

The effect of tracking technique on the quality of user experience for augmented reality mobile navigation

Yoones A. Sekhavat¹  · Jeffrey Parsons²

Received: 19 September 2016 / Revised: 23 April 2017 / Accepted: 3 May 2017 /
Published online: 12 May 2017
© Springer Science+Business Media New York 2017

Abstract Augmented reality (AR) technology is changing the way we interact with the world by making the journey seamless and interactive. This is done by layering digital enhancements over an existing reality or real life scenario. However, employing AR technologies in wayfinding and navigation does not automatically bring positive experiences. We argue that tracking techniques are important in mobile AR navigation and fundamentally affect the quality of user experience. In this paper, we propose two different tourist AR applications based on different tracking techniques. In addition to the analysis of the applications, we conducted a user evaluation to study the effect of different AR tracking techniques on the performance of users and the quality of user experience.

Keywords Augmented reality · Mobile navigation · AR tracking · User experience · Quality of user experience

1 Introduction

With the advent of Augmented Reality (AR), which makes it possible to create a composite view rooted in both real and virtual worlds, it is not strange to see people interacting with

✉ Yoones A. Sekhavat
sekhavat@tabriziau.ac.ir
Jeffrey Parsons
jeffreyp@mun.ca

¹ Faculty of Multimedia, Tabriz Islamic Art University, Azadi Blvd, Hakim Nezami Square, Tabriz 5164736931, East Azerbaijan, Iran

² Faculty of Business Administration, Memorial University of Newfoundland, St. John's, NL A1B 3X5, Canada

things that are not part of the real world. AR is defined as a technology that allows the superimposition of computer-generated multimedia on a user's view of the real world [1]. Such data seems to co-exist with the real world and provides additional knowledge about the environment [10]. AR games like Niantic's Pokemon Go,¹ which gives players a reason to go out into the real world to catch virtual Pokemon, have quickly attracted widespread attention. This technology has considerable potential in the mainstream consumer space, in applications such as gaming [45, 55], health [22], and marketing [44, 46].

According to [52], tourism and travel are the most popular AR applications after games and entertainment. The emergence of mobile devices that bring together all the necessary technologies to sense the environment (such as GPS, compass and gyroscope) and technologies to render multimedia (such as 3D graphics, video, images and text) have made it possible for AR to enter the mass tourism market [29, 48]. AR mobile applications have changed the way people can experience a destination. These applications have the potential to increase awareness of one's immediate unknown surroundings [6] by providing a large amount of relevant information about points of interest (POIs) in a convenient and seamless way, as compared to conventional online search travel guides [31]. Much previous research has shown the potential of using AR to guide people in unfamiliar places through information augmented on real data. Finding points of interest, navigation, and receiving information about POIs are the most widely used features in mobile tourism applications [12]. Such applications not only reduce the mental effort required to navigate and discover unknown environments, but also provide travelers with complementary multimedia information about their surroundings.

As discussed in [57], employing AR technologies in wayfinding and navigation applications does not automatically bring positive experiences. Although the potential of AR has been shown in many previous studies, effective and usable AR application design is still in its infancy [9]. There are issues related to the design of AR applications and adaptation of this technology in mobile applications. One important issue is AR tracking, which fundamentally affects how users interact with the application. By *tracking* we mean the method of identifying a user's location and orientation in order to properly align virtual objects to the real world environment. This process of locating a user in an environment is critical to the accuracy of AR applications. The lack of wide adoption by the general public [39] is another important issue, as some people still prefer traditional forms of information such as paper-based maps and guidebooks.

In this paper, we study how different AR tracking techniques affect the quality of user experience in wayfinding. More specifically, we compare the effects of using location-based and marker-based AR on the performance and user acceptance of AR applications. In location-based AR, tracking and sensing of the environment is performed through geo-location information (e.g., GPS and compass data), while in marker-based AR, easily detectable predefined signs in the environment are used as specific labels to decide what information must be augmented on real data. We implemented a prototype of location-based augmented reality (LAR) and a prototype of marker-based augmented reality (MAR). We propose novel manual and automatic techniques to configure AR data based on the location of a user. To evaluate, we investigate how the tracking method affects the quality of user experience in visiting points of interest in TIAU (the main campus of Tabriz Islamic Art University, which is a historical site). To the best of our knowledge, this the first study

¹www.pokemongo.com

to evaluate the quality of user experience from an AR tracking technique point of view. In summary, the contributions of this paper are:

- Proposing user-friendly and informative location-based and marker-based AR applications for wayfinding.
- Developing novel techniques for manual and automatic configuration of POIs based on the clustering of these points in a walkable area.
- Directly observing tourists' interactions with our prototypes to study how the AR tracking technique affects user performance, measured in terms of errors in finding points of interest and time required to do so.
- Conducting a field experiment to study the effect of different AR tracking techniques on the quality of user experience.

2 Related work

In this section, we review previous research on AR navigation systems to indicate how our work contributes to the state-of-the-art. We review both outdoor and indoor navigation apps considering commercial and open source frameworks and applications.

2.1 Outdoor AR navigation apps

City Lens,² WAM,³ ARNav Geocaching,⁴ Layar,⁵ and Wikitude Navigation⁶ are among commercial apps for AR navigation. City Lens employs the camera of a mobile device to show nearby POIs, as well as information about them in AR style. Features of City Line include setting the line of site, storing favorite searches, and the ability to use the application in both portrait and landscape. This application also allows a user to select a category (e.g., food) to avoid showing unrelated items when looking for a particular item. WAM (World around me) is another AR Navigation app that adds useful information (e.g., names of places, ratings and distance) to all the places around a user. Features of this application include the ability to rank places, issue voice commands, and represent direction to POIs. ARNav Geocaching is an Android outdoor game app that shows directions in AR mode. This application makes it possible to select a search category to avoid cluttering screens full of POI tags. Layar combines visual cues captured by the camera with GPS information and compass to determine the location and the orientation of the mobile device. This information is used to overlay POIs on top of the picture coming from the camera. A commercial product, Wikitude Navigation, is a GPS-based guidance application that is integrated with around 7000 individual content providers to inform users of their surroundings. Radio Frequency Identification (RFID) is also used in AR navigation apps. In [8], an audio message is activated once an RFID tag embedded in a footpath is identified. The main problem of using RFID tags is delay incurred in identify tags. In [54], RFID tags enhanced with spatial coordinates are used to describe the pre-defined surroundings. The system proposed in

²<http://betalabs.nokia.com/trials/nokia-citylens-for-windows-phone>

³<http://worldaroundmeapp.com/>

⁴<http://arnav.eu/products-services/>

⁵<http://worldaroundmeapp.com/>

⁶<http://www.wikitude.com/>

[15] determines whether a POI is inside the field of view, independent of the rotation of the device. Nokia's MARA project [16] is an outdoor mobile AR application that employs a GPS and an inertial sensor to provide location and orientation.

2.2 Indoor AR navigation apps

The IZat system from Qualcomm⁷ is an indoor navigation and positioning system that employs Wi-Fi based calculations to localize devices. Google Indoor Maps⁸ uses an indoor navigation algorithm that exploits Wi-Fi access points and mobile towers to determine the location of a user. Place Lab [33] is a research prototype using radio beacons, information about beacon locations and clients, to estimate their location. NAVVIS⁹ is an indoor positioning system that works based on a large database of images of indoor places. Pictures taken by a user from his surroundings are compared with these images to determine the location and the orientation of a user. In InfSoft,¹⁰ Wi-Fi, Bluetooth and GPS information are used to localize objects. This information is used to present navigational information in AR style. In Theodolite,¹¹ compass, GPS, map, photo/movie camera, rangefinder, and two-axis inclinometer information are combined to determine the location and the orientation of a user. In the viewfinder of this application, information about position, altitude, bearing, range, and inclination are shown on the live camera. In [50], local descriptors are used to recognize and track object placement. This is performed by extracting features from a video frame and then matching them against existing features. Image matching is also used in [56] to determine the location of a user. Image retrieval algorithms based on robust local descriptors are used by [51] to locate the user in AR Navigation. In this system, captured images are matched against a database of highly relevant features. In PhoneGuide [4], a neural network is trained to recognize colors employed as a museum guide.

2.3 Presenting navigation information

Maps have a long tradition in navigation apps. Unlike AR navigation apps, the problem with text-based navigation information is the difficulty of reading text while walking. In [2], a system is proposed that combines textual instructions, maps and geo-tagged photos for navigation. In [21] geo-tagged photos are augmented with 2D directional arrows. However, as discussed in [42], this system suffers from mismatching photos to the real environment. Similar to our work, they proposed a system that renders several views of augmented images of landmarks with 3D directional arrows. In [23], a system is proposed that augments maps with street scene photographs. According to [36], participants prefer an AR view to a map view. Using eye tracking technology, they found that the AR view is mainly used at complex decision points (before and after road intersections). In [25], the potential of haptic-enabled navigation applications in comparison to visual and audio-based applications is studied. They showed that, although elderly people may benefit from audio navigation, this system has several shortcomings as it relies too much on street names. In the Navig system [30], visually impaired people are guided through semantic audio rendering. AR view is also used

⁷<https://www.qualcomm.com/products/izat>

⁸<https://www.google.com/maps/about/partners/indoormaps/>

⁹<http://www.navvis.com/>

¹⁰<https://www.infsoft.com/>

¹¹<http://hrtapps.com/theodolite/>

in Navigation apps to guide elderly people in wayfinding, in which physical information from objects are transformed into an AR view [20]. The findings of [14] show the advantages of representing three-dimensional virtual objects in a realistic environment to improve the wayfinding process of patients. Unlike NAVVIS and InfSoft, which just show arrows on the screen to help navigation, our prototypes overlays POIs on the live camera based on user's context. Our prototypes also provide a rich user experience through a user friendly interactivity with virtual objects on AR view.

A comparative table highlighting the organization of existing approaches for AR navigation is shown in Table 1. Summarizing previous work in navigation apps, it is not clear which tracking parameters affect navigation performance and user experience. Although subjective rating for maps is typically better, recent studies have shown promising results for AR and photograph-based navigation [19]. On the other hand, navigation performance in an auditory mode is not well considered by users.

3 Marker-based vs. location-based AR

In terms of tracking and sensing the environment to decide what information to augment, mobile AR applications are classified into location-based and marker-based [27]. Location-based AR (a.k.a, marker-less or position-based AR) is generally used in outdoor environments, where compass, accelerometer, and GPS data are used to identify the location of a user. Then, based on this geo-location data, corresponding multimedia information is retrieved and rendered over the camera view. In location-based AR applications in the field of tourism, users are able to explore their surroundings by adding new layers of synthesized data to reality. Arusma,¹² and Layar¹³ are examples of location-based AR applications that employ geo-location information to recognize the surrounding environment and augment information on real images captured from the camera without the need to scan target images. In contrast, marker-based AR is generally used indoors, where easily detectable and predefined signs in the environment are used as specific labels to register the position of data that is augmented [7, 34]. In this technique, the marker can be an image, where the image recognition algorithms compute the rotation and translation of the image detected in relation to the camera of the smartphone. This indicates where to augment the multimedia data.

