Future Generation Computer Systems ▮ (▮▮▮▮) ▮▮▮–▮▮▮

Contents lists available at ScienceDirect

Future Generation Computer Systems

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

Exploring the potential of a mobile eye tracker as an intuitive indoor pointing device: A case study in cultural heritage

Moayad Mokatren, Tsvi Kuflik *, Ilan Shimshoni

The Department of Information Systems, University of Haifa, Mount Carmel, Haifa 31905, Israel

HIGHLIGHTS

- Analysis and examination of the potential use of mobile eye tracker in a museum is presented.
- A mobile museum visitors guide that uses a mobile eye tracker as a pointing device is described.
- A user study comparing the use of a museum visitors guide that uses an eye tracker and a conventional one is presented.

ARTICLE INFO

Article history:
Received 26 February 2017
Received in revised form 3 June 2017
Accepted 1 July 2017
Available online xxxx

Keywords:
Mobile eye tracking
Museum visitors guide

ABSTRACT

Current technology offers a variety of ways for context-aware information delivery to mobile users. The most challenging aspect, however, is to determine what the user is interested in. The user's position is the best available hint, but if we know what the user is looking at and what his or her gazing profile is, we can narrow down the possibly relevant objects of interest. With the advent of mobile and ubiquitous computing, it is time to explore the potential of mobile eye tracking technology for natural, intelligent interactions between users and their smart environment, not only for specific tasks, but also for the more ambitious goal of integrating eye tracking into the process of inferring mobile users' interests, for the purpose of providing them with relevant services, a research area that has received little attention so far.

In this work, we examine the potential of integrating a mobile eye tracker, as a natural interaction device, into an audio guide system for museum visitors. Using it as a pointing device enables the system to reason unobtrusively about the user's focus of attention and to deliver relevant information about it as needed. To realize this goal, we integrated an image-matching based technique for indoor positioning and an eye-gaze detection technique to identify the user's focus of attention into two different versions of a mobile audio guide: (1) a proactive version that delivers information automatically whenever user interest is detected, and (2) a reactive version that notifies the user about the availability of this information, thus giving the user more control over information delivery. Furthermore, we developed a conventional museum visitors' mobile guide system using a smartphone and low-energy Bluetooth beacons for positioning; this guide was used as a reference system.

The three museum visitors' guides were evaluated in realistic settings at the Hecht[1] Museum, a small museum, located at the University of Haifa that has both archeological and art collections. The experimental evaluation compared the contribution of the three versions of the audio guide to the visit experience. The results showed that the mobile eye tracking technology, although unfamiliar, and perhaps even immature, was accepted by the participants. The mobile eye tracker audio guide was perceived as preferable to the conventional museum mobile guide, especially with regard to learning during the visit. Furthermore, with regard to proactivity in context-aware systems, the results showed that the participants like to be in control, and that most of them preferred the reactive version of the mobile eye tracker audio guide over the proactive one.

* Corresponding author.

E-mail address: tsvikak@is.haifa.ac.il (T. Kuflik).

[1] http://mushecht.haifa.ac.il/Default_eng.aspx.

http://dx.doi.org/10.1016/j.future.2017.07.007
0167-739X/

1. Introduction

For most of us, vision is our main sense for gathering information. When we want to gather information about or express interest in something in our environment, the first thing we do is look at it. However, the only information we get in this way is what we see: size, shape, color, distance, etc. Nowadays, a lot of information about the objects that we see is available online and is easily accessible. Theoretically, it is only a click away, a query away, or available by simply activating the mobile device, writing the query, submitting it, scrolling through the results list, selecting the relevant one, and accessing the relevant page. This is, however, a complicated set of actions to perform in a mobile scenario, when immediate, personalized, and context-aware information is desired. Current technology offers a variety of ways to deliver information to mobile users. Context awareness is the general term describing the attempt to deliver relevant information at the relevant time and place to the user. Most context-aware services nowadays make use of the communication and computational power (and sensors) of the users' mobile devices (mostly smartphones). In addition, they interact with their users mainly by their mobile device's touch screens, which have a few major limitations: they are limited in size, the users must look at them during the interaction, and they have to use a keyboard or select icons. Although voice commands can be used to activate applications, this option is still very limited.

A major challenge in the mobile scenario is to know exactly what the user is interested in. In classic human–computer interactions, the users use a pointing device, most commonly a mouse or, in the case of a touch screen, a finger. However, this is becoming a major challenge in the mobile setting, as noted by Calvo and Perugini [1], who surveyed novel pointing approaches for wearable computing. The user's position is the best hint, accompanied by his or her orientation. Still, there are many possibly interesting objects near and around the user. If we know what the user is looking at, and what the specific user's gazing profile is, then we can narrow down the possibly relevant objects of interest and better serve the user with relevant service/information.

Given the current performance of our mobile devices, we should be able to gain seamless access to information of interest, without the need to take pictures or submit queries and look for results, which are the prevailing methods of interaction with our mobile devices. As we move towards "cognition-aware computing" [2], it becomes clearer that eye-gaze based interaction should and will play a major role in human–computer interaction (HCI) before/until brain computer interaction methods will become a reality [3]. The study of eye movements started almost 100 years ago. Jacob and Karn [4] presented a brief history of techniques that were used to detect eye movements. The major works on this topic dealt with usability, and one of the important works was begun by Fitts et al. [5] in 1947, when they began using motion picture cameras to study the movements of pilots' eyes as they used cockpit control and instruments to land an airplane. "It is clear that the concept of using eye tracking to shed light on usability issues has been around since before computer interfaces, as we know them" [4]. In recent years, commercial mobile eye trackers that enable us to detect what someone is looking at have become available [6]. Moreover, eye tracking and image based object recognition technology have reached a reliable degree of maturity: it is now possible to develop a system based on this technology, precisely identifying what the user is looking at [7]. We shall refer to this field by reviewing techniques for image matching and extend them for location-awareness use, and we will follow the approach of "What you look at is what you get" [8].

The museum visit experience has been changing over the last two decades. With the progress of technology and the spread of handheld devices, many systems were developed to support the museum visitor and enhance the museum visit experience. The purpose of such systems was to encourage the visitors to use devices that provide multimedia content rather than use guide books, and therefore focus on the exhibits instead of flipping through pages in the guide book [9–12]. With the advent of mobile and ubiquitous computing, it is time to explore the potential of this technology for natural, intelligent interactions between users and their smart environment, not only for specific tasks, but also for the more ambitious goal of integrating eye tracking into the process of inferring mobile users' interests and preferences, for the purpose of providing them with relevant services and developing a better user model to enhance their experience, an area that has received little attention so far. This work aims at exploring the potential of mobile eye tracking technology in enhancing the museum visit experience by integrating and extending these technologies into a mobile museum visitors' guide system, so as to enable the use of machine vision to identify visitor positions and objects of their interest, in order to deliver personalized information.

In this study, we addressed the following questions:

- Q1: How can we use a mobile eye tracker to identify the user's location and object of interest?
- Q2: How can we integrate a mobile eye tracker as a pointing device in a system that delivers information to the museum visitor?
- Q3: To what extent does the use of a mobile eye tracker in an audio guide contribute to the museum visit experience?

2. Background and related work

2.1. Background on eye tracking and computer vision

Eye tracking is an active area of research that has seen significant progress over the years. However, as Hansen and Ji noted in their survey [13] of eye-tracking research, "Despite active research and significant progress in the last 30 years, eye detection and tracking remains challenging due to the individuality of eyes, occlusion, variability in scale, location, and light conditions". They concluded that "The tendency to produce mobile and low-cost systems may increase the ways in which eye tracking technology can be applied to mainstream applications, but may also lead to less accurate gaze tracking. While high accuracy may not be needed for such applications, mobile systems must be able to cope with higher noise levels than eye trackers for indoor use". Relatively inexpensive, easy to use mobile eye trackers have appeared in recent years. In 2015, Yousefi et al. [14] surveyed a large variety of such mobile eye tracking applications and technologies for aviation, marketing, learning, medicine, and other fields, and predicted that such applications would continue to appear. As noted, most existing mobile eye trackers are intended for specific applications and tasks.

Modern mobile eye trackers usually record video of the scenes for further analysis using a front camera [7]. With the advent of computer vision technology, we can exploit this camera to develop a positioning tool. An image matching procedure can be used to identify the museum visitor's location/position by comparing the front camera scene with a set of known dataset images. In that way, position can be identified in a pre-defined environment. Furthermore, the gaze data can be used to infer the user's attention in a specific scene, thus making it possible to deliver personalized information related to a specific object in the scene. Consider a device consisting of a forward-looking camera and an eye tracker. The device takes a picture while the user is fixating on a certain position within the image. The challenge is to recognize the object in the scene and deliver content related to this object to the user.

Please cite this article in press as: M. Mokatren, et al., Exploring the potential of a mobile eye tracker as an intuitive indoor pointing device: A case study in cultural heritage, Future Generation Computer Systems (2017), http://dx.doi.org/10.1016/j.future.2017.07.007.

When an image is taken, and given to the algorithm together with a set of database images, the goal is to find the image that shows the same scene as the test image. The algorithm should work in cluttered scenes (scenes from which objects have been removed or added), where the images are not taken from the same pose and with varying illumination. In this work, we used local image features that are unaffected by nearby clutter or partial occlusion. The features are at least partially invariant to illumination, 3D projective transforms, and common object variations.

