Attention allocation of traffic environments of international visitors during virtual city walks

Kristof Nevelsteen

SADL, KULeuven; Celestijnenlaan 200E, 3001 Heverlee, Belgium
Tel: 0032/16329727

Kristof.nevelsteen@sadl.kuleuven.be

Keywords: 'real world' images; attention allocation

Abstract.
The analysis procedure of an eye-tracking experiment on mobile mapping image (Google Street View) is presented. International visitors virtually navigated culture historic cities in Flanders. Routes of their walk were mapped and compared with their attention allocation during their walk. The focus lies on the atomization of coupling the fixation data and the recorded video-file of the virtual walk. The advantages and disadvantages of the technique are described.

1 Introduction

More and more people arrive in cities which are unknown for them. By arriving they automatically become part of the local systems (e.g. the local traffic system) and they are confronted with environments different from their familiar ones. International visitors who visit for a short time should be able to navigate smooth to capture quick and efficient the essence of the city and to enjoy their trip. Navigation is the science of determining our position and our course of motion (Hofmann-Wellenhof et al., 2003). When the destination is not located in the near vicinity, navigation demands wayfinding (Mengue-Topio et al., 2011). Montello (2001) conceptually separates navigation behavior into locomotion and wayfinding. Locomotion refers to navigation behavior in response to current sensory-motor input of the immediate surroundings and is therefore strongly related to the act of physically moving through space. Wayfinding tasks, in contrast, aim at reaching destinations beyond the current sensory horizon and involve some representation of the environment, decision making and/or planning processes. The relation between decision making and gaze behavior has been studied in non-spatial contexts, and has been shown to be closely tied to preferences (e.g. Simion and Shimojo, 2007; Armel et al., 2008). But the question of how gaze behavior relates to wayfinding-related tasks such as memorizing and recalling path choices and movement decision along a route, or deciding between path alternatives at junctions when exploring novel environments, has been addressed only in few studies (e.g. Schuchard et al., 2006; Vembar et al., 2004; Spiers and Maguire, 2008) and they used particularly virtual environments (Wiener et al., 2011). Currently, it is also unknown which characteristics of the environment viewers are attending to when deciding between path alternatives in the context of wayfinding. The features people attend to when inspecting images of scenes are investigated in numerous studies, but again mainly in a non-spatial context (for an overview see Henderson, 2003). Emo, 2012 suggests that studying wayfinding behavior with photographic images is possible as the participants were responding to the photographs as if they were in situ and promotes the benefits of modeling real world wayfinding behavior using photographs.

'What features do international visitors attend to during a 'real world' city walk in an unknown environment?' is the main research question of this study. The results presented are preliminary and the discussion about the analysis procedure is equally important.

2 Materials and Methods

2.1 General Experiment Design

A test was designed to let newly arrived visitors visit two culture historic cities in Flanders in a virtual way, but showing real world images. To explain differences in perception and route choice, personal characteristics of the participants were recorded via a short survey before starting the actual test. A static eye-tracking device (Tobii T120, screen size: 17", screen resolution: 1280 x 1024) recorded their fixations during the virtual walks in Google Street View. Street View was used to have real world images. Virtual computer designed environments have benefits when analyzing the data but the study goal was to capture attended features in realistic (unfortunately still static and not dynamic) environments. Emo (2012) found that real world photographs can be used to model wayfinding tasks. The result is a quantitative dataset of participants' attention allocation, which can be statistically analyzed.

2.2 The Test

International students were asked to participate in the test, the moment they arrived at the international students' office of the university. 45 participants (Western Europe: 35%, Eastern Europe: 15%, Africa: 9%, Latin-America: 22%, Asia: 19%) entered the study. Two (Leuven and Bruges) city walks were performed online in Google Street View. Participants were positioned circa 70 centimeter from the screen (as instructed in the eye-tracker manual) and were navigated by the guiding researcher to the railway station and were then left freely to navigate the city with the central market as end goal. They had a paper tourist-map available with two tourist attractions per city marked on. Interactions with the paper map were not tracked by the eye-tracker. Participants could visit those locations but were not obligated to do so. It was pointed out that it was not a race, there was no time limit to reach the old market (average walk was circa 25 minutes). No further instructions were given. During the virtual walk participants' eye-movements were recorded simultaneous with the Google Street view images ending in a video-file of their chosen route.

2.3 Analysis Procedure

Figure 1 shows the analysis procedure. Tobii (eye-tracker) software produces a table of all fixations attributed with x,y screen coordinates, a timestamp (milliseconds since the start of the virtual walk) and several fixation metrics. The coordinates were used to map the fixations centers (figure 1, upper left) and the fixation metrics of the table were linked to this map. The fixation centers were then buffered to create 'viewing areas'. Since the fixation data comes with a timestamp, the general 'viewing areas' map was then cut in individual maps every 2.5 second and converted into '.tiff images', with the viewing areas colored black (RGB 0,0,0) and the background white (RGB 255,255,255). An example is showed in figure 1, upper right. The above described sub-process was automated via Python scripting and uses Arcpy (= predefined ArcGIS statements that can be used in regular Python scripting to use ArcGIS functions).