In addition to target images used as markers in marker-based AR, 3D object tracking techniques may also be used. Such systems rely on natural features instead of fiducial marks. For example, [37] employed a descriptor of the line segment feature in 3D objects to compute the camera pose estimation. However, due to the complexity of tracking, which compromises accuracy, this tracking technique is not widely used. In some hybrid approaches, such as [35], computer vision techniques are used to improve accuracy in identifying POIs. In particular, GPS and compass data are integrated with the spatial relation between the reference images and the POIs, assisting and correcting the GPS data. In [43], several tracking sensors, including edge-based tracker, gyroscope and measurements of gravity and magnetic field were integrated. Although hybrid tracking techniques are reported as robust and accurate, they are expensive and require complex computations.

In terms of interaction, location-based AR provides a more interactive and dynamic experience than marker-based AR, which depends on a certain label and restricts the movement

¹²<http://www.aurasma.com>

¹³<https://www.layar.com>

Table 1 A comparative table highlighting the organization of existing approaches for AR navigation

	AR Representation	Finding locations	Specific Features
City Lens	POI tags, Information about POIs, Directions	Mobile data / wifi	Line of site for visible POIs, Storing favorite searches, Both portrait and landscape
WAM	Names of places, Ratings, Distance and directions	GPS / wifi	Rank places, Issue voice commands, Search in categories
ARNav Geocaching	Direction to POI, Turn-by-turn navigation route presented as virtual arrows	GPS and inertial sensors	Record your track in GPX and later replay it
Layar	Direction to POI, Favorite route direction	Compass, GPS and markers	Compass in augmented reality mode
Wikitude Navigation	Direction to POI, 3D models in the augmented reality scene	Image recognition and tracking technology	Integrated with content providers to inform users of their surroundings
RFID-based system [8]	Voice is used to guide the users to navigate	Using RFID to activate audio messages Delays to identify tags	The tags store location information that is within a proximity range
Nokia's MARA	Restaurants menu, a persons blog, historical information about a building	GPS and inertial sensors	Using RFID to describe the pre-defined surroundings
IZat	Show who and what are around you, whether you're indoors or out	GPS, GLONASS, and BeiDou satellites, Wi-Fi	Gives more directions & context Finding friends located in Android's Location settings
Google Indoor Maps	Zoom in to see the indoor floor plan of a building	wifi and mobile towers	Search within the building, Easily recognizable symbols represent different POIs
Place Lab [33]	POI tags, Information about POIs, direction to POI	Radio beacons, information about beacon locations	Cell phones estimate their position by listening for the cell IDs of radio beacons
NAVVIS	POI tags Information about POIs, direction to POI	Large database of images of indoor places	Indoor navigation without Wifi or beacons Access your data on any device
InfSoft	POI tags, Information about POIs direction to POI	Bluetooth beacons for indoor positioning	Special hardware for indoor tracking
Camera tracking [50]	-POI tags, Information about POIs direction to POI	Identifying local descriptors matching them against features	The system can handle scenes with moving elements, occlusions and illumination changes
Camera pose [21]	A directional arrow to display dynamic information of surroundings	Determining landmarks and matching them	Uses the location estimate to select the search area in the floor plan
Navig system [30]	The system integrates an adapts geographic information system	Combining satellite, image, and other sensor information	Optimize performance for visually impaired users

Table 1 (continued)

	AR Representation	Finding locations	Specific Features
MAR	POI tags and Information about POIs as well as directions	Uses markers to indicate locations and the directions	Can be used indoor and outdoor Separate AR view and Map view
LAR	POI tags and Information about POIs as well as directions	GPS, compass, accelerometer, gyroscope	Used indoor and Uses automatic clustering of POIs to determine what is be augmented

of the user [34, 41]. However, various difficulties have hindered the full exploitation of the potential of location-based approach. First, as discussed in [26], marker-less AR is resource-intensive, as it pulls power on the mobile devices. This is due to the fact that using tracking sensors, including GPS compass and accelerometer, is expensive [27]. Second, location-based AR is limited due to sensor accuracy. This can lead to poor alignment of real and augmented objects with each other. For example, the wrong information may be augmented on the images of two restaurants with different ratings that are located beside each other. On the other hand, marker-based applications recognize objects more accurately. As discussed in [28], marker-based capture systems are quite popular due to efficiency and accuracy.

We argue that tracking is a major research issue in AR, and depending on the application and the field of use, different tracking techniques might be more appropriate. The focus of this research is to study how different AR tracking techniques might affect the user performance and experience in the field of tourism.

4 Theoretical background and hypotheses

User experience (UX) is the experience a product creates for those who use it. According to the ISO's definition [24], UX focuses on a person's perception and the responses resulting from the use of a product or service. Accordingly, the quality of experience is a measure of a user's experiences with a product or service

Different factors affect the quality of user experience using AR applications; among these we focus on interaction between users and physical environments. In the AR literature, this issue has been studied in terms of "Presence" [3], "Sense of Place" [5] and "Situation Awareness" [13]. In this section, we elaborate these concepts and discuss how they affect the overall user experience. In AR applications, tracking technique is an important issue in mobile AR navigation that fundamentally affects all of these parameters.

In digital tourism, the purpose of AR is mixing the real world with digital content to enhance the experience of visitors [3]. A critical question for AR in the field of tourism and navigation is "Can AR technology result in feeling more presence in experiencing a place?" If we answer this question according to the definition of presence by Lombard and Ditton (1997), which is "disappearance of mediation" or "perceptual illusion of non-mediation", any technology is a barrier, and achieving presence requires removing the sense that technology mediates the experience. However, according to several researchers, AR has been able to increase user satisfaction. We argue this is due to the fact that AR is able to add layers of information to enhance experiencing the place, where this information is not

available in traditional experience. In other words, although AR requires mediation to see the world, it enriches the information, resulting in a higher level of presence.

As discussed in [3], a system that anticipates the needs of users at run-time, but that does not require a significant cognitive effort to perform the tasks, can increase the sense of presence. We argue that, by increasing the sense of presence, users can intuitively transform their intentions to actions. This can increase user performance in executing the tasks. We argue that LAR, by automatically representing relevant information regarding the context without requiring the user to take any action, allows users to perform tasks more efficiently with fewer errors. However, MAR requires manually scanning the markers without anticipating the needs of users. As a result, MAR is expected to require more cognitive effort than LAR to experience the environment. Consequently, we expect to achieve better performance to perform navigation using the LAR tracking technique than using the MAR tracking technique. We argue this is due to the flexible and intuitive interaction between users and the environment in LAR. Specifically, we posit hypotheses H1 and H2 in terms of performance as follows:

H1: Participants will take less time to find and visit POIs using LAR than using MAR.

H2: Participants will make fewer errors in finding and visiting POIs using LAR than using MAR.

Sense of place, which is defined as the feelings of attachment that people develop regarding a place [5], is also an important factor in increasing the quality of user experience in visiting a location. To achieve this goal in our applications, coupling AR information and the actual environment is crucial. We argue that, in the case of LAR, this coupling is conducted in a flexible and interactive way, as there is no need to be in a specific location to observe the AR information. Based on these foundations, we posit hypothesis H3:

H3: Participants will report a higher level of perceived quality of AR user experience using LAR than using MAR.

The level of user acceptance and the use of technology strongly depends on the usefulness of the system being used. In spite of tremendous capabilities of mobile AR navigation applications that define new forms of interaction between users and their surrounding environment, some mobile AR applications have been criticized as “useless” [40]. Although much research has addressed the technical and perceptual issues of mobile AR navigation, “situation awareness” is not studied sufficiently. Understanding situation awareness in the context of mobile AR-navigation can guide the design of mobile AR guidance applications. In [13], “situation” is defined as “a set of environmental conditions and system states with which the participant is interacting that can be characterized uniquely by a set of information, knowledge, and response options” (p. 34). Accordingly, “awareness” in this context is the information associated with a “situation” [13]. In the context of mobile AR navigation, users explore and navigate to a destination by interacting with their surroundings. Situation Awareness theory can provide a framework for understanding the way people interact with the environment that can help in designing mobile AR to assist users interactions. The implications of considering situational awareness are effective interaction of users with the surrounding environment and real-time representation of augmented data, both of which directly affect the usefulness of a system. In terms of these parameters, LAR is expected to be more successful as this method shows related computer-generated context once a user reaches a specific location. Based on these foundations, we posit hypothesis H4 as follows:

H4: Participants will report a higher level of subjective acceptance and intention to use the AR technology using LAR than using MAR.



Fig. 1 Additional information describing a point of interest is augmented on the image captured by the camera which is shown in a callout

Although the accuracy of GPS depends on many conditions such as ionospheric effects, according to Official U.S. Government information about the Global Positioning System,¹⁴ in the worst case, this system can achieve accuracy of 7.8 meters at a 95% confidence level. We argue this level of accuracy does not negatively affect our location based system for two reasons. First, generally the average distance between POIs is considerably more than this. For example, in the case of our experiments, the minimum point-to-point distance between any two POIs is 36 meters and the average is 58 meters. Second, in our application, the AR view works based on AR rules (expressions that states what information must be augmented on the camera view given the current location and orientation of a user). Each rule is assigned to an area (i.e. a cluster), which is triggered when a user stands in that area. Since a user is assigned to a cluster based on her distance from the center of that cluster, this level of accuracy does not negatively affect the overall accuracy of LAR.

5 Implementation and research methodology

We propose working prototypes of LAR and MAR wayfinding applications. LAR and MAR are two Android mobile applications that have exactly the same user interface with a difference in tracking technique. The primary goal of these applications is providing support for the visual exploration of POIs. Using these applications, users can easily find POIs and gain information about them. This is achieved through two important features. First, once looking through the camera, POIs and complementary information regarding them are augmented on top of the image captured from the camera (See Fig. 1). Second, the application directs users to target POIs by providing visual signs, as illustrated in Fig. 2. This section describes various features of MAR and LAR and provides scenarios of using these applications in the field of tourism.

¹⁴<http://www.gps.gov>

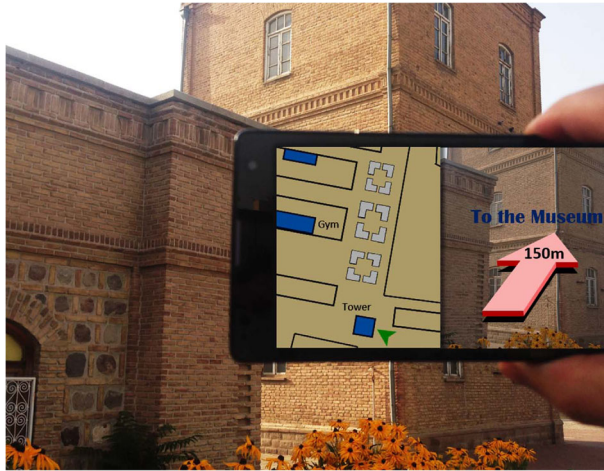


Fig. 2 A 3D arrow representing the direction to a point of interest (Museum in this case) is augmented on the image captured by the camera

5.1 User interface and interaction

Two main parts of the user interface are camera view and map view. The map view provides a map of the current location, including walkable paths, buildings, and POIs marked on the map. A user is able to filter and select POIs. On the other hand, the camera view shows the image captured by the camera augmented with the titles of POIs. At any time, the complete information about a POI is shown in the camera view. In the case of searching for a point of interest, directions to that point are presented on the camera view. Map view and camera view are shown simultaneously on the smartphone screen to provide overview and details at the same time. However, users can hide either of these views to have more space for the other view. In terms of interaction, this architecture supports users in following Shneidermans’s advice which is “overview first, zoom and filter, then details on demand” [49].