The features must also be sufficiently distinctive to identify specific objects among many alternatives. Several types of local features have been developed. The most popular type of feature is SIFT [15] but others also exist (e.g. SURF [16], BRISK [17], and ORB [18]). The SIFT features are invariant to image scaling, translation, and rotation, and partially invariant to illumination changes and affine or 3D projection. When the SIFT algorithm is applied to an image, it produces a set of SIFT features with their descriptors. By matching the descriptors recovered from the test image to the ones recovered from the image set, a set of possible matches between the features is recovered. Images with a large number of possible matches are candidates for the matched scene. However, the matched scene might still be incorrect especially if the number of matches is small. To overcome this problem, we can exploit the geometric relationships between the positions of the matched features in the two images (a fundamental matrix in the general case and a homography matrix for planar scenes). This matrix is usually computed using a robust estimation procedure from the RANSAC family [19–26]. These algorithms can be run on the test image and on each of the images from the dataset. If the algorithm succeeds, then usually the two images are of the same scene.

If, however, the image database is large, such a procedure can be time consuming and cannot be run in real time. To address this problem, several algorithms have been proposed for scene recognition from a test image. Using various techniques, the number of possible images from the database is reduced and the aforementioned algorithms only need to be run on the remaining ones. Some of the algorithms develop a single descriptor for the whole image and recognize the scene using similarity measures between the descriptors [27]. Others use convolutional neural net classifiers for deciding whether two images belong to the same scene [28]. Statistics on matches between local descriptors can also be used for scene recognition [29]. Once the database image has been recovered, it will be accompanied by a set of matches between the test image and one database image. At this stage, the user's fixation point comes into play. The designer of the application can mark in advance positions on the database images where objects of interest are visible. In order for the object of interest to be recognized, the fixation point on the test image has to be transformed into a point on the database image and, if that point is close to one of the marked objects of interest, the content related to that object can be delivered. Using the matched points and the recovered geometric relationships between the images (fundamental matrix or homography), the transformation of the fixation point is computed.

2.2. Technology and CH

Over the last 20 years, cultural heritage has been a favored domain for personalization research. For years, researchers have experimented with the cutting-edge technology of the day. Now, with the convergence of internet and wireless technology, and the increasing adoption of the Web as a platform for the publication of information, cultural heritage material can be exploited by a museum visitor before, during and after the visit, with different goals and requirements in each phase. However, cultural heritage sites have a huge amount of information to present, which must be filtered and personalized for easy access. Personalization of cultural heritage information requires a system that can model the user (e.g., interest, knowledge, and other personal characteristics), as well as contextual aspects, select the most appropriate content, and deliver it in the most suitable way. It should be noted that achieving this result is extremely challenging in the case of first-time users, such as tourists who visit a cultural heritage site for the first time.

The museum environment has many limitations, such as the restriction not to make noise, not to talk loudly, not to touch anything, etc. Obviously, mobile guides for museum visitors should complement rather than replace traditional interpretation methods [30]. Under these limitations, Cheverst et al. [12] proposed two key requirements for such guides, the first of which is flexibility. The system is expected to be sufficiently flexible to enable visitors to explore, and learn about, a museum in their own way, including controlling their own pace of interaction with the system. The second requirement is context-awareness, meaning that the information presented to the visitors should be tailored to their personal or environmental context. The personal context includes the visitor's interests, the visitor's current location, and exhibits already visited, while the environmental context includes the time of day and the opening hours of the museum.

Adrissino et al. [9] have argued that the evolution and convergence of technologies, together with the needs expressed in recent museum research, open new opportunities for personalization research, which has the potential to improve the presentation of information, the exploration of content interesting for the specific user, and collaboration among users having similar interests, as well as adapt to heterogeneous user contexts and devices.

2.3. Ubiquitous computing and HCI

Weiser's vision [31] of ubiquitous computing, with its invisible yet attentive computing environment that provides the right information to the right person at the right time, is an exciting vision of how to evolve computer technology. Within a ubiquitous computing environment, the computing elements and their intercommunication are largely hidden from the user, and the technology is not readily visible – it is worn or embedded in building infrastructure – and is spoken with and related to. In a follow-up to Weiser's vision, and with the maturation of mobile technology, the idea of context awareness emerged. "Context and context-awareness provide computing environments with the ability to usefully adapt the services or information they provide. It is the ability to implicitly sense and automatically derive the user needs that separates context-aware applications from traditionally designed applications, and this makes them more attentive, responsive, and aware of their user's identity, and their user's environment" [32].

Interaction between users and computers occurs at the user interface, including both hardware and software. As computers become mobile and invisible, designing the interaction between humans and computers becomes more and more challenging. Interaction design means designing interactive products to support people in their everyday and working lives. Because HCI concerns a human and a machine in conjunction, designing a user interface requires knowledge of both the human and the machine side: knowledge about communication theory, graphic disciplines, social sciences, and cognitive psychology, on the one hand, and knowledge about computer graphics techniques, operating systems, and programming languages, on the other hand. Huang et al. [33] discussed the challenges in HCI design for mobile devices. The limitations of current mobile devices, such as limited input/output/screen size and inconvenient navigation through hierarchical menus, are well known and mostly a result of the device dimensions. Hence, applying implicit interaction, using eye-gaze as a natural interaction method, may help overcome some of these challenges.

Please cite this article in press as: M. Mokatren, et al., Exploring the potential of a mobile eye tracker as an intuitive indoor pointing device: A case study in cultural heritage, Future Generation Computer Systems (2017), http://dx.doi.org/10.1016/j.future.2017.07.007.

2.4. Related work

Many studies were conducted on detecting eye movements before the appearance of computer interfaces as we know them today. Robert and Jacob [8] presented techniques for local calibration of an eye tracker. This technique produces a mapping of the eye movement and eye wandering measures. In addition, they presented a technique for fixation recognition with respect to extracting data from noisy, jittery, error-filled streams and for addressing the problem of "Midas touch", where the eye tracking system is "misled" by people inadvertently looking at an item they are not interested in. Jacob and Karn [4] presented a list of promising eye tracking metrics for data analysis:

- Gaze duration—cumulative duration and average spatial location of a series of consecutive fixations within an area of interest.
- Gaze rate—number of gazes per minute on each area of interest.
- Number of fixations on each area of interest.
- Number of fixations overall.
- Scan path—sequence of fixations.
- Number of involuntary and number of voluntary fixations (short fixations and long fixations should be well defined in terms of millisecond units).

Brône et al. [34] have implemented effective new methods for analyzing gaze data collected with eye tracking devices and shown how to integrate it with object recognition algorithms. They presented a series of arguments as to why an object-based approach may provide a significant advantage, in terms of analytical precision. In order to identify the area of interest (AOI), they attached physical markers to each AOI. They presented some limitations of this technique, such as the challenge of the installation. We used their lessons in our study by defining the object of interest (OOI) (this term being more appropriate to our museum visit scenario) "digitally" on the images in the database.

Pfeiffer et al. [35] presented the EyeSee3D method. They combined geometric modeling with inexpensive 3D marker tracking to align virtual proxies with the real-world objects, so that fixations on objects of interest can be classified automatically while supporting a completely free moving participant. During the analysis of pose estimation accuracy, they found that the marker detection failed when the participant looked sideways and there was simply no marker within view, or due to swift head movements or extreme position changes. Ohm et al. [36] tried to determine where people look when navigating in a large-scale indoor environment and what objects can assist them in finding their way. They conducted a user study and assessed the visual attraction of objects with an eye tracker. Their findings show that functional landmarks like doors and stairs are most likely to be looked at and named as landmarks.

Beugher et al. [37] presented a novel method for the automatic analysis of mobile eye-tracking data in natural environments and for processing this data by applying object, face, and person detection algorithms. The obtained detection results were satisfactory for most of the objects. However, large scale variations resulted in a lower detection rate (for objects that were looked at both from very far away and from close by.)

Schrammel et al. [38,39] studied attentional behavior of users on the move. They discussed the unique potential and challenges of using eye tracking in mobile settings and demonstrated the ability to use it to study the attention on advertising media in two different situations: within a digital display on public transportation and towards logos on a pedestrian shopping street; they also presented ideas for a general attention model based on eye gaze. Kiefer et al. [40] also explored the possibility of identifying user attention by eye tracking in the tourism setting, by examining, for example, when a tourist gets bored looking at a city panorama. This scenario may be of specific interest to us in future work, as locations or objects that attract more or less interest may be used to model user interest and trigger further services/information later on.

Nakano and Ishii [41] studied the use of eye gaze as an indicator of user engagement, trying also to adapt it to individual users. Engagement may be used as an indicator of interest, and the ability to adapt engagement detection to individual users may also enable us to infer interest and build/adapt a user model using this information. Furthermore, Ma et al. [42] demonstrated an initial ability to extract user models based on eye gaze of users viewing videos.

The use of handheld devices as a multimedia guidebook in museums has led to improvement in the museum visit experience. Research has confirmed the hypothesis that a portable computer with an interactive multimedia application has the potential to enhance interpretation and to become a new tool for interpreting museum collections [43].

Studies about integration of multimedia guidebooks with eye tracking have already been conducted in the context of museums and cultural heritage sites. Museum Guide 2.0 [44] was presented as a framework for delivering multimedia content to museum visitors that runs on a handheld device and uses the SMI viewX eye tracker and object recognition techniques. The visitor can hear audio information when looking at an exhibit. A user study was conducted in a laboratory setting, but no real museum was involved. We extended this work by integrating the eye tracker into a real museum visitors' guide and experiment on it in a realistic setting.

