Conference Paper

**Identifying User Tasks in Map Based-Pedestrian Navigation from Eye Tracking Data**

**Author(s):**
Liao, Hua; Dong, Weihua; Gartner, Georg; Liu, Huiping

**Publication Date:**
2018-01-15

**Permanent Link:**
https://doi.org/10.3929/ethz-b-000225606

**Rights / License:**
Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International

This page was generated automatically upon download from the ETH Zurich Research Collection. For more information please consult the Terms of use.
Identifying User Tasks in Map Based-Pedestrian Navigation from Eye Tracking Data

Hua Liao ¹,², Weihua Dong ¹, Georg Gartner ², Huiping Liu ¹
liaohua@mail.bnu.edu.cn

¹ 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
² Department of Geodesy and Geoinformation, Vienna University of Technology, Vienna, Austria

Abstract. Eye movement data convey a wealth of information that can be used to inspect human behavior and cognitive processes. In this study, we explored to identify user tasks in map-based pedestrian navigation using eye movement data. We collected 44 participants’ eye movement data by conducting an eye tracking experiment in real-world environments. We trained and cross-validated a Random Forests classifier to classify 7 common user tasks using 564 features. The preliminary results show the classifier can achieve an overall accuracy of 48%.

Keywords. wayfinding, eye movement, Random Forests

1. Introduction

It is demonstrated that high-level (top-down) factors such as user tasks can guide visual attention and lead to different eye movement patterns (Yarbus et al. 1967). For example, in the well-known Yarbus’s experiment, participants’ eye movements were recorded when they were viewing a painting under different tasks such as remembering locations of the people and objects in the painting. The results revealed significant differences of eye movement patterns produced by seven different tasks. Can the Yarbus’s experiment process be inverted? That is, is it possible to detect user tasks (only) from eye movement data?

While many studies have explored detecting user tasks in laboratory (Boisvert and Bruce 2016), here we report an ongoing study focusing on
detecting users tasks in real-world environments. Specifically, we aim to identify user tasks in map-based pedestrian navigation from eye movement data using machine learning methods. Addressing such real-world tasks is of great importance to design adaptive map interfaces and gaze-based human-computer systems for navigation (Kiefer et al. 2013).

2. Methods

2.1. Eye movement data collection

We conducted a real-world pedestrian navigation experiment to collect eye movement data. We used SMI eye tracking glasses (https://www.smi-vision.com) with a sampling rate of 60Hz. The scene video and sound were recorded synchronously. All of the data were stored in a Thinkpad laptop that was connected to the eye tracker.

2.2. Participants

Forty-four participants (20 females and 24 males) from various backgrounds (e.g., geography, engineering, arts, management, etc.) were recruited to participate in the experiment. They were undergraduate, graduate, or PhD students. The mean age of the participants was 23 (SD=2.5). All participants had normal or corrected-to-normal vision. None of them reported eye diseases.

2.3. Procedure

After calibration, participants were given a set of A4 papers with instructions and maps printed on it. They were required to finish two routes. In each route, they were required to perform 7 tasks orderly. Instructions of these tasks are described below. The participants were told that there was no lower nor up time limit for all tasks unless otherwise stated in the instructions. They were required to say ‘Begin’ when they were ready to begin a task and say ‘I have found it’ when they finished the task.

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 the north in the environment.”

Task 2 (local environment object search). “Please find an object in the nearby environment.” The target object of Route 1 was a sculpture and Route 2 was a labeled sign.

Task 3 (environment free view). “Please view the surrounding environment freely for 30 seconds. When the time is up, you will be informed.”
Task 4 (map target search). “Please search a labeled target on the map on the next page.” There are three trials for this task.

Task 5 (route memorizing). “Please memorize the route on the next page for 30 seconds.”

Task 6 (walking to the destination). “Please walk from here to the destination along the route you have just memorized. You can look at the map if necessary during the navigation.”

Task 7 (map free view). “Please view the map on the next page freely for 30 seconds. When the time is up, you will be informed.”

After finishing these tasks for both routes, the participants were required to finish an questionnaire to evaluate their sense of direction (Hegarty et al. 2002), familiarity of the routes, and the task load of the experiment.

2.4. Data preprocessing

5 participants were excluded from analysis because of calibration failure or recording failure. The recorded eye movement data, video, and sound were segmented for each task. We further conducted a trial selection process to select trials that meet the following criteria: (1) tracking ratio was greater than 80 percent; and (2) length was greater than 10 seconds. This process resulted in 520 trials (60 task 1, 50 task 2, 80 task 3, 80 task 4, 70 task 5, 60 task 6, and 120 task 7). For trials that were longer than 10 seconds, only the first 10 seconds of data were used.

2.5. Feature extraction

We extracted a total of 564 features to characterize statistical, spatial and temporal variations of eye movements as described below.

**Basic statistical features.** This type of features is based on the basic eye movements (fixations and saccades), blinks, and pupil diameters. They are commonly used eye movement indicators in literature. 35 basic statistical features were calculated including 3 frequency features (fixation frequency, saccade frequency, and blink frequency) and 32 features resulted from mean, maximum, minimum, and skewness of 2 fixation indicators (fixation duration and fixation dispersion), 4 saccade indicators (saccade duration, saccade amplitude, saccade velocity, and saccade latency), 1 pupil diameter indicator (pupil diameter), and 1 blink indicator (blink duration).