5.1.1 Map view

The map view represents a network of paths a user can move and navigate by walking or driving, and a set of POIs shown on them. The map view provides a “big picture” overview of the current position, as well as the locations of POIs. Users are able to identify POIs they are interested in visiting. The current position of the user and the direction in which a user is heading are shown on the map view. POIs are selectable in this map, where the selected POI starts to blink after selection. The map view, including all details, is shown in Fig. 3. For the map view, one alternative option could be using Google Maps¹⁵ and corresponding APIs. However, the level of detail is not the same in all places around the world. In our case, TIAU campus is shown as a brown area on the Google Maps, and there is no information about the buildings and walkable areas. For this reason, we implemented our own map view of this campus.

¹⁵<https://maps.google.com>

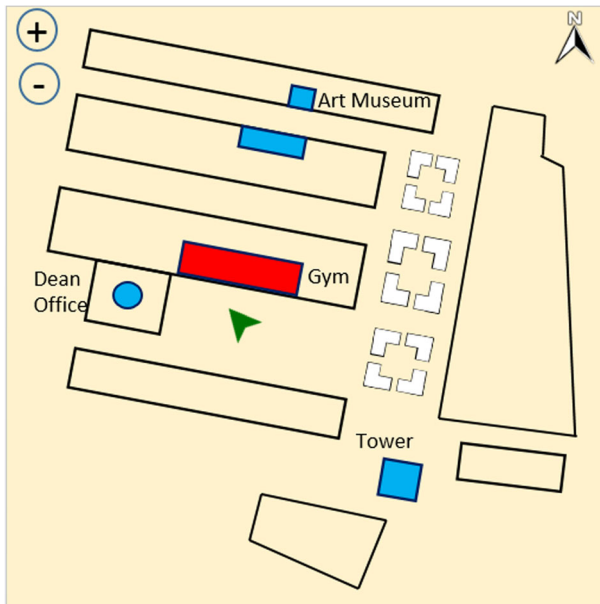


Fig. 3 An snapshot of map view representing POIs and their locations on the map

Visual encoding Numerous studies in the field of information visualization have shown that good color coding is an effective way to reduce visual search time. In our prototypes, colors are selected by following the opponent process theory of colors [18]. According to this theory, red vs. green, blue vs. yellow and black vs. white are three opponent channels, where the human visual processing system can easily differentiate between colors in each channel in comparison to other color pairs. Based on this theory, yellowish colors are used as a background of the map view, while POIs are shown in blue. Selected POIs are shown in pale blue while unselected POIs are shown in dark blue. The titles of POIs are shown near each POI position on the map. The size of the font of titles is compatible with the zooming scale. However, there is a limit for maximum and minimum font sizes to prevent extremely large or small text on the screen. In the case of LAR, an icon is added to indicate if the GPS and compass sensors are active. This helps to decrease confusion and improve the usability of the application.

Interaction Interaction is an important element of any visualized mobile application. Minimizing cognitive overload is a key factor in interacting with mobile tourism applications. In LAR, when a tourist is moving the application automatically provides updated and relevant content regarding the geo-location information of the users. LAR and MAR allow users to interact with the system to gain required information. Map view allows conventional map navigation features, including panning and zooming. This way, a user is able to explore the overview of the environment to find the POIs. Zooming is available only in five predefined scales.

Users have three options to select a POI to discover. As an intuitive action, users can select a POI on the map by simply clicking on the POI on the map. This way, the color of selected POI turns from dark blue to pale blue. The drawback of this method is that only nearby POIs which are visible on the map are selectable. To address the problem, two other

options are provided for the users. First, by tapping on the top of the map view, a search box appears allowing a user to search for a POI. This way users can look for a POI that is not currently on the map view. Second, users can select the POI from the list of POIs sorted alphabetically or based on the minimum distance to the current location of the user. These two options allow exploring and finding POIs that are not currently visible in the map view.

5.1.2 Camera view

The camera view is the AR view of our prototypes in which synthesized data in from various multimedia is shown on the image captured by the camera. The goal of the camera view is mixing real and virtual information to enhance the perception of a user from the current location and POIs. In LAR, which is activated and tracked using geo-location data, once the smartphone is held in front of a POI, the title of that POI is shown in the screen (See Fig. 4). If needed, a user can tap on the information icon near the title of the POI to see complementary information regarding this POI including description, a slide show of photos, or a video (See Fig. 5). Another important feature of the camera view is showing the direction to the selected POI using 3D visual signs and arrows augmented on the captured image (See Fig. 6). In LAR, by turning the device to different directions, the visual signs are changed according to the compass data received from the device. In addition, the distance to a final POI is shown nearby the visual signs. Drawing the visual signs that begin from the bottom of the screen and ends at the middle of the screen, provides a good sense of the direction a user should move in to reach the destination. By default, directions from the current position to the closest POI is shown in the camera view once the show navigation option is activated.

Visual encoding In the camera view, the size of text is used to encode the distance between the current location of a user and the point of interest the user sees through the camera view (See Fig. 4). In particular, closer POIs are shown with larger fonts than those that are located further away. This method provides a sense of how far the user is from the selected POI tagged in the camera view. Within the camera view, where the title of POIs are shown on the image captured by camera, different font colors are used to show the title of POIs.



Fig. 4 Points of interest (POIs) are tagged on the camera view when looking at them through the camera of the device



Fig. 5 The description and a slide show of photos of a POI are shown on the screen once tapping on the information icon of the POI

However, to avoid confusion in finding directions, the same pink color is used for visual signs no matter how far is the selected POI. For the same reason, there is no zooming option in the camera view.

Interaction In LAR, when a user points the device towards POIs, additional virtual information is overlaid on top of the real-world camera view through virtual tags. By tapping on the POI tags on the screen, a short text describing the POI is shown in a callout with a light background. In addition, depending on the data available for the POI, a slide show of photos, a video, or a 3D model is shown near the text information. Displaying multimedia data

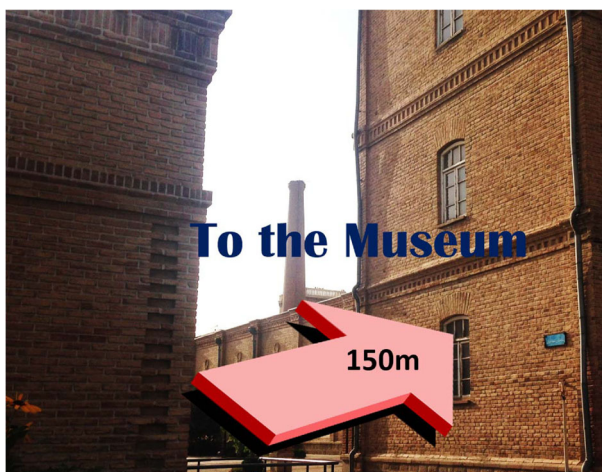


Fig. 6 A 3D arrow representing the direction to a point of interest (Museum) is augmented on the image captured by the camera

explaining the details of a POI provides a better perception of the point of interest. Tapping on a POI on the camera view also selects the current POI. Then, a user can select navigate button to show the direction to that POI.

5.2 Architecture

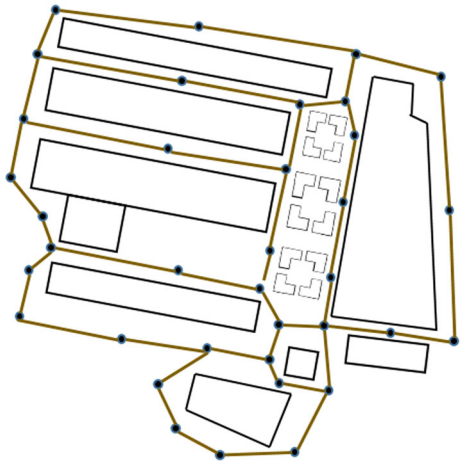
In MAR, once the scanned marker is recognized, since the location of the marker is known, the current position of the user on the map view is identified. In addition, since the camera directly points towards a marker, we know in which direction the user is facing. However, for LAR, GPS data must be captured to indicate the current position of the user. In addition, compass data must be retrieved to determine the direction in which the user is heading. This information is crucial in determining what multimedia must be represented. In practice, a user may stand in any point in the walkable area and look to his surrounding from any viewing angle. As a result, the system must be able to indicate what POIs can be seen in each position. Due to the large number of possible positions, we must find a way to handle this large space of points. We propose and use manual and automatic techniques to reduce the large space of points, as elaborated in the following.

5.2.1 Manual configuration

In the manual configuration, hypothetical paths in the walkable area that form a network (graph) of paths are created (See Fig. 7). The idea behind this technique is to assigning any point in the walkable area to the closest edge in the path graph. An edge in this graph represents an area where, from any point on this edge, POIs can be seen from almost the same viewing angles. Ideally, different viewing angles must be set to each point in the walkable area. However, this edge mapping technique makes it possible to handle the complexity of unlimited points in the walkable area.

In this technique, the position of the user on the walkable path is mapped to one of the edges in this graph at runtime, independent of the distance from this point to existing POIs. The walkable path network, is a graph denoted $G = (N, E)$. In this graph, N is a set of predefined nodes, where each node is a geographical point for which GPS data is already

Fig. 7 The network of walkable paths for the TIAU campus based on the 2D map of this campus including all nodes and edges



known. E is a set of edges between these nodes, where each edge represents a walkable path between two points. Working on a 2D map, each point p is a pair $p(i, j)$, where i is the latitude, and j is the longitude.

In MAR, the target image indicates what information must be shown on the camera view. However, in LAR, the position of a user as well as the direction this user is facing (i.e., measured in terms of the deviation angle from north) indicates what to show. In the latter case, this can be defined in terms of a set of AR rules. An AR rule is an expression that states what information must be augmented on the camera view given the current location of a user and the direction this user is facing. However, because of the large space of possible combinations of points, and facing angles, it is not practical to assign a rule for each combination of (latitude, longitude) and facing angles. To address this problem, we used an *Edge* map function to reduce the large space of points defined as follows:

Definition 1 (Edge Map Function) An edge map function is a function that maps a point $p(i, j)$ to the closest edge in the network of walkable paths, denoted:

$$Edge(p(i, j)) = \operatorname{argmin}_{e_i \in E} distance(e_i, p(i, j))$$

In this definition, the minimum distance is acceptable only when it is less than a predefined threshold ($distance(e_i, p(i, j)) < t$). Otherwise, $p(i, j)$ is not assigned to an edge. Figure 8 shows an example of how a point in different position is mapped to an edge in the walkable path network.