As we have seen, there is a large body of work about monitoring and analyzing user eye gaze in general and some also in cultural heritage. Moreover, the appearance of mobile eye trackers opens new opportunities for research in mobile scenarios. It was also demonstrated on several occasions that eye gaze may be useful in enhancing a user model, as it might make it possible to identify user attention (and interests). In mobile scenarios, when users also carry smartphones equipped with various sensors, implicit user modeling can be carried out by integrating signals from various sensors, including the new eye-gaze sensor, to better model the user and offer better personalized services. Sensors like GPS, compasses, accelerometers and voice detectors have thus far been used to model user context and interest, (see, for example, [45]). Mobile scenarios in fact cover a wide variety of activities, from jogging to shopping to cultural heritage. The tasks in each scenario are different and user attention differs according to the task. Bulling and Gallersen [46] discuss some of the characteristics and challenges of mobile eye-tracking given the technological progress in the field and, specifically, how these characteristics make eye movements a distinct information source about the user's context. Giannopoulos et al. [47] presented viGaze—an eye tracking framework that was demonstrated in a supermarket. It allows the dynamic creation and design of virtual shelves, their enhancement with audio and visual information, as well as the design and enablement of gaze-based interactions (explicit and implicit) that can take place between the users and the virtual shelves. Using a prototype implementation of the framework, they conducted a user study that demonstrated its feasibility in the context of an instrumented retail environment. They concluded that the ideas generalize easily to different kinds of instrumented environments.

Although much research has been conducted on monitoring, analyzing, and using eye gaze to infer user interest, little attention has been paid so far to user gazing behavior "on the go". This scenario poses major challenges as it involves splitting attention

Please cite this article in press as: M. Mokatren, et al., Exploring the potential of a mobile eye tracker as an intuitive indoor pointing device: A case study in cultural heritage, Future Generation Computer Systems (2017), http://dx.doi.org/10.1016/j.future.2017.07.007.

Fig. 1. Pupil eye-tracker (http://pupil-labs.com/pupil).

between several tasks at the same time—avoiding obstacles, gathering information, and paying attention to whatever seems relevant. While user behavior has been monitored and analyzed in various ways in smart environments, using a variety of sensors, this has hardly ever been done for eye gaze.

3. Tools and methods

For the purpose of this study, a commercial mobile eye tracker, the Pupil-Dev eye tracker [7], was integrated into a mobile museum visitors' guide system as a positioning tool and for focus of attention detection, both using computer vision techniques. It comprises a lightweight eye tracking headset, an open source software framework for mobile eye tracking, as well as a graphical user interface to play back and visualize video and gaze data. It features high-resolution scene and eye cameras for monocular and binocular gaze estimation. We used the monocular version (30_{hz}) as an input device for inferring the object of interest (OOI) (see Fig. 1).

The software and GUI are platform-independent and offer real-time pupil detection and tracking, calibration, and accurate gaze estimation. Results of a performance evaluation show that Pupil can provide an average gaze estimation accuracy of 0.6° of visual angle with a processing pipeline latency of only 0.045 s [7].

A key challenge in using mobile technology for supporting museum visitors is figuring out what they are interested in. This may be achieved by tracking where the visitors are and the time they spend there [48]. A more challenging aspect of the problem is to identify exactly what they are looking at [49]. The developed system addresses the two aforementioned challenges – it identifies user focus of attention accurately, and it does so unobtrusively. The system exploits and extends an image-based positioning technique (described later in Section 4) to deliver audio information about exhibits in the museum. A visitor wears the mobile eye tracker, which is connected to a laptop (carried in a backpack), and gazes steadily at an exhibit for approximately three seconds while standing in one place, after which the image-based positioning procedure starts, location/position and point of interest are identified, and audio information regarding the desired exhibit is delivered. To meet the goal of unobtrusiveness, two assumptions were made regarding the interaction of the user with the system: the 3-s gazing period that triggers the positioning system is long enough to avoid the "Midas touch" problem, and also long enough to ensure that the user is not moving but standing and looking at an exhibit. A simple "stop" gesture was adopted for starting/stopping the presentations. Further studies on how the user interacts with the system are obviously required, possibly taking into consideration the metrics suggested by Jacob and Karn in [4] and listed in Section 2.4.

The system (see Fig. 2) consists of three main modules: a position locator, an OOI identifier, and a broadcaster. The modules and the databases are local on the computer for reasons of speed, as well as to reduce the latency of delivering information to the user/visitor. The flow of input/output is as follows: The mobile eye tracker streams a captured scene frame from the world and a fixation point, after which SIFT features of the frame are extracted and sent to the position locator together with the fixation point. The position locator module matches the current features with a predefined set of descriptors (that are extracted from dataset images beforehand). Once there is a match, the position locator streams the position ID together with the fixation point to the OOI identifier, which identifies the object of interest using the fixation point. Finally, the OOI identifier passes the object ID to the broadcaster, which finds the appropriate audio file and broadcasts it to the user. We have implemented two versions of the mobile eye tracker based audio guide:

1. Proactive: After the position of the visitor and point/object of interest are identified, a "beep" sound is played, and the audio information about the exhibit is delivered immediately after.
2. Reactive: After the position of the visitor and point/object of interest are identified, a "beep" sound is played, and the system waits for a mid-air gestural action ("stop sign"). After the user makes the appropriate gesture, the audio information is delivered.

For both versions, the same mid-air gestural action of a "stop sign" was used to stop the audio presentation. We implemented this feature using Dense Optical Flow, which was presented in [50]. We did so by looking at three continuous frames, each divided into 100 blocks (10×10), and from each block we took one point, calculated its optical flow magnitude and counted it if it was in the threshold range. In our case the range is between 10 to 150 pixels. We call these points violation points. We determined that a "stop sign" gesture was made if three consecutive frames had zero violation points.

Since the system is to be used in a real-time scenario, the following functional requirements should be met:

- Response time: The system should be responsive enough to deliver the desired information within interactive time. According to Card et al. [51], 10 s is about the limit for keeping the user's attention focused. In our system, we set a maximum time limit of 5 s for delivering the desired information.
- Accuracy: The system should be accurate enough to deliver the correct information at the right time. This means the correct information should be delivered when the visitor is standing in front of the exhibit, in accordance with the *what you look at is what you get* approach.

The performance of the eye-tracker was first evaluated in three user studies, after which the system was evaluated in a separate user study in a realistic setting. 22 students from the University of Haifa participated in the latter study, which was conducted in the Hecht Museum, a small museum at the University of Haifa that has both archeological and art collections. The study included an orientation session to explain the use of the eye tracker, followed by a tour of the museum with the eye tracker, which was connected to a laptop carried in a backpack, with audio content about the exhibits delivered via headphones.

Please cite this article in press as: M. Mokatren, et al., Exploring the potential of a mobile eye tracker as an intuitive indoor pointing device: A case study in cultural heritage, Future Generation Computer Systems (2017), http://dx.doi.org/10.1016/j.future.2017.07.007.

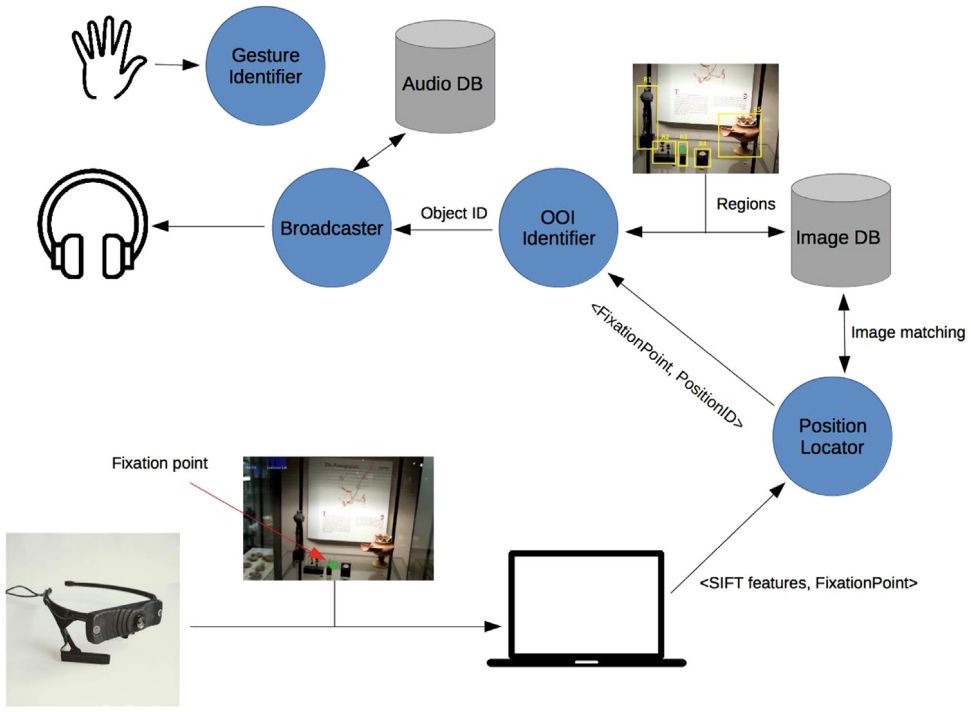


Fig. 2. The process of OOI identification and information delivery.

Table 1
Number of misses per grid cell.

Cell #	6	18	19	23	24
# of misses	5	5	3	5	5

Visitor experience when using two different mobile guides was also compared: an audio guide that uses the mobile eye tracker and a conventional mobile guide that runs on a smartphone and uses BLE (Bluetooth low energy) Estimote beacons [52] to identify the position of the visitor. Whenever the visitor reached a predefined area of interest, a multimedia presentation was delivered.

4. Assessing the performance of the Pupil-Dev eye tracker in realistic settings

To assess the accuracy of the mobile eye-tracker device in realistic settings, we first had to design and develop the system to work and to be evaluated in the required operational range. To this end, we conducted three preliminary user studies.

