *In*: *Proceedings of the 17th International Conference on Human-Computer Interaction with Mobile Devices and Services*. Copenhagen, New York: ACM, 337–346.

Gkonos, C., *et al*., 2017. Maps, vibration or gaze? Comparison of novel navigation assistance in indoor and outdoor environments. *Journal of Location Based Services*, 1–21. doi:10.1080/17489725.2017.1323125

Goldberg, J.H., *et al*., 2002. Eye tracking in web search tasks: design implications. *In*: *Proceedings of the 2002 symposium on Eye tracking research & applications*. New Orleans, LA, 51–58.

Goldberg, J.H. and Kotval, X.P., 1999. Computer interface evaluation using eye movements: methods and constructs. *International Journal of Industrial Ergonomics*, 24 (6), 631–645. doi:10.1016/S0169-8141(98)00068-7

Golledge, R.G., 1999. Human wayfinding and cognitive maps. *In*: R.G. Golledge, ed. *Wayfinding behavior: cognitive mapping and other spatial processes*. Baltimore, MD: The Johns Hopkins University Press, 5–45.

Gong, L., *et al*., 2015. Inferring trip purposes and uncovering travel patterns from taxi trajectory data. *Cartography and Geographic Information Science*, 43 (2), 103–114. doi:10.1080/15230406.2015.1014424

Greene, M.R., *et al*., 2012. Reconsidering Yarbus: A failure to predict observers' task from eye movement patterns. *Vision Research*, 62, 1–8. doi:10.1016/j.visres.2012.03.019

Haji-Abolhassani, A. and Clark, J.J., 2014. An inverse Yarbus process: predicting observers' task from eye movement patterns. *Vision Research*, 103, 127–142. doi:10.1016/j.visres.2014.08.014

Hart, S.G. and Staveland, L.E., 1988. Development of NASA-TLX (Task Load Index): results of empirical and theoretical research. *Advances in Psychology*, 52, 139–183.

Hegarty, M., *et al*., 2002. Development of a self-report measure of environmental spatial ability. *Intelligence*, 30 (5), 425–447. doi:10.1016/S0160-2896(02)00116-2

Henderson, J.M., 2003. Human gaze control during real-world scene perception. *Trends in Cognitive Sciences*, 7 (11), 498–504. doi:10.1016/j.tics.2003.09.006

Huang, H., *et al*., 2014. AffectRoute–considering people's affective responses to environments for enhancing route-planning services. *International Journal of Geographical Information Science*, 28 (12), 2456–2473. doi:10.1080/13658816.2014.931585

Huang, H. and Gartner, G., 2012. Collective intelligence-based route recommendation for assisting pedestrian wayfinding in the era of Web 2.0. *Journal of Location Based Services*, 6 (1), 1–21. doi:10.1080/17489725.2011.625302

Jiang, B. and Yao, X., 2006. Location-based services and GIS in perspective. *Computers, Environment and Urban Systems*, 30 (6), 712–725. doi:10.1016/j.compenvurbsys.2006.02.003

Just, M.A. and Carpenter, P.A., 1976. Eye fixations and cognitive processes. *Cognitive Psychology*, 8 (4), 441–480. doi:10.1016/0010-0285(76)90015-3

Kanan, C., *et al*., 2014. Predicting an observer's task using multi-fixation pattern analysis. *In*: *Proceedings of the symposium on eye tracking research and applications*. Safety Harbor, FL, 287–290.

Kiefer, P., *et al*., 2013. Using eye movements to recognize activities on cartographic maps. *In*: *Proceedings of the 21st ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*. Orlando, FL, 478–481.

Kiefer, P., *et al*., 2017. Eye tracking for spatial research: cognition, computation, challenges. *Spatial Cognition & Computation*, 17 (1–2), 1–19. doi:10.1080/13875868.2016.1254634

Komogortsev, O.V., *et al*., 2010. Standardization of automated analyses of oculomotor fixation and saccadic behaviors. *IEEE Transactions on Biomedical Engineering*, 57 (11), 2635–2645. doi:10.1109/TBME.2010.2057429

Lander, C., *et al*., 2017. Inferring landmarks for pedestrian navigation from mobile eye-tracking data and Google street view. *In*: *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems*. Denver, CO, 2721–2729.