Another simplification technique we used is that, instead of stating an AR rule for any deviation angle, a rule is stated for a range of angles. This is a reasonable assumption since a small change in the deviation angle does not affect what information to show. For example, standing in a road in front of a building, after a small change in the deviation angle (e.g., between -20 and 20 degrees), we are still looking the same building, and the title of that building must be shown on the screen. We argue this technique not only simplifies the data entry process, but also improves the performance of the application.

Definition 2 (Tag Show Rule) A tag show rule is a quadruple $(e_i, min_ang, max_ang, L)$ representing what POIs L must be shown on the screen if the corresponding edge of the current position of the use is e_i and deviation angle from North is between min_ang and max_ang degrees.

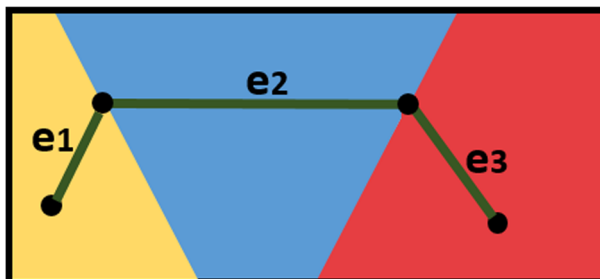


Fig. 8 Given the network of walkable paths (including 4 nodes and three edges in this figure, the *Edge* map function maps any point in the yellow section to e_1 , any point in the blue section to e_2 , and any point in the red section to e_3

We define $PI = \{l_1, \dots, l_k\}$ as a finite set of all POIs, where each $l_i \in PI$ has a title, position $p(i, j)$, description, and a set of images or videos. TSR is denoted the finite set of all tag show rules in the system.

Definition 3 (Navigation Show Rule) A navigation show rule is a quadruple $(e_i, min_ang, max_ang, S)$ representing what navigation signs S must be shown on the screen, if the corresponding edge of the current position of the user is e_i , and deviation angle from the north direction is min_ang and max_ang degrees.

Navigation signs $SN = \{s_1, \dots, s_n\}$ are a finite set of 3D arrows used to guide a user depending the current position of the user and the direction the user is heading. NSR is the finite set of all navigation show rules in the system.

We define $Tags$ as an AR functions indicating what POIs must be tagged in the camera view given the current position of the users and existing tag show rules.

$$Tags(p(i, j), a, tr) = \{l_i \in PI | \exists r \in TSR, Edge(p(i, j)) = e_i, r.min_ang < a < r.max_ang, dist(p(i, j), l_i) < tr\}$$

In the $Tags$ function that represents the main functionality of our prototypes, $p(i, j)$ is the current position of the user, a is the deviation angle of the user from North, and tr is the maximum threshold distance between the user and the point of interest that can be shown.

We also define $Navs$ as an AR function indicating what navigation signs must be augmented on the camera view given the current position of the users as well as existing navigation show rules.

$$Navs(p(i, j), a) = \{ns \in NS | \exists r \in NSR, Edge(p(i, j)) = e_i, r.min_ang < a < r.max_ang\}$$

Manually forming the graph of paths including different edges and then defining tag show rules by manually assigning visible POIs from each edge is a time consuming process that may hinder the application of this technique in practice. In the prototype we propose, these rules are manually created for a small historic site, while it might be impossible to perform this process manually for a large area. This issue becomes even more problematic when the person who manually creates tag show rules must take into account more than one POI. To address this problem, we propose a clustering based technique which is elaborated in the following.

5.2.2 Clustering based configuration

To address the problem of manually creating the network of walkable paths, as well as tag show rules, we propose an alternative solution based on clustering. In this technique, the walkable area is partitioned automatically to different areas (clusters), where the same tag show rules are assigned to each cluster. In this technique, instead of mapping the position of a user to an edge in the path graph, the position of the user indicates to which cluster it belongs. Consequently, the tag show rules of that cluster are extracted and applied for the user in that position.

In comparison to the edge map function, clustering the walkable area is performed automatically, so it can take into account any number of POIs. We use the k-means algorithm for clustering the points of a walkable area. Defining a proper distance function to measure the difference between points is crucial to form the clusters in this algorithm. We partitioned the walkable area into squares one square meter in size, where the collection of these

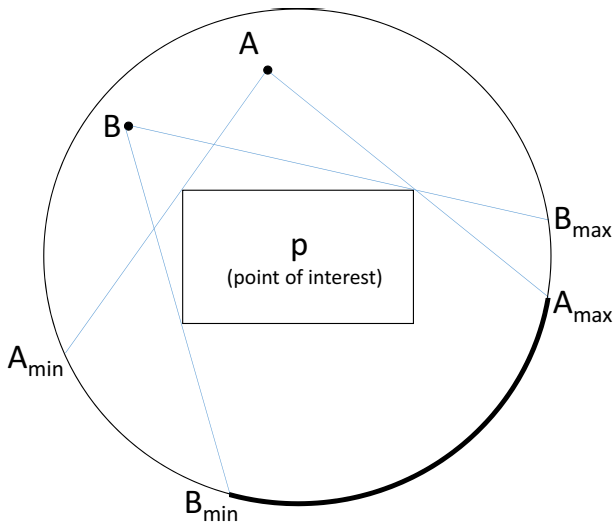


Fig. 9 Given two points A and B , and the hypothetical rectangle around the point of interest (p), $\widehat{B_{min}A_{max}}$ is the overlap between the viewing angle of p from A and B

squares are the items that are clustered. In the clustering algorithm, each point was assigned to a cluster centroid that yields the least within cluster sum of squares. In the standard k-means algorithm, the euclidean distance function is used to measure the distance between points. However, in our technique, we had to use a different distance function that takes into account the overlap between viewing angles for a point of interest from two points.

Given a point of interest (p) in Fig. 9, and two points A and B , the distance between A and B is computed as:

$$d(A, B, p) = (\text{overlap between } \widehat{A_{min}A_{max}} \text{ and } \widehat{B_{min}B_{max}}) / (\widehat{A_{min}A_{max}} + \widehat{B_{min}B_{max}}) / 2$$

In this formula, $d(A, B, p)$ is a distance between two points A and B given a single point of interest I . In the case of more than one point of interest, we need to find the accumulative distance, which we define as:

$$dist(A, B, P) = \sum_{p_i \in P} d(A, B, p_i) / |P|$$

where P is a set of all POIs. In order to simplify the shape of the POIs to find the overlap angles, we assume there is a hypothetical rectangle around any arbitrary shaped point of interest $p_i \in P$. This way, we simplify determining the minimum and maximum viewing angles. In particular, to compute max and min viewing angles for a given point of interest, all intersection between the lines from that point to all corners of that rectangle are calculated. Given two points $p_1 = (x_1, y_1)$ and $p_2 = (x_2, y_2)$, the angle between the horizontal line and the line passing these points is calculated as:

$$\text{atan2}(y_2 - y_1) / (x_2 - x_1)$$

Using the k-means clustering algorithm and the distance function ($dist$), we cluster the collection of points in the walkable area. Determining the number of clusters is a challenging issue in clustering that might significantly affect the application of this technique. To show the effect of the number of clusters and the number of POIs on clustering, we performed a

set of experiments. The visual representations of the clusters for 1, 2 and 3 POIs for 10, 20 and 30 clusters are shown in Figs. 10, 11 and 12.

To find the appropriate number of clusters as an input to our clustering algorithm, we performed a set of experiments to tune the number of clusters based on the different POIs. To this end, we need a measure to evaluate the quality of the clustering regarding the number of clusters. Since any point in the cluster is finally assigned to the centroid of that cluster, we measure accuracy of a cluster as an average distance between any points in that cluster from the centroid of that cluster.

$$AvgError(C) = \sum_{c_i \in C} \sum_{p_i \in P} dist(p_i, o) / |P|$$

where C is the collection of all clusters. P is the collection of all points in c_i , and o is the centroid of the cluster c_i . To avoid screen cluttering because of a large number of POI tags on the screen, we limited the maximum number of POIs that can be shown on the screen to 4 points. Given 1, 2, and 4 POIs, we varied the number of clusters from 5 to 50, and we measured the accuracy of clustering based on *AvgError* (Figs. 13 and 14).

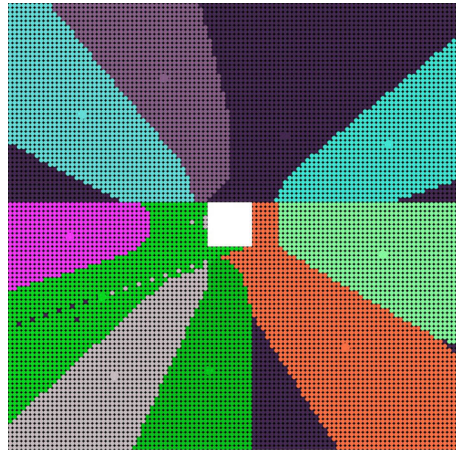
One might argue real time computing minimum and maximum viewing angles and dynamically indicating visible POIs would be more accurate than assigning any points in the cluster to the viewing angles of the centroid point. Although this is possible, it is not viable due to heavy processing that must consider all POIs. The advantage of the clustering technique is that all computations are done off-line, while at runtime only corresponding rules are extracted from a hash table based on the position of the user.

5.3 Implementation

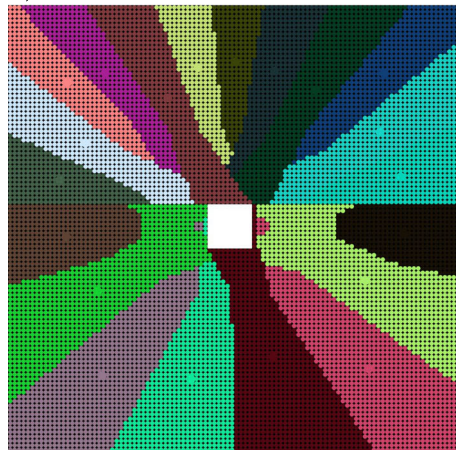
Developing AR rules (including tag show rules and navigation show rules), as well as implementing *Tags* and *Navs* functions, has been performed in the implementation phase. LAR is an android application implemented in Java using off the shelf libraries including the Ketai library to capture sensor data such as GPS data and compass data. This prototype was developed in *Processing*, one of the powerful tools based on Java extended with many information visualization libraries. We also used the video libraries of *Processing* to process images captured by the camera. We employed the *Qualcomm Vuforia* module in the *Unity* game engine to develop MAR. This module was selected to implement MAR because it offers the option of using a wide variety of customized markers. MAR and LAR have exactly the same user interface. Once either MAR or LAR is launched, the same window, including map view and camera view, is shown on the screen. Geo-locations of landmarks as well as nodes in the network of walkable paths were retrieved manually from <http://www.mapcoordinates.net> by students recruited for this purpose. AI features were implemented using behavior trees [47].

In the case of MAR, markers used for tracking play an important role in engaging users. Assigning appropriate images can improve the communication process. Instead of using binary black and white markers, a graphic artist was recruited to design attractive and informative markers based on the features of POIs. We argue using such graphic metaphors helps users to better understand the interface, functions and interactions. The logo of our application is placed on the top left corner of the markers representing that these are AR traceable markers. Markers were tested with different graphical features to ensure robust webcam recognition by the AR application. A total of 38 markers were installed in different places in TIAU campus. The locations of the markers were selected in a way that covers directions to all POIs from all edges in the network of walkable paths.