4.1. User study 1: looking at grid cells

Five students from the University of Haifa, without any visual disabilities, participated in this study (average age is 22), the goal of which was to determine the accuracy of the detection of a predefined POI. The students were asked to look at a wall-mounted grid from a distance of 2 m and track a finger while using the eye tracker (see Fig. 3, left). Standing at a fixed point, they were asked to look for approximately 3 s at each cell the finger pointed at. On average, the eye tracker detected fixation with an accuracy of ∼80% (most of the missed fixations were in the edges/corners—see Table 1 for details). In addition, the average fixation point error, in terms of distance from the center of the cell, was approximately 5 cm.

During the study, we encountered several practical problems. The first is that the eye tracker was not fitted individually to each participant. The device consists of two cameras, the first for delivering the scene and the second directed to the right eye for detecting fixations. When the device did not fit properly, the vision range decreased and parts of the pupil fell outside the area of the captured frame (see Fig. 3 (right) for example), as no fixations were detected. Another limitation was that tall people have to step back from the object (to keep it in the camera's field of view), which affects the accuracy.

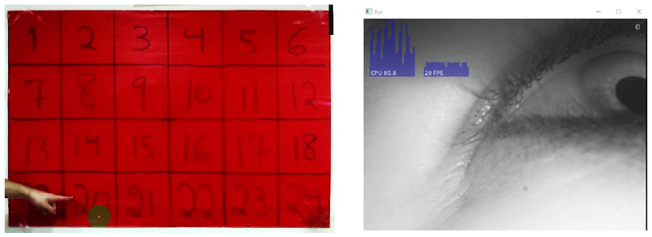


Fig. 3. Left: Screen capture from user study 1. The finger points at a cell at which the participant was asked to look. The green circle is the fixation point returned by the eye tracker. The size of each grid cell is 20 × 20 cm. Right: Screen capture from eye camera. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

4.2. User study 2: looking at an exhibit

In this study, we examined the accuracy of the eye tracker in a realistic setting. One participant (1.79 m tall) was asked to look at exhibits in the museum. Several exhibits where chosen with different factors and constraints (see Figs. 4 and 5). The main constraint in this case was the distance from the exhibit, since the visual range increases when the distance grows, and we have to cover all the objects that we are interested in. Table 2 presents height of the objects from the floor and the distance of the participant from each object. The next step was to examine fixation accuracy after making sure that the participant is standing at the correct distance from the exhibit. The participant was asked to look at different points in the exhibit/scene. In the gallery exhibits, the scan path

Table 2
Experiment details—we considered the three glass shelves on the far left of the vitrine shown in Fig. 5 (right).

Exhibit type	Width (cm)	Height (cm)	Height from floor (cm)	Stand distance (cm)
Vitrine shelf	80	25	150	150
	80	15	120	230
	80	20	90	310
	80	15	40	390
Gallery	60	67	150	200

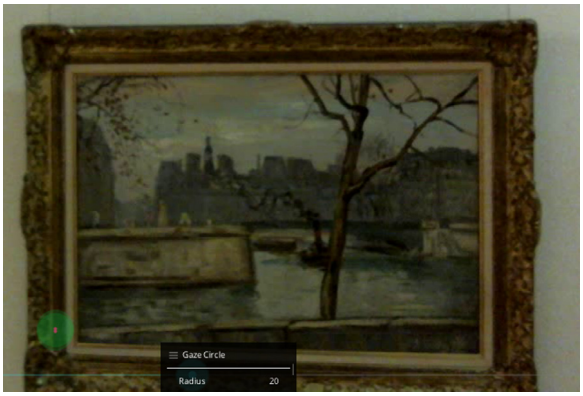


Fig. 4. Gallery exhibition.

was set to be the four corners of the picture and finally the center of it. Regarding the vitrine exhibits, for each jug, one point at the center was defined.

Not surprisingly, we obtained 100% accuracy in the art wing since all the pictures are placed at an ideal height. The archeological wing is considered a more challenging environment, since objects are placed at different heights and differ in size. Specifically, when the user has to tilt his or her head to look down, we noticed poor performance. As this poor performance is due to a limitation of the current device, we did not consider low-height exhibits in our experiments. More challenging exhibits are those that are placed in harsh lighting conditions: conditions that change drastically during the day, as a result of changing sunlight. Hence, in the case of the archeological wing, we estimated that about 60% of the exhibits are detectable with the current device.

As the goal of the study was to explore the potential of the device in a realistic setting, being able to detect 60% of the exhibits seemed good enough for our purpose. We assume that as the technology improves, the current limitations will be reduced or even eliminated completely.

4.3. User study 3: image-based positioning

4.3.1. Preparation

In this study, we wished to answer the question: **How can we use a mobile eye tracker to identify the location and the object of interest?** We implemented an image-based positioning technique to identify the visitor's position and object of interest in a predefined museum layout. We used the SIFT algorithm [15] to match the current scene's camera to a set of images from a predefined dataset for locating the visitor's position. What remained after locating the visitor's location is to infer his/her object of interest. Since we have a matched image from the dataset, transforming the fixation point that we get from the eye tracker will lead us to a point in the dataset image. Hence, with predefined regions/labels in every dataset image, we infer the visitor's object of interest.

A visitor entering the museum can stop/stand in front of each exhibit at different viewpoints (in terms of distance and angle). Consequently, preparing the dataset that represents the museum

Table 3
Standing distances.

Exhibit type	Distance (cm)
Vitrine shelves	50–70
Art gallery	70–100
Small statues	50–70
Large exhibits	100–150

layout plays a crucial role in obtaining accurate matching results. One option might be to capture several images from different viewpoints for each exhibit. Time complexity is the limitation of this option, since the image-to-image matching procedure requires massive amounts of computation, which can cause a delay in delivering the current position. To identify typical viewpoints, 10 regular museum visitors were observed when visiting the Hecht Museum, and their standing distance from each exhibit was measured for four types of exhibits: vitrine shelves (Fig. 5, right), art galleries (Fig. 4), small statues (Fig. 5, left) and large exhibits (Fig. 7). The distances are presented in Table 3. During the observations, we ignored the angle between the visitor and the exhibit, because several such frontal-viewing angles are possible. We therefore just considered the distances.

Once typical distances were known, images of the exhibits were taken and assigned distinct label values (image ID), and a set of rectangular regions within the images (around objects) was defined and assigned a distinct ID.

The matching procedure for location identification and interest detection was done in four steps:

1. An eye-tracker scene camera frame was taken (Fig. 6 (left)) after the user focused on an object (looked at it steadily for three seconds). This was done by tracking good features in three frames within the three-second period (for each second we stored one frame) using the Lucas–Kanade method for optical flow [53].
2. Image-to-image matching was applied using SIFT features. Two images (the current frame and the dataset frame) are said to match if the number of matched features divided by the total number of features is higher than a threshold. (in our case we chose the threshold $= 0.12$). The result is an image that matches the current location (Fig. 6 (right)).
3. A mapping transformation was obtained to transform the fixation point identified in the scene camera of the eye-tracker to a suitable/matched point in the image that exists in our dataset with labeled regions (see Fig. 6, right), since the viewpoint from which the objects were photographed can differ in the two images. For example, one image might be rotated relative to the other or one zoomed in/out because the visitor's distance from the object differed from the distance from which the data-set image was taken. The mapping transformation was obtained using the homography matrix, which was computed using a robust estimation procedure from the RANSAC family [19].
4. The final step of finding the object is simple now that we have obtained mapped fixation points and labeled regions. What remains is to determine which object (if any) the point corresponds to.

Please cite this article in press as: M. Mokatren, et al., Exploring the potential of a mobile eye tracker as an intuitive indoor pointing device: A case study in cultural heritage, Future Generation Computer Systems (2017), http://dx.doi.org/10.1016/j.future.2017.07.007.

Fig. 5. Small statue exhibit (left). Backlit vitrine exhibit (right).

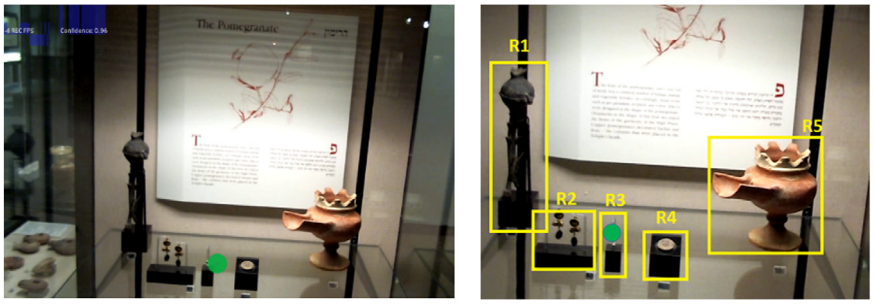


Fig. 6. Left: Example of an image taken by the scene camera of the eye-tracker. The green point is the fixation point. Right: Image-to-image matching. The yellow rectangles are the regions around each object. The green point is the fixation point after transformation from the left image is performed. The corresponding region would be R3. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

As the matching of the camera scene image with every image from the dataset is time consuming, we optimized the process by starting it from images near the visitor's current location. To that end, we represented the dataset using a graph, where each node represents the exhibit's image/label and the arc values represent the physical distance between two neighboring exhibits.