Liao, H., *et al*., 2017b. Exploring differences of visual attention in pedestrian navigation when using 2D maps and 3D geo-browsers. *Cartography and Geographic Information Science*, 44 (6), 474–490. doi:10.1080/15230406.2016.1174886

Liao, H., *et al*., 2018. Measuring the influence of map label density on perceived complexity: a user study using eye tracking. *Cartography and Geographic Information Science*, 1–19. doi:10.1080/15230406.2018.1434016

Liao, H. and Dong, W., 2017a. An exploratory study investigating gender effects on using 3D maps for spatial orientation in wayfinding. *ISPRS International Journal of Geo-Information*, 6 (3), 1–19. doi:10.3390/ijgi6030060

Liu, B., *et al*., 2013. Learning geographical preferences for point-of-interest recommendation. *In*: *Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining*. Chicago, Illinois, 1043–1051.

Montello, D. and Raubal, M., 2012. Functions and applications of spatial cognition. *In*: D.W.A.L. Nadel, ed. *Handbook of Spatial Cognition*. Washington, DC: APA, 249–264.

Montello, D.R., 2005. Navigation. *In*: P. Shah and A. Miyake, eds. *The Cambridge handbook of visuospatial thinking*. New York, NY: Cambridge University Press, 257–294.

Ohm, C., *et al*., 2014. Where is the landmark? Eye tracking studies in large-scale indoor environments. *In*: Peter Kiefer *et al*., eds. *Proceedings of the 2nd International Workshop on Eye Tracking for Spatial Research (in conjunction with GIScience 2014)*. Vienna, Austria, 47–51.

Ohm, C., *et al*., 2017. Evaluating indoor pedestrian navigation interfaces using mobile eye tracking. *Spatial Cognition & Computation*, 17 (1–2), 89–120. doi:10.1080/13875868.2016.1219913

Ooms, K., *et al*., 2012. Investigating the effectiveness of an efficient label placement method using eye movement data. *The Cartographic Journal*, 49 (3), 234–246. doi:10.1179/1743277412Y.0000000010

Rayner, K., 2009. Eye movements and attention in reading, scene perception, and visual search. *The Quarterly Journal of Experimental Psychology*, 62 (8), 1457–1506.

Salvucci, D.D. and Goldberg, J.H., 2000. Identifying fixations and saccades in eye-tracking protocols. *In*: *Proceedings of the 2000 symposium on Eye tracking research & applications*. 71–78.

Schrom-Feiertag, H., *et al*., 2017. Evaluation of indoor guidance systems using eye tracking in an immersive virtual environment. *Spatial Cognition & Computation*, 17 (1–2), 163–183. doi:10.1080/13875868.2016.1228654

Silverman, B.W., 1986. *Density estimation for statistics and data analysis*. Florida: CRC press.

Spiers, H.J. and Maguire, E.A., 2008. The dynamic nature of cognition during wayfinding. *Journal of Environmental Psychology*, 28 (3), 232–249. doi:10.1016/j.jenvp.2008.02.006

Steil, J. and Bulling, A., 2015. Discovery of everyday human activities from long-term visual behaviour using topic models. *In*: *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing*. Osaka, Japan., 75–85.

Viaene, P., *et al*., 2016. Examining the validity of the total dwell time of eye fixations to identify landmarks in a building. *Journal of Eye Movement Research*, 9 (3), 1–11.

Wang, S., *et al*., 2016. Visualizing the intellectual structure of eye movement research in cartography. *ISPRS International Journal of Geo-Information*, 5 (10), 1–22. doi:10.3390/ijgi5100168

Wolfe, J.M., *et al*., 2003. Changing your mind: on the contributions of top-down and bottom-up guidance in visual search for feature singletons. *Journal of Experimental Psychology: Human Perception and Performance*, 29 (2), 483–502.

Wu, X., *et al*., 2008. Top 10 algorithms in data mining. *Knowledge & Information Systems*, 14 (1), 1–37. doi:10.1007/s10115-007-0114-2

Yarbus, A.L., *et al*., 1967. *Eye movements and vision*. New York, NY: Plenum press.