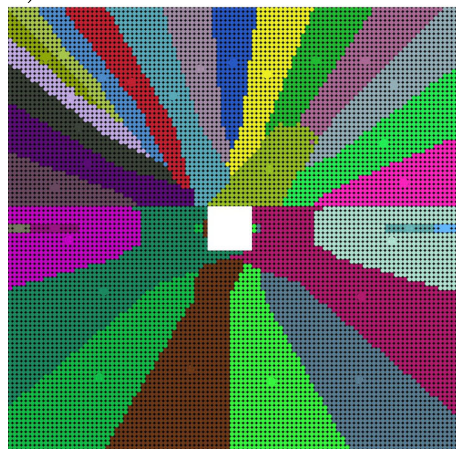
Fig. 10 Visual representation of clusters in single POI for 10, 20 and 30 clusters



a) 1 POIs and 10 clusters

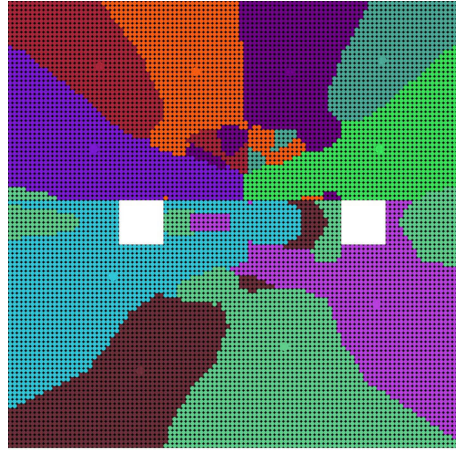


b) 1 POIs and 20 clusters

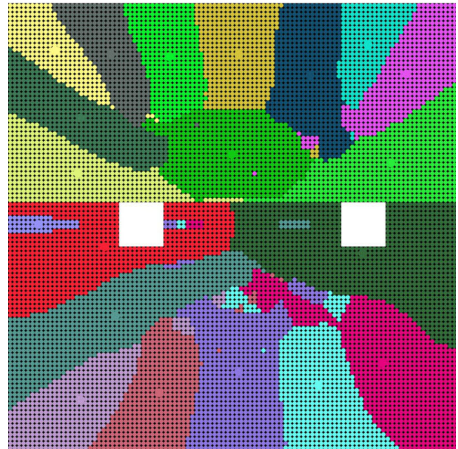


c) 1 POIs and 30 clusters

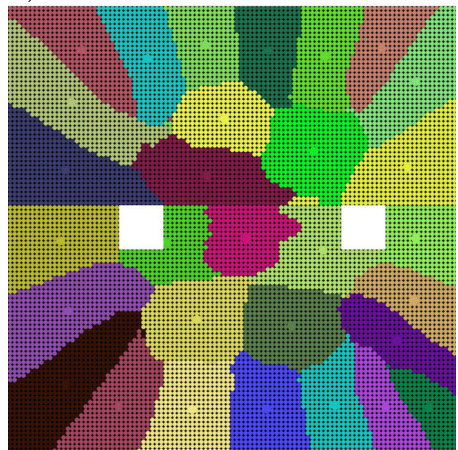
Fig. 11 Visual representation of clusters in 2 POIs for 10, 20 and 30 clusters



a) 2 POIs and 10 clusters

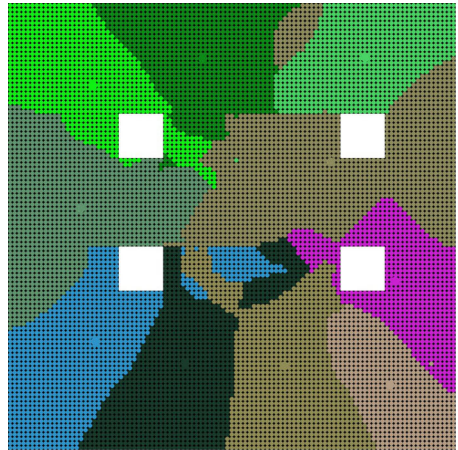


b) 2 POIs and 20 clusters

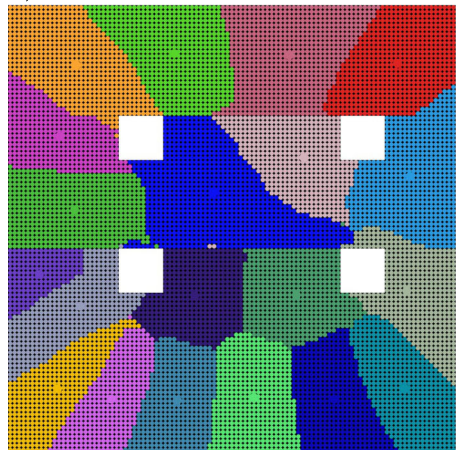


c) 2 POIs and 30 clusters

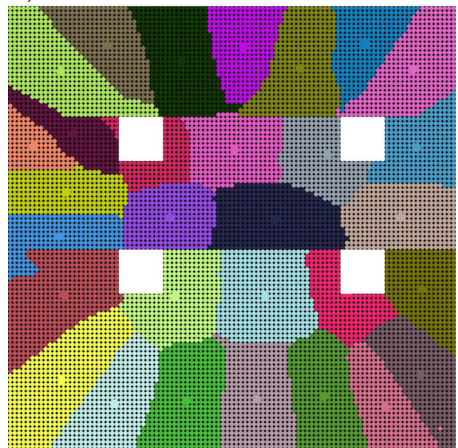
Fig. 12 Visual representation of clusters in 4 POIs for 10, 20 and 30 clusters



a) 4 POIs and 10 clusters



b) 4 POIs and 20 clusters



c) 4 POIs and 30 clusters

Total Clustering Errors vs. Clustering Time



Fig. 13 Additional information describing a point of interest is augmented on the image captured by the camera which is shown in a callout

As discussed in [31], AR applications are prone to the *occlusion* problem. That is, some contents may visually cover the AR content that is being displayed. This may result in losing some valuable information. In the camera view, the titles of POIs are augmented on the real data. It is possible that the titles of POIs cover each other when they are in the same direction. To avoid this, similar to what has been done in CorfuAR [32], we filter the number of POIs that can be shown on the camera view based on their distance from the current location. This a reasonable decision, as the intent of the camera view is not to show all POIs in the camera view. Generally, users are interested to see what is nearby, rather than

Total Clustering Errors vs. Clustering Time, #POI = 2



Fig. 14 A 3D arrow representing the direction to a point of interest (Museum in this case) is augmented on the image captured by the camera

see all POIs in one view. Although this method reduces the probability of occlusion, it is still possible that that POIs in the same direction are shown on each other. To address this problem, we first select the subset of POIs filtered based on their distance and sort them. After drawing the closest POIs, the area used to draw this POI is marked as reserved. Then, the location of the next item on the camera view is indicated. If there is an intersection between this area and the area already marked as reserved, up and left, up and right, down and left, or down and right shifts are randomly performed on the display location of POIs until there is no intersection between drawing areas.

In the case of online architecture, the information augmented on the real data is fetched from the server once the tracking sensor decides what must be shown. Although this can add flexibility to keep the content updated and also help in managing the application, there are a few major issues that hinder the application of this approach. First, not all locations in the cities are covered by Internet networks. Moreover, data transfer can be considerable especially for large multimedia content. Offline applications are a viable solution to address these problems. The LAR and MAR prototypes proposed in this paper are offline applications in which multimedia contents augmented on real data are already available in the application on the mobile device.

6 User evaluation

The goal of user evaluation is to measure quantitative and qualitative data related to using different AR tracking in tourism applications. According to [11], AR applications must be evaluated based on *performance* (examining the user task performance within AR applications), *perception* (understanding how human perception and cognition operate in AR contexts) and *usability* (identifying issues related to system usability). In this paper, we measure *time to task completion* and *error counts* as quantitative measures to evaluate task performance. In terms of perception, we measure expected quality of user experience of AR based on the technique proposed in [38]. To evaluate usability, we measure acceptance and use of technology in a consumer context using UTAUT2 proposed in [53]. Through this evaluation, we aim to find how the tracking technique affects mobile AR-navigation.

6.1 Study design

A user evaluation in a real tourist navigation setting was designed to study our prototypes. Since the main goal of the prototypes is helping users to navigate and find POIs, the task of finding three given POIs and physically visiting them was devised to be performed by the participants. Independent variables in this study are the type of tracking system (LAR: Location-based Augmented Reality, and MAR: Marker-based Augmented Reality) and the number of existing POIs in the site. The goal of considering the type of system as an independent variable was to determine if the tracking type has an effect on performance, perception and usability measures. The number of POIs corresponds to the complexity of the task as finding a given POI will be more difficult as the number of POIs in the site increases. We varied the number of existing POIs, choosing levels of 10, 25 and 50. Note that the number of POIs to be found by the participants remained constant (i.e., 3 POIs).

To address the potential learning effect, we used a between-subjects design, where each participant was assigned to only one of the groups. This way, we removed the order in which users work with different systems as a potential confound. More specifically, according to our grouping policy, 6 groups (based on the type of system \times the complexity of task), were

formed with 15 participants in each group. Considering a 5-minute introductory session to explain the experiment and trying the prototypes, the average time of experiment for each user was 25 minutes. Participants were told to finish their given tasks in 15-20 minutes, and 91% of visitors finished their visits within this constraint.

TIAU campus was selected as a case study to evaluate the effectiveness of the proposed prototypes and to compare the effect of different tracking techniques on the quality of user experience. This campus is a historical site with an area of 36,000 square meters. The buildings in this campus are part of an old leather factory which was renovated as a university campus. Many landmarks and POIs are located on the campus, attracting many tourists to visit.

To eliminate the type of device on which the application is used as an independent variable, we asked participants to use our device. We used Huawei Honor 3, with body size of $(5.24 \times 2.65 \times 0.39)$ inches, screen size of 4.7 inches, and resolution of 720×1280 pixels. The study started after obtaining informed consent from participants. A pre-study questionnaire was filled by each participant to measure educational level of participants and prior experience with AR. To make sure all participants were familiar with the concept of AR, a brief overview of AR and an instruction to use MAR and LAR were provided. Because of the between-subjects design of the research, each participant used only one of the prototypes (LAR or MAR). Consequently, each participant was introduced to only one of the prototypes at the beginning of the study.

The details of the task were explained to the participants and we made sure they knew what they were required to do. Users were asked to start the task from the same entrance of the campus. Only questions about the instructions to use the app were answered during each task. After task completion, participants completed a short questionnaire regarding their confidence in the results they have obtained as well as the ease of completing the task.

6.2 Participants

A total of 90 individuals of different ages and diverse origins visiting the TIAU campus were recruited to participate in the study. Standing in the main entrance of TIAU campus, we asked visitors if they want to participate in the study. They were told the study conditions and the duration of the experiment. The only preconditions to participating in the study were having experience using smartphones and being a first time visitor to the TIAU campus. The pre-study questionnaire verified a similar level of prior experience with AR. Among participants, 13% were familiar with the concept of AR, while just two participants had already experience of using AR applications. Sixty-five participants were male and 25 were female. Ages varied between 15 and 54 years, however 75% of participants were young adults aged between 20 and 34. Only 45% of the participants reported using mobile phones for navigation. Although *GoogleMaps* is active in Iran, it is not as complete as the maps in some other developed countries. We argue this issue is one of the reasons for the low rate of using phones for mobile navigation in Iran. Based on the technology-related questions, most participants can be considered highly technologically oriented. Participants were recruited among public visitors visiting the historical campus of TIAU and those who were interested in participating in the study. They were compensated with a visual book about Tabriz.