4.3.2. Evaluation of the accuracy and matching time

We observed museum visitors and noticed that a visitor who enters a museum might walk around or stand and look at an exhibit. We have to distinguish between these two cases. To recognize the event of looking at an exhibit, we set a time interval (three seconds) of looking at a scene. We use this as a trigger for starting a matching procedure, comparing the image of the scene with a set of existing position-representing images (that were taken beforehand, by the same type of camera, at every position from several different angles, based on our observations of visitors' behavior). The matching procedure yields a set of scores, and the image with the highest score is selected as representing the visitor's current position. During the study conducted in the Hecht Museum, one person was asked to walk around the museum and look at exhibits. When he looked steadily at an exhibit for about three seconds, the image based positioning procedure started and the position and the object of interest was identified. With a database of images taken from 24 positions and representing 18 exhibits, the process took 1.5 s on average. The experimental results are presented in Table 4. It is clear the positions were correctly identified in most cases. However, there were two positions/exhibits where the performance was poor or even failed most of the time. Examining these cases revealed that low-accuracy results were obtained for exhibits with unusual features, e.g., those placed in such a way that the visitor can look at them from a distance and from a wide variety of angles (Figs. 7 and 8 for example), a scenario that requires many reference images. A possible solution is to add several images from different distances and angles for each such exhibit. Exhibits with different lighting conditions (especially nearby windows) that may affect the image-matching process will also require many reference images. A possible solution is to add several images taken at different times of day. In our system, we do not consider these cases.

5. Empirical study

Once the performance of the eye tracker was studied and the systems developed, we evaluated them in a realistic setting. The research questions we were interested in exploring were:

Please cite this article in press as: M. Mokatren, et al., Exploring the potential of a mobile eye tracker as an intuitive indoor pointing device: A case study in cultural heritage, Future Generation Computer Systems (2017), http://dx.doi.org/10.1016/j.future.2017.07.007.

Table 4
Accuracy of exhibit matching.

Item #	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
Accuracy	1	1	1	1	0.42	1	1	1	1	1	0.11	1	1	1	1	1	1	1

Fig. 7. Exhibit E5. A large exhibit that requires additional dataset images from different viewpoints.

Fig. 8. Exhibit E11. Additional dataset images are required for each object.

(Q3a) To what extent does a proactive version of the visitor's audio guide contribute to the visitor experience in museums, compared with the contribution of a reactive version?
(Q3b) To what extent does the use of a mobile eye tracker in an audio guide contribute to the visitor experience in museums, compared with the contribution of conventional mobile guide?
The hypotheses for Q3a were:
H0: The proactive and reactive versions will not differ significantly in terms of their contribution to visitor experience in museums.
H1: The proactive and reactive versions will differ significantly in terms of their visitor experience in museums.
The hypotheses for Q3b were:
H0: The mobile eye-tracker based audio guide and the smartphone based mobile guide will not differ significantly in terms of their contribution to visitor experience in museums.
H1: The mobile-eye tracker based audio guide and the smartphone based mobile guide will differ significantly in terms of their contribution to visitor experience in museums.

5.1. Participants

Twenty-two students from the University of Haifa participated in the study, some of whom were randomly invited, and some of whom were occasional visitors who happened to be at the museum during the experiments. 12 participants were females and 10 participants were males, with an average age of 24.45 years (SD = 4.415). The choice of university students, who are not characteristic of visitors to the Hecht Museum, may impact dependent variables, such as average age or background knowledge, which could influence the experimental results. However, we made this choice because regular visitors to the Hecht Museum are mainly groups of senior citizens and classes of school children.

5.2. Experiment setup

To test the hypotheses, experiments manipulating and measuring variables under controlled conditions were carried out. The independent variable was the type of system while the dependent variables were:

Please cite this article in press as: M. Mokatren, et al., Exploring the potential of a mobile eye tracker as an intuitive indoor pointing device: A case study in cultural heritage, Future Generation Computer Systems (2017), http://dx.doi.org/10.1016/j.future.2017.07.007.

(a) Usability factors, as measured by means of the SUS questionnaire [54], and (b) a subjective assessment of the visitor experience, as measured (in a set of three additional questionnaires) by user preference. The goal of the subjective assessment was to determine, whether the users felt that the guide was an effective way to get information and learn about the objects in the exhibit, and whether the system was sufficiently intuitive.

5.3. Procedure

The study took about an hour and a half and was performed as a randomized *counterbalanced, within-groups* study to eliminate the learning effect.

The evaluation procedure was organized as follows:

1. It began with a brief introduction to the study, after which the participants were asked to complete a personal and background questionnaire.
2. Then the participants were given a short demonstration of each system and its features. The participants were also instructed how to perform the calibration process and, for the reactive version, how to interact with the guide.
3. The participants were next requested to visit the exhibits in the archeological wing, using the visitor's guide systems. The visit started with the calibration process. Then the visitors were instructed to follow a pre-defined path in the museum and to look at particular objects (about which we have information).
4. During the experiment, the participants filled out the SUS questionnaire three times—(once for every system they experienced). At the end of the experiment they filled out three additional questionnaires (1) an individual questionnaire whose purpose was to compare visitor experience while using the two different versions of the mobile eye-tracker based audio guide; (2) an individual questionnaire regarding user acceptance and accuracy of the gaze-based interface and (3) an individual questionnaire whose purpose was to compare user experience while using the conventional mobile guide and the preferred version of the mobile eye-tracker based audio guide.
5. Finally, the participants were briefly interviewed and answered two open questions:
   - How was your museum visit experience when using the mobile eye tracker audio guide?
   - What do you think about the way the system interacts with the user (the gaze-based interface)?

5.4. Experimental results

The three museum visitor's guide systems obtained high usability scores: (1) mean = 86.47 and SD = 7.96 for the proactive version of the mobile eye-tracker audio guide; (2) mean = 86.36 and SD = 11.84 for the reactive version of mobile eye tracker audio guide; and (3) mean = 93.75 and SD = 5.7 for the conventional smartphone-based mobile.

As can be seen in Fig. 9 (left), there is no real difference between the proactive and the reactive mobile eye tracker audio guide but there are differences between the smartphone based system and the eye-tracker based systems. This was confirmed by the Friedman test with Bonferroni correction. The null hypothesis that the distributions of all three cases are the same was rejected ($\chi^2 = 9.829$, $p = 0.007$). We found that there was no significant difference between the proactive and reactive versions ($p = 0.706$), but there were significant differences between the proactive and the smartphone versions ($p = 0.016$), and between the reactive version and the smartphone versions ($p = 0.048$), both significant at alpha (<0.05).

In addition to the usability study, the user's subjective assessment of the proactive or reactive system was analyzed (see Fig. 10). The following aspects of user preference were assessed:

1. *Preferred version for overall museum visit*: 13 participants preferred the reactive version compared with 9 participants who preferred the proactive version.
2. *Effectiveness for getting information and learning*: 11 participants preferred the reactive version compared with 8 who preferred the proactive version. 2 participants had no preference.
3. *Intuitiveness*: 18 participants preferred the proactive version compared to 3 participants who preferred the reactive version. One person had no preference.

As can be seen in Fig. 10, there are slight differences between the proactive and the reactive systems with respect to users' preferences and perceived effectiveness, where the reactive system outperformed the proactive one: however, a binomial test showed that the differences are not significant ($p = 0.523$ for the preferred version and $p = 0.648$ for effectiveness). Still, the proactive system seemed to be much more intuitive than the reactive one and the binomial statistical test confirmed this observation ($p = 0.001$).

Next, the **preferred** mobile eye tracker audio guide was compared with a conventional mobile guide on a smartphone (see Fig. 11).

1. *Preferred guide for the overall museum visit*: 16 participants preferred the mobile eye tracker audio guide compared to 6 who preferred the conventional mobile guide on a smartphone.
2. *Ease of use and intuitiveness*: 12 participants preferred the conventional mobile guide on a smartphone compared to 10 participants who preferred the mobile eye-tracker audio guide.
3. *Learning*: 17 participants preferred the mobile eye-tracker audio guide compared to 5 participants who preferred the conventional mobile guide on a smartphone.

As can be seen in Fig. 11, the mobile eye tracker was considered the preferred guide, and it also outperformed the smartphone with respect to learning. These observations were confirmed by a binomial statistical test that showed significant differences between the systems ($p = 0.52$, which is marginally significant, for the preferred version, and $p = 0.017$ for learning). However, there was no real difference between the systems with respect to ease of use, and the binomial statistical test confirmed this observation ($p = 0.832$).

We were interested in analyzing the acceptance and accuracy of the gaze-based interface, in order to evaluate the potential of using the mobile eye tracker as a pointing device. To that end, we presented the participants with a four-question questionnaire, where each question had a five-point Likert scale response. Figs. 12–15 present the responses of the participants to the questions regarding the gaze-based interface: they liked the interface (Fig. 12), the calibration process did not bother them much (Fig. 13), they learned about objects of interest (Fig. 14), and they usually got information about the correct object (Fig. 15).

The final step of the study was a brief, five-minute interview, conducted at the end of the one-and-a-half-hour study, where the overall visitor experience was briefly discussed and the participants were asked about their museum experience using the mobile eye-tracker audio guide and to give their opinion about the way the system interacts with the user.

The answers to these open-ended questions were transcribed by the interviewer. Given the conditions of this part of the study,

Please cite this article in press as: M. Mokatren, et al., Exploring the potential of a mobile eye tracker as an intuitive indoor pointing device: A case study in cultural heritage, Future Generation Computer Systems (2017), http://dx.doi.org/10.1016/j.future.2017.07.007.