Because of the slow internet connection, the screen recorded video-file was not smooth. Breaking the video-files (5 frames/second) in individual frames revealed series of very similar frames. On average 13 frames in a row could be considered similar so one frame per 2.5 seconds was extracted from the large file representing the image seen during those 2.5 seconds (figure 1, lower right). Bias because of the changing stimuli with changes in position of the participant is limited in this way.

The frames and the viewing images (.tiff images) are then subtracted from each other (figure 1, lower left). Both the individual frame creation and image subtraction was automated via Matlab code.

This result is a large number (every 2.5 seconds) of images per visitor, which show the features in the image fixated by the participant during those 2.5 seconds (luminance map (Holmqvist 2001)). Because the data were attributed with fixation metrics, this procedure can be repeated for different fixation metrics. So is it for example possible to only use viewing areas with a fixation duration above a certain threshold to find feature categories which received attention for a long time. Because the procedure is for a large part automated via Matlab and Python scripting a large amount of data could be processed.

The luminance maps are then manually converted into viewing tables. This approach was chosen since the demarcation of Areas of interest (AOI) manually is an endless job and automated tracking of certain features (e.g. traffic signs) in changing real world photographs is not yet fully up to date. Different feature categories (e.g. traffic signs, statues, street name signs, tourist relevant infrastructure, etc.) were predefined and when a visitor fixated on a feature of that category it was marked in the table. This is very time consuming and the major disadvantage of the procedure used. Automated feature/object recognition in images would be useful for this part. By converting the image into tables, statistics can be performed upon them.

The recorded video-files were also used to map the routes of the participants. Routes were mapped manually in QGIS using Open Streetmap as base map. The routes were mapped to later compare the attention allocation of the participants with their navigation-strategy and route choice.

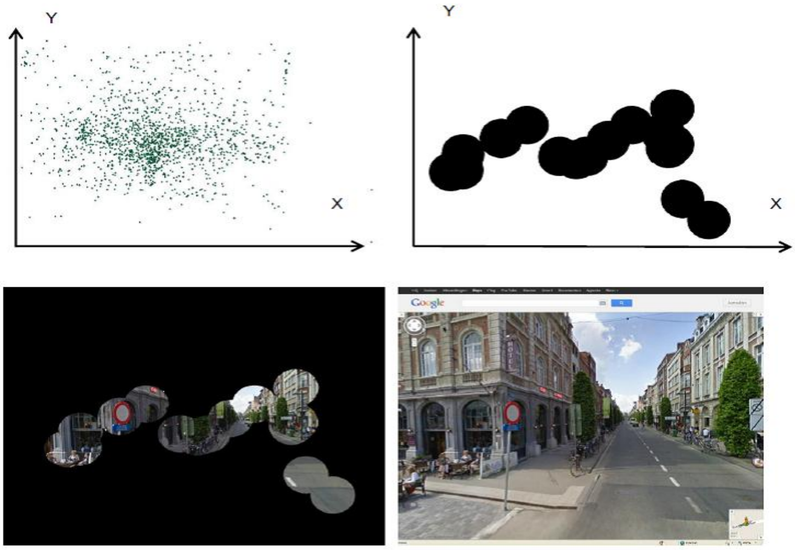

Fig. 1. Analysis procedure

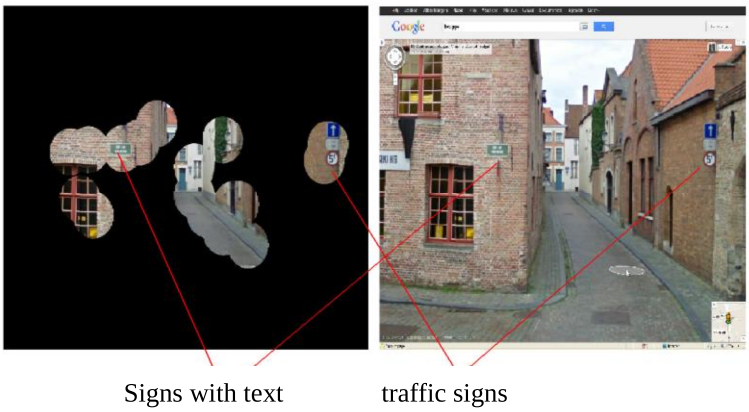

Fig. 2. Example of often attended features

3 Results

The results described are not fully finalized and still open for improvement. The main idea is to compare route choice and navigation strategy with the attention allocation of the participants. The navigation strategy was determined from the constant communication with the participants during the virtual walks. These 'interviews' were digitized and via cluster analysis, groups based on similar navigation strategies were defined. Per participant a list of followed streets was generated in Qgis. Based on this list another cluster analysis was performed to find categories of participants with comparable actual route choice. These two clusters were then compared. The most important findings are:

- There is big difference in how people think they navigate a city and how they actually chose their routes.
- Origin of the participant determines the route-choice. Asian people are for example not so adventurous.
- Pre-knowledge of the city has no influence on the chosen routes.