6.2.1 Performance

To measure the performance of prototypes, dependent variables were *time to task completion*, automatically measured by the application by tapping on the start button in the

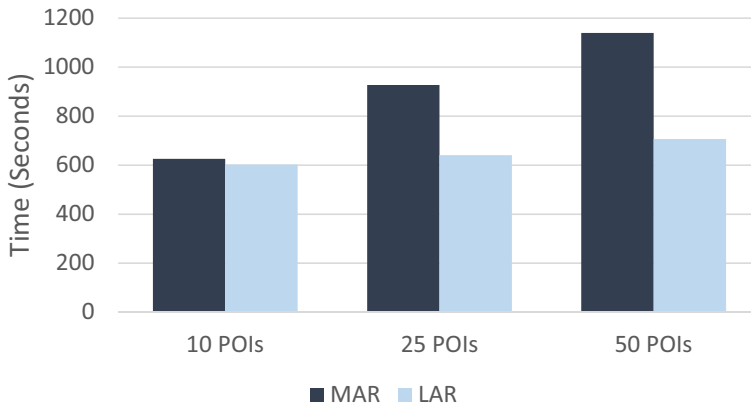


Fig. 15 The average time to task completion for finding three given POIs among different number of POIs using MAR and LAR

beginning and the finish button at the end of the experiment, and *error counts*, that is finding a wrong POI or not finding a POI. As a 2×3 between-subjects design (type of system \times the complexity of task), each participant performed only one of the scenarios. The primary analysis was to determine which tracking technique leads to better performance in terms of time to task completion and error rates. We also aimed to verify whether the number of existing POIs has an impact on the performance of participants.

Time to task completion The average time to task completion for each scenario is illustrated in Fig. 15. As shown in this figure, time to task completion for MAR is more than LAR in all scenarios. In particular, on average the time to task completion for MAR was 4% more for 10 POIs, 45% more for 25 POIs and 61% more for 50 POIs. Although the average time to task completion for finding given POIs among different number of POIs using MAR and LAR are shown in Figure 15, pair-wise comparison of time to task completion for the devised task using ANOVA representing statistical confidence is also shown in Table 1. The results show statistically significant difference between 25 POIs pair and 50 POIs pair, but not for the 10 POI pair. As a result, we conclude that H1 (Participants will take less time to find and visit POIs using LAR than using MAR) is supported for a large number of POIs. That is, participants were able to perform the task faster when they used LAR than MAR to find and visit given POIs among more than 25 POIs (Table 2).

In terms of time to task completion, we also measured the scalability of the prototypes. In the case of LAR, using the time to completion with 10 POIs as a baseline, on average there was 6% increase in the time to task completion using the 25 POIs, and 17% increase using 50 POIs. A pair-wise analysis of this data using ANOVA (see Table 3) found no statistical

Table 2 Pair-wise comparison of time to task completion for the devised Task using ANOVA

Comparison of prototypes	ANOVA Results
10 POIs - LAR(M = 602, SD = 94) vs. MAR(M = 625, SD = 104)	$F = 0.4144, p = 0.524$
25 POIs - LAR(M = 640, SD = 93) vs. MAR(M = 926, SD = 91)	$F = 71.8152, p < 0.05$
50 POIs - LAR(M = 706, SD = 123) vs. MAR(M = 1139, SD = 177)	$F = 60.3828, p < 0.05$

Table 3 Pair-wise comparison of different POIs set in terms of time to task completion for LAR

Comparison of different POIs set	ANOVA Results
10 POIs (M = 602, SD = 94) vs. 25 POIs (M = 640, SD = 93)	$F = 1.2665, p = 0.2699$
25 POIs (M = 640, SD = 93) vs. 50 POIs (M = 706, SD = 122)	$F = 2.7371, p = 0.1092$
10 POIs (M = 602, SD = 94) vs. 50 POIs (M = 706, SD = 122)	$F = 0.3328, p = 0.5685$

significance in the differences. As a result, we conclude that there are no differences in time to task completion regardless of the number of POIs for LAR within the range we considered. That is, even though the number of POIs to examine increased by five times, the time to find POIs did not increase. In the case of MAR, a pair-wise analysis of this data using ANOVA (see Table 4) found statistically significance differences between all pairs. As a result, we conclude that MAR is not scalable in terms of the time to task completion with increasing the number of existing POIs in the site.

According to the background of participants, only 45% of the participants reported using mobile phones for navigation. We performed an analysis to find if the performance varies based on this factor in either LAR or MAR. According to the box plots for LAR shown in Fig. 16, boxes overlap with both medians (LAR-10-N with LAR-10-Y, LAR-25-N with LAR-25-Y, LAR-50-N with LAR-50-Y), representing there is no statistically significant difference between the performance of those who had used (Y) and those who had not used (N) mobile phones for navigation in the case of 10, 25 and 50 POIs for LAR. Accordingly, there is no statistically significant difference between the performance of the participants in the case of MAR in terms of using/not using mobile phones for navigation (See Fig. 17).

Error counts Two types of errors were identified based on the participants' performance: (1) participants identifying and visiting a wrong POI, and (2) participants not being able to identify a given POI. The average number of each of these errors is illustrated in Fig. 18. Because of the extremely low error rate, for the purposes of statistical verification, the errors were grouped together and a single ANOVA analysis was performed. According to the results of ANOVA ($F = 1.8064, p = 0.1897$), there is no statistical significance in the error rates between MAR and LAR. As a result, we conclude that H2 (Participants will make fewer errors in finding and visiting POIs using LAR than using MAR) is not supported. In addition, there was no statistical significance in the error rates within LAR and MAR as the number of POIs increases. This leads to the conclusion that participants will make a similar number of errors in finding and visiting POIs, regardless of the number of existing POIs in the site using either LAR or MAR.

Perception In [24], user experience (UX) is defined as “a person’s perceptions and responses that result from the use or anticipated use of a product, system or service”. This is a subjective and holistic concept describing the experience resulting from the interaction

Table 4 Pair-wise comparison of different POIs set in terms of time to task completion for MAR

Comparison of different POIs set	ANOVA Results
10 POIs (M = 625, SD = 104) vs. 25 POIs (M = 926, SD = 91)	$F = 70.4050, p < 0.05$
25 POIs (M = 926, SD = 91) vs. 50 POIs (M = 1139, SD = 177)	$F = 17.0460, p < 0.05$
10 POIs (M = 625, SD = 104) vs. 50 POIs (M = 1139, SD = 177)	$F = 93.6063, p < 0.05$

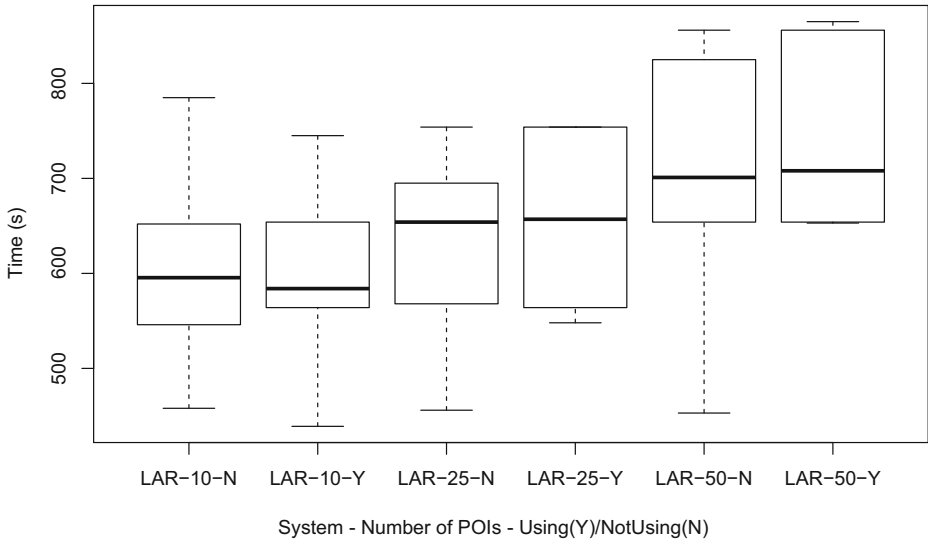


Fig. 16 The comparison between the performance of the users who use (Y)/do not use (N) mobile phones for navigation. The overlap between (Y) and (N) in each group (10, 25 and 50 POIs)

with a technological product or service. In [17], the basic parameters of the quality of user experience are identified as utility, joy, appeal, and aesthetics. Although general level framework of user experience is considered widely in the literature, there is a lack of user experience research in specific fields of technology like mobile AR.

In this paper, we use the technique and the questionnaires proposed in [38] to highlight the relevance of expected quality of user experience regarding the use of AR applications.

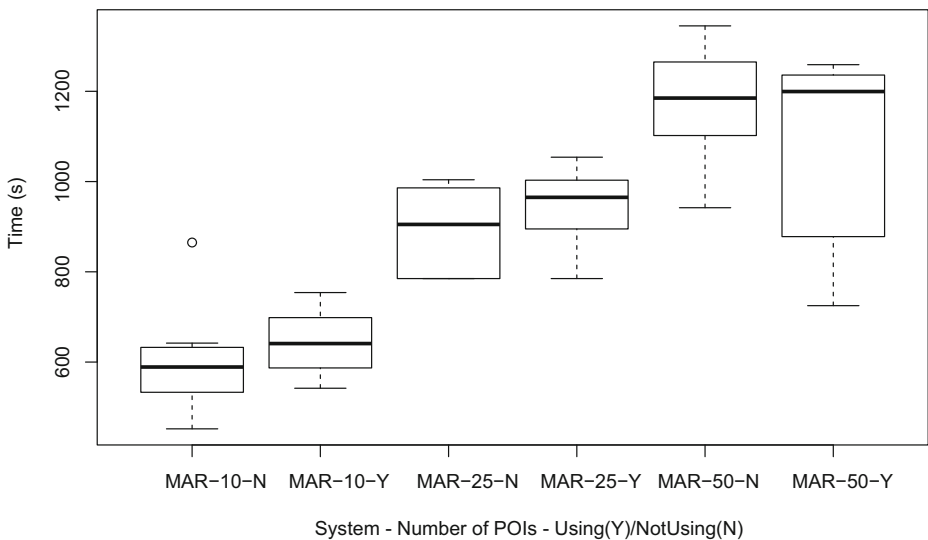


Fig. 17 The comparison between the performance of the users who use (Y)/do not use (N) mobile phones for navigation. The overlap between (Y) and (N) in each group (10, 25 and 50 POIs)