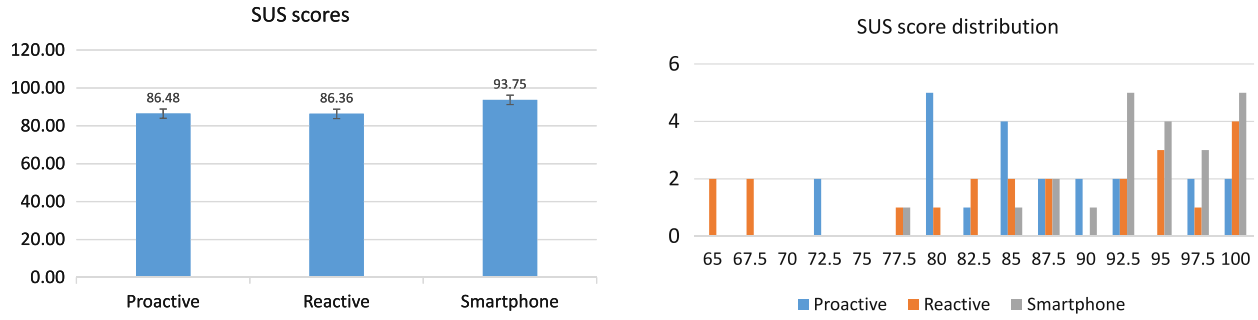


Fig. 9. SUS scores for the three museum visitor's guides (left) and score distributions (right).

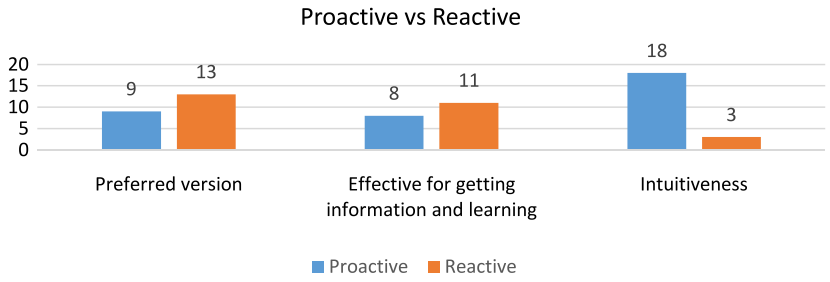


Fig. 10. Comparison between proactive and reactive versions of the mobile eye-tracker audio guide.

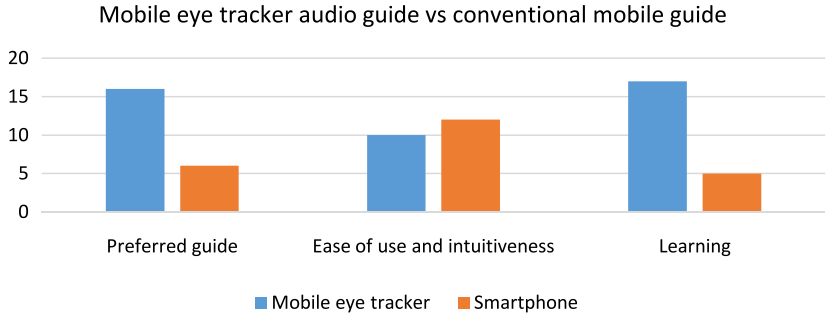


Fig. 11. Comparison between mobile eye tracker audio guide and conventional mobile guide on smartphone regarding the museum visit experience.

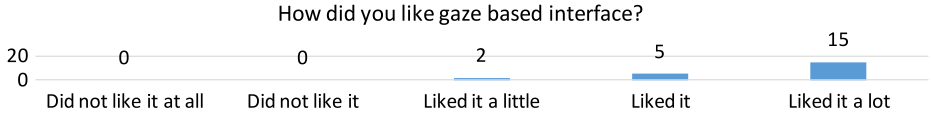


Fig. 12. Responses to the question, How did you like the gaze based interface?

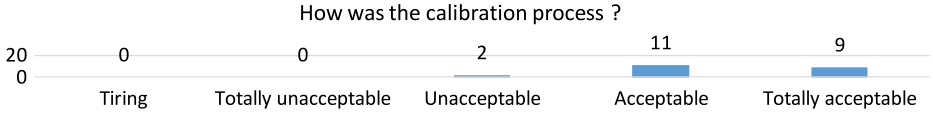


Fig. 13. Responses to the question: To what extent was the calibration process acceptable to you?

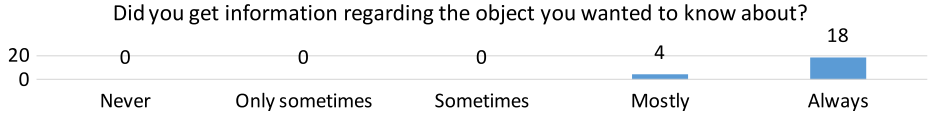


Fig. 14. Responses to the question: Did you get information regarding the object you wanted to know about?

Please cite this article in press as: M. Mokatren, et al., Exploring the potential of a mobile eye tracker as an intuitive indoor pointing device: A case study in cultural heritage, Future Generation Computer Systems (2017), http://dx.doi.org/10.1016/j.future.2017.07.007.

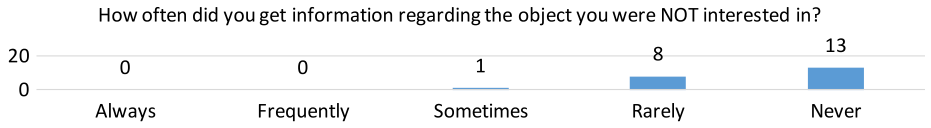


Fig. 15. Responses to the question: How often did you get information regarding the object you were not interested in?

we see this step as merely complementary. Data analysis was carried out in accordance with grounded theory analysis principles [55]. With regard to the first open question, two categories of response were identified: those relating to visit enrichment ("Very cool and interesting, tempting me to listen to information even if the exhibit seems uninteresting to me"; "It enriched the experience, I'll remember the visit for a long time") and those relating to the use of futuristic technology ("Groovy, felt futuristic, it's what I always hoped museum visits should be like"; "I felt I'm living in 2020, it felt like a journey in time"). With regard to the second open question, the answers tended to be very short and, repetitive, and two categories of response were also identified: those relating to the intuitiveness of the system and those relating to the long delay ("3-s is too long").

6. Discussion

We developed and evaluated a mobile museum visitor's guide that uses an eye tracker as a pointing device. The current technology, while still limited and premature for daily use, has the potential to be used for experimentation in realistic settings. The evaluation results show that the system got satisfactory scores ("very good" scores for both the proactive and the reactive versions). Not surprisingly, the conventional mobile guide got a higher usability score, probably because the mobile eye-tracking technology is still not mature enough and the experimental procedure was a bit cumbersome as compared to smartphones, which are used by the visitors on a daily basis and completely familiar to them.

In contrast to its lower usability score, the mobile eye-tracker audio guide was shown to be preferable to the conventional smartphone guide with regard to *overall museum visit experience*. Participants who preferred the mobile eye tracker audio guide indeed attributed their preference to experience-related aspects of the guide: "It is mainly because of the experience"; "I prefer to get audio information while looking"; "It's because of the ability to get information while walking "; "It's more comfortable to use, there's no need to play presentations and to scroll down on the screen"; "Less effort, more accurate, more control and gives information per object"; "It's because of the innovative technology"; "You need to search for the object in the museum when using the smartphone"; "It gives an opportunity to 'live the museum' or 'to feel the museum'"; "It's more efficient, but what about if we were in a larger museum? Why should I have to search for the location of every object when using the smartphone? It's a waste of time!". Participants who preferred the conventional mobile guide attributed their preference not to overall experience but to familiarity: "I know to use the smartphone better, it's more intuitive"; "I have more control over the smartphone"; "Wearing the mobile eye tracker is overload"; "The use of the smartphone is more comfortable, you are not limited in where you stand"; "It's quicker". With regard to *ease of use and intuitiveness*, the results showed no real difference between the mobile eye-tracker audio guide and the conventional one.

With regard to *learning*, the results showed that most of the participants preferred the mobile eye tracker audio guide over the conventional one. Answering the open-ended question, these participants justified their preference with statements such as: "I'm in focus, nothing in my hand"; "It's more accurate and better for self-learning"; "You don't need to perform a lot of actions, just to look"; "More intuitive, more interesting"; "Easier to get information with"; "I'm getting information while looking"; "More control over the objects I'm interested in". In contrast, those participants who did prefer the conventional mobile guide said: "More control in which I can move backwards and forwards in the presentation"; "Showing the presentation is preferable to just listening to audio"; "I would use the mobile eye tracker audio guide when it adds new things to the exhibit, like augmented reality".

With regard to the *preferred guide for the overall museum visit,* the results showed that most of the participants preferred the reactive version over the proactive version of the mobile eye-tracker audio guide (even though there was no difference in the usability questionnaire). The open-ended questions show that the main reasons for preferring the reactive version are that the users feel more in control when using this version and they can decide when to play the audio. According to Lanir et al. [56], museum visitors feel less in control when using proactive context-aware systems. Our reactive version was developed as an attempt to overcome user aversion to such proactive systems. Nonetheless, our interaction design was not sufficiently accepted as an effective method of user control. It was suggested that the three-second "stop" gesture be replaced with a button or a sensor on the mobile eye-tracker device. Therefore, while this mobile technology is still too premature for daily use, once improved it can be easily adopted and used as a natural pointing device.

With regard to the *effectiveness of getting information and learning*, the questionnaire results showed no real differences between the two versions of the mobile eye-tracker audio guide.

Finally, with regard to *intuitiveness*, the questionnaire results showed that most of the participants preferred the proactive version over the reactive version. The open-ended questions indicate that this preference is mainly due to the simplicity of the interaction and the fact that it requires less effort.