**Fixation density features.** In order to capture the spatial distribution of fixations, we calculate the fixation density images using Gaussian kernel density estimation (Silverman 1986). The resulting 2D density images were then down-sampled to $20 \times 20$ or $1 \times 400$ feature vectors.
Saccade direction features. We first divided saccade directions into 4 and 8 cardinal directions. For each direction, we computed the number of saccades (N=12); the mean, maximum, and minimum values of saccade amplitude and saccade duration (N=72). This results in 84 features in total. This type of features was inspired by Kiefer et al. (2013).

Time slicing statistical features. To capture temporal variations of eye movement, we divided each data sample into small time segments with a time bin size of two seconds. We then computed the mean, max, and min values of fixation duration, fixation dispersion, and pupil diameter, resulting in 45 features.

2.6. Classification and cross-validation
We adopted random forests for classification. A pilot study showed random forests could achieve higher accuracy than SVM in this study (Figure 1). We used the open-source Python implementation from scikit-learn library (Garreta and Moncecchi 2013). 10-fold cross-validation method was adopted.

![Figure 1. Comparison of accuracy between Random Forests and SVM.](http://doi.org/10.3929/ethz-b-000225606)

3. Preliminary Results
The confusion matrix of the classification is shown in Table 1. The results show that 249 out of 520 trials are classified correctly. The overall accuracy is 48% (chance level=1/7 ≈ 14%). The mean kappa coefficient is 0.40 (SD=0.09). Task 3 (environment free view) has the highest recall and precision which is 78% and 61%, respectively, followed by Task 6 (walking to the destination) with a recall of 73% and precision of 58%. This is probably because of the difference of spatial distribution of fixations: during free view,
participants fixated a larger area than the other tasks, while during walking, participants attended at the center of the visual field for most of the time.

However, the classifier failed to distinguish Task 1 (self-localization and orientation). Both the recall and precision are zero. A deeper look at the results reveals that 25 out of 60 samples of Task 1 has been misclassified into Task 4 (map target search). This is partly because participants needed to search where they were on the map, which is very similar to the map target search task.

It is interesting to note that the non-free view tasks tend to be misclassified into free view tasks (either environment free view or map free view). For example, 22 out of 50 samples of Task 2 (local environment object search) have been misclassified into Task 3 (environment free view); 24 out 60 samples of Task 1, 30 out of 80 samples of task 4 (map target search), and 46 out of 70 samples of Task 6 (route memorizing) have been misclassified into Task 7 (map free view).

<table>
<thead>
<tr>
<th>Actual task</th>
<th>T1</th>
<th>T2</th>
<th>T3</th>
<th>T4</th>
<th>T5</th>
<th>T6</th>
<th>T7</th>
<th>Sum</th>
<th>Precision</th>
</tr>
</thead>
<tbody>
<tr>
<td>T1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>4</td>
<td>6</td>
<td>0.00</td>
</tr>
<tr>
<td>T2</td>
<td>3</td>
<td>18</td>
<td>8</td>
<td>1</td>
<td>0</td>
<td>2</td>
<td>1</td>
<td>33</td>
<td>0.55</td>
</tr>
<tr>
<td>T3</td>
<td>4</td>
<td>22</td>
<td>62</td>
<td>1</td>
<td>0</td>
<td>8</td>
<td>5</td>
<td>102</td>
<td>0.61</td>
</tr>
<tr>
<td>T4</td>
<td>25</td>
<td>1</td>
<td>1</td>
<td>38</td>
<td>4</td>
<td>2</td>
<td>17</td>
<td>88</td>
<td>0.43</td>
</tr>
<tr>
<td>T5</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>4</td>
<td>19</td>
<td>0</td>
<td>15</td>
<td>41</td>
<td>0.46</td>
</tr>
<tr>
<td>T6</td>
<td>1</td>
<td>8</td>
<td>8</td>
<td>4</td>
<td>1</td>
<td>44</td>
<td>10</td>
<td>76</td>
<td>0.58</td>
</tr>
<tr>
<td>T7</td>
<td>24</td>
<td>1</td>
<td>1</td>
<td>30</td>
<td>46</td>
<td>4</td>
<td>68</td>
<td>174</td>
<td>0.39</td>
</tr>
</tbody>
</table>

| Sum         | 60 | 50 | 80 | 80 | 70 | 60 | 120 |     |           |
| Recall      | 0.00 | 0.36 | 0.78 | 0.48 | 0.27 | 0.73 | 0.57 |     |           |

*Table 1*. Confusion matrix of the classification. Overall accuracy=48%.

**4. Summary and Outlook**

This ongoing study aims to use eye movement data to infer user tasks in map-based pedestrian navigation in real-world environments. We employed Random Forests to classify 7 common tasks from 520 trials using 564 eye movement features. Our preliminary results show the classifier can achieve an overall accuracy of 48% with a kappa coefficient of 0.40. Currently, we are investigating the influence of time window size, familiarity, and individual difference on the classification performance and participants’ visual behavior during the navigation.
References