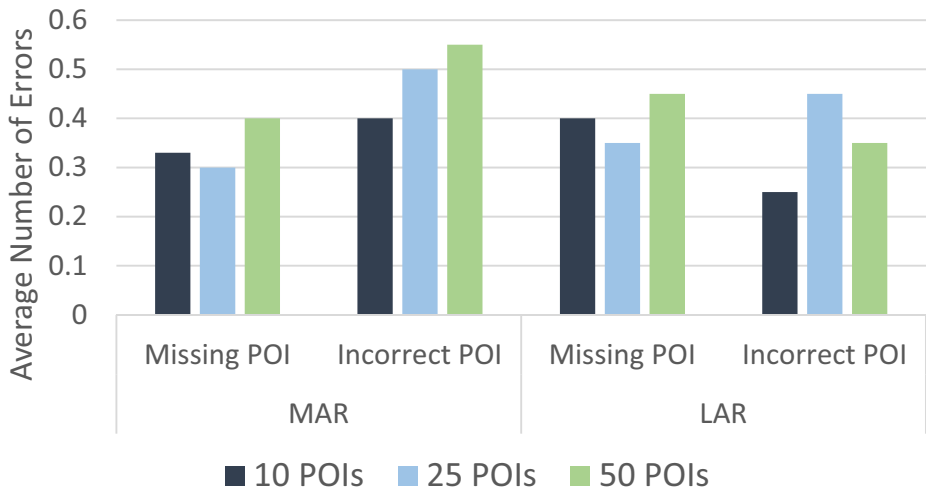


Fig. 18 The average number of each of the different types of errors that were made during the tasks

This questionnaire originally includes 18 questions regarding the expected user experience of AR applications. We modified the questions as statements to express to what extent these expectations are fulfilled using our prototypes. We also categorized the questions to context awareness (CA1, CA2), quality of experience (QE1, QE2, QE3, QE4, QE5), self-expressiveness (SE1, SE2, SE3) and cognitive effort (CE1, CE2, CE3). The questions were also translated to Persian. Four questions regarding the quality of communication with others using AR were removed as they were not applicable in our case. An English version of the questionnaire used in the study is shown in Table 5.

A pair-wise Wilcoxon-Mann-Whitney test was conducted on the subjective data from this questionnaire to identify if there were any statistically significant difference as a result

Table 5 The questionnaire used to evaluate the perception in terms of user experience

Questions	
Context Awareness	CA1: Attaches new meanings to places, services, products etc CA2: Increased my understandability of my environment and its objects
Quality of Experience	QE1: Offered me new experiences QE2: Brought significant benefits to my everyday life QE3: The system was captivating QE4: I got a lot from augmented reality services QE5: The system Delighted me
Self-Expressiveness	SE1: Helped to inspire my creativity SE2: The system helped me to develop myself SE3: The system offered me a suitable way to express myself
Cognitive Efforts	CE1: The system was easy to use CE2: Did not require too much of my attention CE3: Interfered my understanding of what is real and what is virtual

of completing the task with different prototypes. We combined the results of tasks with different number of POIs. Consequently, we tested the results of 45 participants for LAR and 45 participants for MAR. All items were measured using a seven-point Likert scale, with the anchors being “strongly disagree” and “strongly agree”. The test found a statistically significant difference between the quality of user experience for MAR and LAR ($LAR\ mean = 5.3666$, $MAR\ mean = 4.7111$, $p = 0.0027$). As a result, we conclude that H3 (Participants will report a higher level of quality of user experience using LAR than using MAR) is supported. In a fine-grained analysis, *context awareness* in terms of the effect of increasing the extent of available information from real life objects as well as increasing the users’ understanding of the nearby environment, was significantly higher in LAR than MAR ($LAR\ mean = 5.4333$, $MAR\ mean = 4.6222$, $p = 0.0003$). Although in terms of the *quality of experience*, participants were positive towards the presented scenarios using both LAR and MAR, participants reported higher quality of experience for LAR than MAR ($LAR\ mean = 5.8111$, $MAR\ mean = 5.1555$, $p = 0.004$). However, for self-expressiveness, ANOVA found no statistical significance in the differences between the self-expressiveness of MAR and LAR ($LAR\ mean = 5.4555$, $MAR\ mean = 5.2$, $p = 0.2588$). In terms of cognitive effort (which was reverse coded, so that higher numbers imply lower cognitive effort), there was a big difference between the mean cognitive effort of MAR and LAR ($LAR\ mean = 5.6777$, $MAR\ mean = 4.8666$, $p = 0.0007$). We attribute this to the convenience of using LAR, as POI tags are shown automatically on the screen just by holding the camera in front a POI. In the case of LAR, the switch between AR and the physical world can be conducted smoothly as this switch is performed automatically without requiring the user to do anything. In addition, misalignment with the AR marker is a problem of MAR applications that might negatively affect the quality of user experience. This can be frustrating if the information represented on the screen goes on and off.

6.2.2 Usability

We measured individual acceptance and use of technology in order to understand to what extent the AR technology can be employed in the field of tourism. In particular, we used UTAUT2 [53], which is a modified version of UTAUT (Unified Theory of Acceptance and Use of Technology). This model includes performance expectancy, effort expectancy, social influence, facilitating conditions, hedonic motivation, price value, habit, and behavioral intention, as factors affecting acceptance and use of technology. Performance expectancy is used to measure to what extent using a technology benefits users in performing certain tasks. Effort expectancy is defined as ease of using technology by users. Social influence is the degree to which users perceive that other people believe they should use a technology. Facilitating conditions cover user perceptions of the facilities available to perform a certain activity. Hedonic motivation is fun or pleasure a user gains from using a technology and has a crucial role in determining technology acceptance and use. Habit is is factor to determine technology acceptance and use, which is defined as the degree to which users tend to conduct activities automatically because of learning. The habit section was dropped as it is not applicable in our case study. The price value was not also considered as our prototypes are free to use. We also dropped the use section since the AR navigation is a pretty new type of application where participants were not aware of any other competing products such as wiktitude in order to compare our prototypes with those products. All items were measured using a seven-point Likert scale, with the anchors being “strongly disagree” and “strongly agree”.

According to the results of a pair-wise Wilcoxon-Mann-Whitney test conducted on the subjective data, there were statistically significant differences as a result of completing the task with MAR and LAR ($LAR\ mean = 5.7222$, $MAR\ mean = 4.9111$, $p = 0.0001$). This leads to the conclusion that H4 (Participants will report a higher level of acceptance and use of technology using LAR in comparison to MAR) is supported. Looking at the details of this study, performance expectancy ($LAR\ mean = 5.5777$, $MAR\ mean = 4.9111$, $p = 0.0037$), effort expectancy ($LAR\ mean = 5.5555$, $MAR\ mean = 4.9555$, $p = 0.0148$), facilitating conditions ($LAR\ mean = 5.4222$, $MAR\ mean = 4.8888$, $p = 0.0031$), and Hedonic Motivation ($LAR\ mean = 5.7777$, $MAR\ mean = 5.0666$, $p = 0.0065$) were significantly higher in LAR than MAR. In the case of behavioral intention, there were no statistical significance in the differences between MAR and LAR. In comparison to other factors, the mean values of behavioral intention for both MAR and LAR are low. According to the comments of participants, most of the participants were reluctant to continually use the application all the time during their visit. They were interested to physically explore the site and use the application just in the case they were interested to find details about a particular POI, or they were interested to find how they can go to a particular POI. The mean values of social behavior for both MAR and LAR are also low ($LAR\ mean = 4.7333$, $MAR\ mean = 3.9555$). We attribute this to the fact that the prototypes are new products that are not used frequently. Consequently, people who are important to the participants, people who influence the behavior of the participant, and people whose opinions that the participant value may not be aware of this application.

7 Conclusion and future work

Augmented reality is an increasingly important class of application that provides value to users in the form of additional information or immersion in a real life setting. In this paper, we propose and compare two augmented reality tracking techniques, location-based (LAR) and marker-based (MAR). We hypothesize that the characteristics of LAR provide advantages that: (1) require less time to locate POIs in a given area; (2) result in fewer errors in locating specific POIs; (3) induce higher perceptions of the quality of user experience; and (4) induce higher levels of acceptance and intentions to use AR technology.

We designed and developed two prototype applications to implement the LAR and MAR approaches. We compared the approaches in a field experiment to test the hypotheses above. Results from the empirical analysis showed that as hypothesized, compared to MAR, subjects using the LAR approach: (1) completed the task of locating three POIs in less time; (2) reported a higher quality user experience; and (3) reported a higher level of acceptance of LAR than MAR. However, contrary to expectations, subjects using LAR had no fewer errors than subjects using MAR (errors were low in both groups) and did not report higher intentions to use LAR than MAR. Overall, these findings suggest that location-based AR produce a range of favorable outcomes compared to MAR.

Our findings need to be interpreted cautiously in light of the limitations of the empirical study. First, our location based solution depends on GPS data, which is power consuming and may not be accurate enough for wayfinding. Second, our experiment dealt with only a single site in which the POIs were artifacts such as buildings; it is not possible to generalize these findings to other settings, such as those in which POIs may be natural phenomena. Third, the experimental task involved finding only three POIs. It is unclear if the restricted nature of the task contributed to the failure to find significant difference in task accuracy between the LAR and MAR experimental conditions. Fourth, our findings may be affected

by the requirement to use a device we provided to participants (rather than their own device). Finally, due to the cross-sectional nature of the study, we measure intentions to use the technology, rather than actual use.

This work both contributes to the design of AR applications and provides empirical evidence of the advantages of location-based tracking. Future research can extend the work reported here in several ways. First, we argue implementing computer vision techniques to detect scenes that are visible in the camera view can lead to a more accurate localization. Such computer vision techniques can be implemented using object detection tools such as those from OpenCV. Second, experiments can be conducted across a range of settings to examine the generalizability of our findings with respect to the benefits of LAR over MAR. Included in such work should be an expansion of the task set to test whether task accuracy depends on complexity as measured by the number of POIs to be located. Additionally, future research can complement our cross-sectional study, which included questions about usage intentions by adopting a longitudinal approach in which ongoing use of LAR-based vs. MAR-based applications is studied to determine whether or not the former leads to greater adoption behavior as familiarity with the application increases.

Acknowledgements This paper is financially supported by Tabriz Islamic Art University based on the research project 1405/1, 2106/02/04.