Like any study, the current study has its limitations. The first limitation is technical. We used a specific eye tracker, examined and mapped its limitations (as explained above), and tried to work within these limitations. It may be that other devices (such as Tobii Pro Glasses 2[2] for instance) are better in terms of performance metrics such as elevation, field of view, accuracy, or latency, but these devices are also much more expensive at this time. An additional limitation stems from the fact that the mobile eye tracker is a *wearable* device, and thus problematic for people who wear glasses. In general, however, we assume that the technology will get better. Hence, while noting the technical limitations, we believe our results are encouraging.

The gap between our exploratory study, which yielded promising results, and the integration of the mobile eye tracker into a real system, also needs to be addressed. The "Midas touch" problem must be more adequately addressed in realistic settings. Correctly identifying the user's object of interest, a problem we solved by setting a high threshold for the decision, might also require us to consider additional factors, such as the options presented in [4]: gaze duration, gaze, number of fixations overall and on each area of interest, scan path and number of involuntary and voluntary fixations.

System errors must also be dealt with: erroneous position identification, erroneous object identification, and positions/objects

[2] https://www.tobiipro.com/product-listing/tobii-pro-glasses-2/.

Please cite this article in press as: M. Mokatren, et al., Exploring the potential of a mobile eye tracker as an intuitive indoor pointing device: A case study in cultural heritage, Future Generation Computer Systems (2017), http://dx.doi.org/10.1016/j.future.2017.07.007.

without information. These are beyond the scope of this exploratory study. Possible solutions to the problem of system errors may include notifying the user that the position or the specific object of interest were not identified or that there is no information about them (content preparation is expensive, so it is reasonable to assume that there will be exhibits for which no information is available). It may also be possible to provide general information about the area/exhibition or the exhibits in front of the visitor, when detailed information is not available.

Finally, we need to design a natural method of gesture-based interaction to start/stop information delivery.

In this study, we focused on exploring the potential of a mobile eye tracker as a pointing device for natural interaction in smart environment. An interesting alternative may be to use a remote eye tracker for this purpose. However, although accurate stationary eye trackers do exist, the use of a remote system poses its own challenges. First and foremost is the challenge of identifying the relevant user, to ensure coherent interaction throughout the visit. A remote system would also require the installation on many stationary eye trackers, which would have to account for factors such as differences in user height or standing distance. Moreover, such systems also pose computational challenges such as inferring the exact fixation point when the user is not standing in front of the eye tracker [57]. Despite these challenges, the use of a remote system is an interesting idea and a possible direction for future work.

7. Conclusions and future work

In this work, we explored the use of a mobile eye tracker as an intuitive pointing device in realistic settings, using cultural heritage as a case study because of the vast amount of information available in museums. We first studied the technical aspects and the limitations of the device we used. Then, we developed a tool for image-based positioning and for detecting objects/points of interest in real-time using computer vision techniques. Finally, we developed a context-aware mobile audio guide system using a mobile eye tracker as a pointing device. We developed and tested two different versions of this guide, proactive and reactive. We evaluated the system in a user study in a realistic setting at the Hecht Museum. The results showed that the mobile eye-tracking technology, even though unfamiliar and possibly immature, was accepted by the participants. The mobile eye-tracker audio guide was perceived as the preferred museum visitors' guide compared to a conventional museum mobile guide, especially with respect to learning. Unsurprisingly, the results also showed that the participants like to be in control, as most of them chose the reactive version of the system.

This study lays the foundations for the use of eye-trackers as a natural HCI pointing device, in real-time mobile scenarios, where there is a need to dynamically and quickly identify the user's focus of attention and act upon it, according to the user's situation. In the cultural heritage setting, visitor movement in space, time spent, information requested, vocal interaction and orientation have been used to infer user interest in museum exhibits and as the social scenario when a group is visiting the museum together [45,58–60]. Adding eye gaze as an additional source of information may greatly enhance the system's ability to pinpoint the user's focus of attention and interest (e.g., on products or exhibits), hence improving the ability to model the user and better personalize the service offered (e.g., exhibit or product information, shopping assistance). According to Majaranta et al. [61], "Advances in the technology open new areas for eye tracking, widening the scope of gaze-based applications. Current hot topics include all kinds of mobile applications and pervasive systems where the user's visual behavior and attention are tracked and used for eye-based interaction everywhere and at any time". In the coming years, when mobile eye-tracking technology becomes affordable, we envision that every person will have a device that employs this technology.

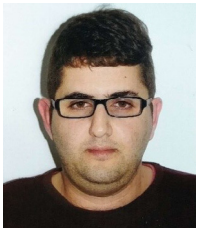
Future work will address improving the accuracy, speed, and interactive design of the mobile eye-tracker audio guide system, dealing also with potential errors and places where no information exists. Furthermore, we will explore the potential of using the mobile eye tracker as an intuitive pointing device in other scenarios, including indoor, outdoor, and urban scenarios. Another interesting future research direction can be to design an overall immersive museum experience, integrating additional novel technologies, such as the real-time generation of personalized coherent presentations, into spatial audio systems.

References

[1] A.A. Calvo, S. Perugini, Pointing devices for wearable computers, Adv. Human–Comput. Interact. 2014 (2014) 10 Article ID 527320.
[2] A. Bulling, T.O. Zander, Cognition-aware computing, IEEE Pervasive Comput. 13 (3) (2014) 80–83.
[3] A. Bulling, R. Dachselt, A. Duchowski, R. Jacob, S. Stellmach, V. Sundstedt, Gaze interaction in the post-WIMP world, in: Extended Abstracts on Human Factors in Computing Systems, CHI'12, ACM, 2012, pp. 1221–1224.
[4] R.J.K. Jacob, K.S. Karn, Eye Tracking in Human–Computer Interaction and Usability Research: Ready To Deliver the Promises, Elsevier Science BV, 2003.
[5] P.M. Fitts, R.E. Jones, J.L. Milton, Eye movements of aircraft pilots during instrument-landing approaches, Aeronaut. Eng. Rev. 9 (2) (1950) 24–29.
[6] K. Hendrickson, K.L. Ailawadi, Six lessons for in-store marketing from six years of mobile eye-tracking research. Shopper marketing and the role of in-store marketing, Rev. Mark. Res. 11 (2014) 57–74.
[7] M. Kassner, W. Patera, A. Bulling, Pupil: an open source platform for pervasive eye tracking and mobile gaze-based interaction, in: Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct Publication, ACM, 2014, pp. 1151–1160.
[8] R.J.K. Jacob, The use of eye movements in human–computer interaction techniques: What you look at is what you get, ACM Trans. Inf. Syst. 9 (3) (1991) 152–169.
[9] L. Ardissono, T. Kuflik, D. Petrelli, Personalization in cultural heritage: the road travelled and the one ahead, User Model. User-Adapt. Interact. 22 (1–2) (2012) 73–99.
[10] S. Stephens, The growth of mobile apps, Mus. Pract. (2010).
[11] S. Billings, Upwardly mobile, Mus. Pract. 46 (2009) 30–34.
[12] K. Cheverst, N. Davies, K. Mitchell, A. Friday, Experiences of developing and deploying a context-aware tourist guide: The GUIDE Project, in: Proc. 6th Annu. Int. Conf. Mobile Comput. Netw., ACM Press, New York, 2000, pp. 20–31.
[13] D.W. Hansen, Q. Ji, In the eye of the beholder: A survey of models for eyes and gaze, IEEE Trans. Pattern Anal. Mach. Intell. 32 (3) (2010) 478–500.
[14] M.V. Yousefi, E.P. Karan, A. Mohammadpour, S. Asadi, Implementing eye tracking technology in the construction process, in: 51st ASC Annual International Conference Proceedings, 2015.
[15] D.G. Lowe, Object recognition from local scale-invariant features, in: Proceedings of the Seventh IEEE International Conference on Computer Vision, Vol. 2, 1999, pp. 1150–1157.
[16] H. Bay, T. Tuytelaars, L. Van Gool, SURF: Speeded up robust features, in: Computer Vision–ECCV, 2006, pp. 404–417.
[17] S. Leutenegger, M. Chli, R.Y. Siegwart, BRISK: Binary robust invariant scalable keypoints, in: Computer Vision, ICCV, 2011 pp. 2548–2555.
[18] E. Rublee, V. Rabaud, K. Konolige, G.R. Bradski, ORB: An efficient alternative to SIFT or SURF, in: Computer Vision, ICCV, 2011, pp. 2564–2571.
[19] M.A. Fischler, R.C. Bolles, Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography, Commun. ACM 24 (6) (1981) 381–395.
[20] O. Chum, J. Matas, J. Kittler, Locally optimized RANSAC, in: Patt. Recog., 2003, pp. 236–243.
[21] K. Lebeda, J. Matas, O. Chum, Fixing the locally optimized RANSAC, in: British Machine Vision Conference, 2012, pp. 1–11.
[22] O. Chum, J. Matas, Matching with PROSAC- progressive sample consensus, in: Proc. IEEE Conf. Comp. Vision Patt. Recog, Vol. I, 2005, pp. 220–226.
[23] A. Brahmachari, S. Sarkar, Hop-diffusion Monte Carlo for epipolar geometry estimation between very wide-baseline images, IEEE Trans. Pattern Anal. Mach. Intell. 35 (3) (2013) 755–762.
[24] L. Goshen, I. Shimshoni, Balanced exploration and exploitation model search for efficient epipolar geometry estimation, IEEE Trans. Pattern Anal. Mach. Intell. 30 (7) (2008) 1230–1242.
[25] R. Raguram, O. Chum, M. Pollefeys, J. Matas, J.M. Frahm, USAC: a universal framework for random sample consensus, IEEE Trans. Pattern Anal. Mach. Intell. 35 (8) (2013) 2022–2038.