The viewing tables were used to find some preliminary results about the attention allocation of visitors navigating unknown cities. The most important ones are:

- Signs with text and traffic signs are often attended features (see figure 3 as an example)
- A city which is more difficult to navigate has attention allocations that are more on features that can help navigating.
- More fixations on navigation related features result in less fixation on traffic related features which can result in dangerous situations

Further analysis will focus on the relation between attended features and navigation strategies, attended features and actual chosen routes and the difference between fixation metrics (e.g. fixation duration, first fixation, etc.)

4 Discussion and Conclusion

More than on the presented results this short paper focuses on the analysis procedure of the eye-tracking data. A possible analysis of eye-tracking data on mobile mapping images like Google Street View, to virtually navigate a city, but based on 'real world' images is presented. Analyzing eye tracking data on changing stimuli is a challenging endeavor. Video-files show the routes participants chose and the eye-tracker device delivers digitally the fixation location on the screen. In essence the video-file could be converted in the standard eye-tracking software to scenes and then it is possible to define areas of interest on those scenes. By defining area of interest it would be possible to extract the fixation metrics on those areas and do statistical tests on those data. For about 3000 minutes of film, this working procedure is however not suitable, because it is too time consuming and so too expensive. The video-files and fixation data can be matched by using standard GIS software and by converting the video-file into separate frames. The problem of changing stimuli with changing position of the participants (by displacement or changing viewing direction) is not fully solved here. The slow internet connection made it possible to bypass the problem to some extent, but is not a valid solution in the long run. Future research will focus on this problem. Virtual, computer created environments do permit automatic tracking of features and hold potential to resolve the problem. However, more interesting in the context of attention allocation research is using real world images which unfortunately don't allow consistent automatic tracking yet. The described approach uses image processing in Matlab to create luminance maps of all the fixations. Since the technique starts from the raw fixation data, attributed with fixation metrics it is possible to experiment with these attributes once you have the scripts to automate the process and on could create different luminance maps which highlight different aspects of an experiment. The major drawback is the post-processing. Once you have the features viewed per time step per participants, the creation of quantitative data by viewing tables is still time consuming. On top, the data are not normalized. Even by working relative, the viewing tables don't give you for example the exact number of traffic signs that could be fixated. Automated feature recognition would be very useful to quantify the number of possible feature to fixate and then use it as normalization.

Eye-tracking on mobile mapping images holds some potential for attention allocation studies in real world scenes, since it can be performed from behind the desk. This is a major advantage when you want to do research with for example children (safety is guaranteed). The degree of game-play is however quiet high and the lack of dynamism are disadvantages compared to in situ 'real world' studies.

References

Armel C., Beaumel A., Rangel A.: Biasing simple choices by manipulating relative visual attention. Judgment and Decision Making, 3 (5), pp. 396-403 (2008)

Emo B.: Wayfinding in real cities: experiments at street corners. Spatial cognition VII. Stachniss C., Schill K. and Uttal D. Berlin Heidelberg, Springer: pp. 232-249 (2012)

Henderson J.: Human gaze control during real-world scene perception, Trends in cognitive sciences, 7 (11), pp. 498-504 (2003)

Hofmann-Wellenhof B., Legat K., Wieser M.: Navigation: Principles of positioning and guidance. Wenen: Springer – Verlag (2003)

Holmqvist K., Nyström M., Andersson R., Dewhurst R., Jarodzka H., Van de Weijer J.: Eye tracking, a comprehensive guide to methods and measures, Oxford University Press (2011)

Mengue-Topio H., Courbois Y., Farran E., Sockeel P.: Route learning and shortcut performance in adults with intellectual disability: a study with virtual environments, Research in Developmental disabilities, 32 (1), pp. 345-352 (2011)

Montello D.R.: Spatial cognition, In international encyclopedia of the social and behavioral sciences, Oxford: Pergamon Press, pp.14771-14775 (2001)

Schuchard R., Connell B., Griffiths P.: An environmental investigation of wayfinding in a nursing home, In proceedings of the 2006 symposium on eye tracking research and applications, San Diego. pp.33 (2006)

Simion C., Shimojo S.: Interrupting the cascade: Orienting contributes to decision making even in the absence of visual stimulation, Perception and Psychophysics, 69 (4), pp. 591-595 (2007)

Vembar D., Iyengar N., Duchowski A., Clark K., Hewitt J., Pauls K.: Effect of visual cues on human performance in navigating through a virtual maze, In Eurographics symposium on virtual environments (2004)

Wiener J.M., Holscher C., Buchner S., Konieczny L.: Gaze behaviour during space perception and spatial decision making, Psychological Research, 76 (6), pp. 713-29 (2011)

Spiers H.J., Maguire E.A.: The dynamic nature of cognition during wayfinding, Journal of Enviromnetal Psychology, 28, pp. 232-249 (2008)