References

1. Azuma R, Baillot Y, Behringer R, Feiner S, Julier S, MacIntyre B (2001) Recent advances in augmented reality. *IEEE Comput Graph Appl* 21(6):34–47
2. Beeharee AK, Steed A (2006) A natural wayfinding exploiting photos in pedestrian navigation systems. In: Proceedings of the 8th conference on human-computer interaction with mobile devices and services. ACM, pp 81–88
3. Benyon D, Quigley A, O’Keefe B, Riva G (2014) Presence and digital tourism. *AI Soc* 29(4):521–529
4. Bruns E, Bimber O (2009) Adaptive training of video sets for image recognition on mobile phones. *Pers Ubiquit Comput* 13(2):165–178
5. Chang YL, Hou HT, Pan CY, Sung YT, Chang KE (2015) Apply an augmented reality in a mobile guidance to increase sense of place for heritage places. *Journal of Educational Technology and Society* 18(2):166–178
6. Chen W (2014) Historical Oslo on a handheld device—a mobile augmented reality application. *Procedia Comput Sci* 35:979–985
7. Cheng KH, Tsai CC (2013) Affordances of augmented reality in science learning: suggestions for future research. *J Sci Educ Technol* 22(4):449–462
8. Chumkamon S, Tuvaphanthaphiphat P, Keeratiwintakorn P (2008) A blind navigation system using RFID for indoor environments. In: 5th international conference on electrical engineering/electronics, computer, telecommunications and information technology, 2008. ECTI-CON 2008, vol 2. IEEE, pp 765–768
9. Chung N, Han H, Joun Y (2015) Tourists’ intention to visit destination: role of augmented reality applications for heritage site. *Comput Hum Behav*
10. Danado J, Dias E, Romão T, Correia N, Trabuco A, Santos C, Câmara A (2003) Mobile augmented reality for environmental management (MARE). In: Eurographics 2003 conference. Granada, Spain
11. Dünser A, Grasset R, Billinghurst M (2008) A survey of evaluation techniques used in augmented reality studies. *Human Interface Technology Laboratory, New Zealand*, pp 5–1
12. Emmanouilidis C, Koutsiamanis RA, Tasidou A (2013) Mobile guides: taxonomy of architectures, context awareness, technologies and applications. *J Netw Comput Appl* 36(1):103–125
13. Endsley MR (2000) Theoretical underpinnings of situation awareness: a critical review. *Situation awareness analysis and measurement*, 3–32
14. Gebril ZM, Tele IMA, Tahir MA, Parhizkar B, Ramachandran A, Lashkari AH (2012) Ubiquitous medical learning using augmented reality based on cognitive information theory. In: *Advances in computer science, engineering & applications*. Springer, Berlin, pp 305–312

15. Geiger P, Schickler M, Pryss R, Schobel J, Reichert M (2014) Location-based mobile augmented reality applications: challenges, examples, lessons learned
16. Greene K, Kahari M, Murphy D (2006) Hyperlinking reality via phones. *Technology Review*, November/December
17. Hassenzahl M, Tractinsky N (2006) User experience—a research agenda. *Behav Inform Technol* 25(2):91–97
18. Hering E (1964) Outlines of a theory of the light sense
19. Hervás R, Bravo J, Fontecha J (2014) An assistive navigation system based on augmented reality and context awareness for people with mild cognitive impairments. *IEEE J Biomed Health Inform* 18(1):368–374
20. Hervás R, Garcia-Lillo A, Bravo J (2011) Mobile augmented reality based on the semantic web applied to ambient assisted living. In: *International workshop on ambient assisted living*. Springer, Berlin, pp 17–24
21. Hile H, Borriello G (2008) Positioning and orientation in indoor environments using camera phones. *IEEE Comput Graph Appl* 28(4)
22. Hsiao KF, Rashvand HF (2015) Data modeling mobile augmented reality: integrated mind and body rehabilitation. *Multimedia Tools Appl* 74(10):3543–3560
23. Ishikawa T, Yamazaki T (2009) Showing where to go by maps or pictures: an empirical case study at subway exits. In: *International conference on spatial information theory*. Springer, Berlin, pp 330–341
24. ISO FDIS 9241-210:2009, Ergonomics of human system interaction – part 210: human-centred design for interactive systems (formerly known as 13407)
25. Jacob R, Mooney P, Winstanley AC (2011) Guided by touch: tactile pedestrian navigation. In: *Proceedings of the 1st international workshop on mobile location-based service*. ACM, pp 11–20
26. Jung S, Kim S, Kim S (2013) Augmented reality-based exhibit information personalized service architecture through spectator’s context analysis. *Inter J Multimedia Ubiquitous Eng* 8(4):313–320
27. Jung T, Chung N, Leue MC (2015) The determinants of recommendations to use augmented reality technologies: the case of a Korean theme park. *Tour Manag* 49:75–86
28. Kapoor P, Ghufuran U, Gupta M, Agarrwal A (2013) Marker-less detection of virtual objects using augmented reality. In: *Proceedings of the conference on advances in communication and control systems-2013*. Atlantis Press
29. Karimi HA (ed.) (2004) *Telegeoinformatics: Location-based computing and services*. CRC Press
30. Katz BF, Kammoun S, Parsehian G, Gutierrez O, Brilhault A, Auvray M, ..., Jouffrais C (2012) NAVIG: augmented reality guidance system for the visually impaired. *Virtual Reality* 16(4):253–269
31. Kourouthanassis P, Boletsis C, Bardaki C, Chasanidou D (2015) Tourists responses to mobile augmented reality travel guides: the role of emotions on adoption behavior. *Pervasive Mob Comput* 18:71–87
32. Kourouthanassis PE, Boletsis C, Lekakos G (2013) Demystifying the design of mobile augmented reality applications. *Multimedia Tools and Applications* 74(3):1045–1066
33. LaMarca A, Chawathe Y, Consolvo S, Hightower J, Smith I, Scott J, ..., Tabert J (2005) Place lab: device positioning using radio beacons in the wild. In: *International conference on pervasive computing*. Springer, Berlin, pp 116–133
34. Lee S, Lim Y, Chun J (2013) 3D interaction in augmented reality with stereo-vision technique. In: *15th international conference on advanced communication technology (ICACT)*, 2013. IEEE, pp 401–405
35. Marimon D, Sarasua C, Carrasco P, Álvarez R, Montesa J, Adamek T, Gascó P (2010) MobiAR: tourist experiences through mobile augmented reality. *Telefonica Research and Development*, Barcelona, Spain
36. Mulloni A, Seichter H, Schmalstieg D (2011) User experiences with augmented reality aided navigation on phones. In: *10th IEEE international symposium on mixed and augmented reality (ISMAR)*, 2011. IEEE, pp 229–230
37. Nakayama Y, Saito H, Shimizu M, Yamaguchi N (2015) Marker-less AR system based on line segment feature. In: *IST/SPIE Electronic imaging*. International Society for Optics and Photonics, pp 93920i–93920i
38. Olsson T, Kärkkäinen T, Lagerstam E, Ventä-Oikkonen L (2012) User evaluation of mobile augmented reality scenarios. *J Ambient Intell Smart Environ* 4(1):29–47
39. Olsson T, Lagerstam E, Kärkkäinen T, Väänänen-Vainio-Mattila K (2013) Expected user experience of mobile augmented reality services: a user study in the context of shopping centres. *Pers Ubiquit Comput* 17(2):287–304
40. Olsson T, Salo M (2011) Online user survey on current mobile augmented reality applications. In: *10th IEEE international symposium on mixed and augmented reality (ISMAR)*, 2011. IEEE, pp 75–84
41. Patkar RS, Singh SP, Birje SV (2013) Marker based augmented reality using Android OS
42. Rehrl K, Häusler E, Leitinger S, Bell D (2014) Pedestrian navigation with augmented reality, voice and digital map: final results from an in situ field study assessing performance and user experience. *J Location Based Serv* 8(2):75–96

43. Reitmayr G, Drummond TW (2006) Going out: robust model-based tracking for outdoor augmented reality. In: IEEE/ACM international symposium on mixed and augmented reality, 2006. ISMAR 2006. IEEE, pp 109–118
44. Sekhavat YA (2016) KioskAR: an augmented reality game as a new business model to present artworks. *International Journal of Computer Games Technology*
45. Sekhavat YA (2016) Nowcasting mobile games ranking using web search query data. *International Journal of Computer Games Technology*
46. Sekhavat YA (2016) Privacy preserving cloth try-on using mobile augmented reality. *IEEE Transactions on Multimedia* 19(5):1041–1049. doi:[10.1109/TMM.2016.2639380](https://doi.org/10.1109/TMM.2016.2639380)
47. Sekhavat YA (2017) Behavior trees for computer games. *Int J Artif Intell Tools* 26(2):1730001–1730028
48. Seo BK, Kim K, Park JI (2011) Augmented reality-based on-site tour guide: a study in Gyeongbokgung. In: *Computer vision–ACCV 2010 workshops*. Springer, Berlin, pp 276–285
49. Shneiderman B (1996) The eyes have it: a task by data type taxonomy for information visualizations. In: *IEEE symposium on visual languages, 1996. Proceedings. IEEE*, pp 336–343
50. Skrypnik I, Lowe DG (2004) Scene modelling, recognition and tracking with invariant image features. In: *Third IEEE and ACM international symposium on mixed and augmented reality, 2004. ISMAR 2004. IEEE*, pp 110–119
51. Takacs G, Chandrasekhar V, Gelfand N, Xiong Y, Chen WC, Bispigiannis T, ..., Girod B (2008) Outdoors augmented reality on mobile phone using loxel-based visual feature organization. In: *Proceedings of the 1st ACM international conference on multimedia information retrieval*. ACM, pp 427–434
52. Tutunea MFS (2013) Augmented reality-state of knowledge, use and experimentation. *The USV Annals of Economics and Public Administration* 13(2(18)):215–227
53. Venkatesh V, Thong JY, Xu X (2012) Consumer acceptance and use of information technology: extending the unified theory of acceptance and use of technology. *MIS Q* 36(1):157–178
54. Willis S, Helal S (2005) RFID Information grid for blind navigation and wayfinding. In: *Ninth IEEE international symposium on wearable computers, 2005. Proceedings. IEEE*, pp 34–37
55. Yamabe T, Nakajima T (2013) Playful training with augmented reality games: case studies towards reality-oriented system design. *Multimedia Tools Appl* 62(1):259–286
56. Yeh T, Tollmar K, Darrell T (2004) Searching the web with mobile images for location recognition. In: *Proceedings of the 2004 IEEE computer society conference on computer vision and pattern recognition, 2004. CVPR 2004, vol 2. IEEE*, pp II–II
57. Yovcheva Z, Buhalis D, Gatzidis C (2013) *Engineering augmented tourism experiences*. Springer, Berlin, pp 24–35



Yoones A. Sekhavat is an Assistant Professor in Faculty of Multimedia in Tabriz Islamic Art University. He holds a PhD in Computer Science from Memorial University of Newfoundland, Canada. He also holds a post-doc from University of Alberta in Canada. His research has been published in *IEEE Transactions on Multimedia*, *IEEE Transactions on Knowledge and Data Engineering*, *International Journal of Artificial Intelligence Tools*, and *International Journal of Computer Games Technology* as well as VLDB, ER, SEGAAH and ICSC Conferences. His research interests include serious games, augmented reality, artificial intelligence for computer games, and recommender systems. He is the head of the cognitive augmented reality lab (www.carlab.ir).



Jeffrey Parsons is University Research Professor and Professor of Information Systems in the Faculty of Business Administration at Memorial University of Newfoundland. His research interests include information integration, crowdsourcing, information quality, conceptual modeling, and recommender systems. His research has been published broadly, including journals such as *Nature*, *Communications of the ACM*, *ACM Transactions on Database Systems*, *IEEE Transactions on Software Engineering*, *Management Science* and *MIS Quarterly*. He is a Senior Editor for *MIS Quarterly*. He has also served as Program Co-chair for a number of conferences, including AMCIS, WITS, and the ER conference. He holds a PhD in Information Systems from the University of British Columbia.