Please cite this article in press as: M. Mokatren, et al., Exploring the potential of a mobile eye tracker as an intuitive indoor pointing device: A case study in cultural heritage, Future Generation Computer Systems (2017), http://dx.doi.org/10.1016/j.future.2017.07.007.

[26] B. Tordoff, D. Murray, Guided sampling and consensus for motion estimation, in: European Conference on Computer Vision, 2002, pp. 82–98.
[27] A. Oliva, A. Torralba, Building the gist of a scene: The role of global image features in recognition, Prog. Brain Res. 155 (2006) 23–36.
[28] B. Zhou, A. Lapedriza, J. Xiao, A. Torralba, A. Oliva, Learning deep features for scene recognition using places database, in: Advances in Neural Information Processing Systems, 2014, pp. 487–495.
[29] M. Brown, S. Süsstrunk, Multi-spectral SIFT for scene category recognition, in: IEEE Conference on Computer Vision and Pattern Recognition, CVPR, 2011, pp. 177–184.
[30] M. Economou, The evaluation of museum multimedia applications: lessons from research, Mus. Manag. Curatorship 17 (2) (1998) 173–187.
[31] M. Weiser, The computer for the 21st century, Sci. Am. 265 (3) (1991) 94–104.
[32] P. Prekop, Paul, M. Mark Burnett, Activities, context and ubiquitous computing, Comput. Commun. 26 (11) (2003) 1168–1176.
[33] K. Huang, Challenges in human–computer interaction design for mobile devices, in: Proceedings of the World Congress on Engineering and Computer Science, Vol. 1, 2009, pp. 20–22.
[34] G. Brône, B. Oben, T. Goedemé, Towards a more effective method for analyzing mobile eye-tracking data: integrating gaze data with object recognition algorithms, in: Proceedings of the 1st International Workshop on Pervasive Eye Tracking & Mobile Eye-based Interaction, 2011, pp. 53–56.
[35] T. Pfeiffer, P. Renner, Eyesee3d: A low-cost approach for analyzing mobile 3D eye tracking data using computer vision and augmented reality technology, in: Proceedings of the Symposium on Eye Tracking Research and Applications, 2014, pp. 369–376.
[36] C. Ohm, M. Müller, B. Ludwig, S. Bienk, Where is the Landmark? Eye Tracking Studies in Large-Scale Indoor Environments, 2014, pp. 47–51.
[37] S. De Beugher, G. Brône, T. Goedemé, Automatic analysis of in-the-wild mobile eye-tracking experiments using object, face and person detection, in: Proceedings of the International Conference on Computer Vision Theory and Applications, VISIGRAPP 2014, Vol. 1, 2014, pp. 625–633.
[38] J. Schrammel, E. Mattheiss, S. Döbelt, L. Paletta, A. Almer, M. Tscheligi, Attentional behavior of users on the move towards pervasive advertising media, in: Pervasive Advertising, Springer, London, 2011, pp. 287–307.
[39] J. Schrammel, G. Regal, M. Tscheligi, Attention approximation of mobile users towards their environment, in: CHI'14 Extended Abstracts on Human Factors in Computing Systems, 2014, pp. 1723–1728.
[40] P. Kiefer, I. Giannopoulos, D. Kremer, C. Schlieder, M. Raubal, Starting to get bored: An outdoor eye tracking study of tourists exploring a city panorama, in: Proceedings of the Symposium on Eye Tracking Research and Applications, 2014, pp. 315–318.
[41] Y.I. Nakano, R. Ishii, Estimating user's engagement from eye-gaze behaviors in human-agent conversations, in: Proceedings of the 15th International Conference on Intelligent User Interfaces, 2010, pp. 139–148.
[42] K.T. Ma, Q. Xu, L. Li, T. Sim, M. Kankanhalli, R. Lim, Eye-2-I: Eye-tracking for just-in-time implicit user profiling. 2015. arXiv preprint arXiv:1507.04441.
[43] A. Hampapur, K. Hyun, R.M. Bolle, Comparison of sequence matching techniques for video copy detection, in: Electronic Imaging, 2002, pp. 194–201.
[44] T. Toyama, T. Kieninger, F. Shafait, A. Dengel, Gaze guided object recognition using a head-mounted eye tracker, in: Proceedings of the Symposium on Eye Tracking Research and Applications, ETRA'12, ACM, New York, NY, USA, 2012, pp. 91–98.
[45] E. Dim, T. Kuflik, Automatic detection of social behavior of museum visitor pairs, ACM Trans. Interact. Intell. Syst. (TiiS) 4 (4) (2014) 17.
[46] A. Bulling, H. Gellersen, Toward mobile eye-based human–computer interaction, IEEE Pervasive Comput. 9 (4) (2010) 8–12.
[47] I. Giannopoulos, J. Schöning, A. Krüger, M. Raubal, Attention as an input modality for Post-WIMP interfaces using the viGaze eye tracking framework, Multimedia Tools Appl. 75 (6) (2016) 2913–2929.
[48] S.S. Yalowitz, K. Bronnenkant, Timing and tracking: unlocking visitor behavior, Visitor Stud. 12 (2009) 47–64.
[49] J.H. Falk, John L.D. Dierking, Learning from Museums: Visitor Experiences and the Making of Meaning, Altamira Press, 2000.
[50] G. Farnebäck, Two-frame motion estimation based on polynomial expansion, in: Scandinavian Conference on Image Analysis, in: LNCS, vol. 2749, Springer, Berlin Heidelberg, 2003, pp. 363–370.
[51] S.K. Card, G.G. Robertson, J.D. Mackinlay, The information visualizer: An information workspace, in: Proc. ACM CHI'91 Conf., 1991, pp. 181–188.
[52] N. Newman, Apple iBeacon technology briefing, J. Direct Data Digit. Mark. Pract. 15 (3) (2014) 222–225.
[53] B.D. Lucas, T. Kanade, An iterative image registration technique with an application to stereo vision, in: Proceedings of Imaging Understanding Workshop, 1981, pp. 121–130.
[54] J. Brooke, SUS-A quick and dirty usability scale, Usability Eval. Ind. 189 (194) (1996) 4–7.
[55] J. Corbin, A. Strauss, Basics of Qualitative Research: Techniques and Procedures for Developing Grounded Theory, fourth ed., Sage, 2015.
[56] J. Lanir, T. Kuflik, A.J. Wecker, O. Stock, M. Zancanaro, Examining proactiveness and choice in a location-aware mobile museum guide, Interact. Comput. 23 (5) (2011) 513–524.
[57] M. Cohen, I. Shimshoni, E. Rivlin, A. Adam, Detecting mutual awareness events, IEEE Trans. Pattern Anal. Mach. Intell. 34 (12) (2012) 2327–2340.
[58] T. Kuflik, J. Lanir, E. Dim, A. Wecker, M. Corra, M. Zancanaro, O. Stock, Indoor positioning in cultural heritage: Challenges and a solution, in: Electrical & Electronics Engineers in Israel (IEEEI), 2012 IEEE 27th Convention of, 2012, pp. 1–5, IEEE.
[59] J. Lanir, T. Kuflik, E. Dim, A.J. Wecker, O. Stock, The influence of a location-aware mobile guide on museum visitors' behavior, Interact. Comput. 25 (6) (2013) 443–460.
[60] I. Beja, J. Lanir, T. Kuflik, Examining factors influencing the disruptiveness of notifications in a mobile museum context, Hum.–Comput. Interact. 30 (5) (2015) 433–472.
[61] P. Majaranta, A. Bulling, Eye tracking and eye-based human–computer interaction, in: S.H. Fairclough, K. Gilleade (Eds.), Advances in Physiological Computing, Springer, London, 2014, pp. 39–65.



Moayad Mokatren is a Ph.D. student in the department of Information Systems at the University of Haifa, Israel. He received his bachelor's degree in Computer Science from Hadassah College, and his master's degree in Information Systems from the University of Haifa. His main research interests are in the fields of human–computer interaction in ubiquitous and smart environments.

Tsvi Kuflik is an associate professor and former head of the Information Systems Department at the University of Haifa, Israel. His main areas of research are Ubiquitous User Modelling and Intelligent User Interfaces. He received his B.Sc. and M.Sc. in computer science and Ph.D. in Information systems from Ben Gurion University of the Negev. Prof. Kuflik is the author of over 200 refereed publications in journals and conferences. For seven years, he is the co-organizer of the series of PATCH workshops focusing on the application of novel technology in cultural heritage and the chair and organizer of many other workshops and conferences, including being the general chair of IUI 2014, and IUI 2017, PC chair of UMAP 2014 and more. Tsvi is a distinguished ACM scientist and a senior IEEE member and the chair elect ACM SIGCHI IUI community.

Ilan Shimshoni received the B.Sc. degree in mathematics and computer science from the Hebrew University Jerusalem, Israel, the M.Sc. degree in computer science from the Weizmann Institute of Science, Rehovot, Israel, and the Ph.D. degree in computer science from the University of Illinois at Urbana-Champaign. Currently, he is an associate professor at the Department of Information Systems, University of Haifa. He served as the chair of the department for four years. His current research interests include computer vision, robotics, and computer graphics. He specializes in multiple view geometry and 3D shape analysis and its applications for the field of archeology. He also conducts research in the field of data mining. He is currently an associate editor of IEEE Transactions on Pattern Analysis and Machine Intelligence.

Please cite this article in press as: M. Mokatren, et al., Exploring the potential of a mobile eye tracker as an intuitive indoor pointing device: A case study in cultural heritage, Future Generation Computer Systems (2017), http://dx.doi.org/10.1016/j.future.2017.07.007.