

Article

Usability Evaluation with Eye Tracking: The Case of a Mobile Augmented Reality Application with Historical Images for Urban Cultural Heritage

Diana Szekely ^{1,*}, Silviu Vert ^{1,*} , Oana Rotaru ¹ and Diana Andone ² 

¹ Communications Department, Politehnica University of Timisoara, 300223 Timisoara, Romania; diana.szekely@student.upt.ro (D.S.); oana.rotaru@student.upt.ro (O.R.)

² eLearning Center, Politehnica University of Timisoara, 300223 Timisoara, Romania; diana.andone@upt.ro

* Correspondence: silviu.vert@upt.ro

Abstract: Eye-tracking technologies have matured significantly in recent years and have become more affordable and easier to use. We investigated how eye-tracking technology can be applied to evaluate the usability of mobile augmented reality applications with historical images for urban cultural heritage. The experiment involved a series of complex user evaluation sessions, combining semi-structured interviews, observations, think-aloud protocol, SUS questionnaire, and product reaction cards, complemented by eye tracking, to gather insights on the Spotlight Timisoara AR mobile application, part of a digital storytelling multiplatform for the city of Timisoara (Romania), soon to be European Capital of Culture in 2023. The results indicate strong and weak aspects of the application, both as expressed by the participants and as derived from analyzing the eye-tracking data. The paper also lists the main challenges we identified in using eye-tracking equipment to evaluate the usability of such mobile augmented reality applications for urban outdoor heritage.

Keywords: augmented reality; cultural heritage; eye tracking; usability testing; user experience



Citation: Szekely, D.; Vert, S.; Rotaru, O.; Andone, D. Usability Evaluation with Eye Tracking: The Case of a Mobile Augmented Reality Application with Historical Images for Urban Cultural Heritage. *Heritage* **2023**, *6*, 3256–3270. <https://doi.org/10.3390/heritage6030172>

Academic Editors: Stella Sylaiou, George Pavlidis and Christos Fidas

Received: 6 March 2023

Accepted: 16 March 2023

Published: 21 March 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The digitalization of cultural heritage has piqued the interest of art enthusiasts as a viable method to preserve, restore, and promote cultural treasures. Among the current technologies, augmented reality (AR) has gained popularity as a concept for integrating and mixing digital aspects into the physical world of consumers [1].

In an earlier systematic review of the literature [2], the authors demonstrated that usability testing of mobile augmented reality applications for cultural heritage has been conducted primarily through interviews, sometimes combined with other well-known methods such as focus groups and observations with the think-aloud protocol. Almost all the studies evaluated outdoor mobile applications with location-based augmented reality.

As a result of this identified stringent need for more comprehensive user testing, the authors have performed a multiplatform usability evaluation [3] of the Spotlight Heritage Timisoara project, a digital storytelling platform for the city of Timisoara (Romania), European Capital of Culture in 2023. This study employed semi-structured interviews, observations, think-aloud protocol, SUS questionnaire, Net Promoter Score, and Product Reaction Cards to gather insights from 105 participants and reveal usability problems in the Spotlight Heritage context. This research revealed, amongst other results, the need to be able to determine some hidden aspects of the behavior of the users that the authors were not able to catch through other methods.

In this study, we go into a deeper usability evaluation, focused on the Spotlight Timisoara AR add-on application, which allows users to scan landmarks from the cultural heritage of Timisoara and display on top how their facades looked in the past. To gain

deeper insight into the user experience of the application, we combine the user testing methods mentioned above with eye tracking.

Eye-tracking evaluation is an established method to determine the user experience of digital products in conjunction with other usability evaluation methods [4]. Eye tracking has previously been used to aid in the evaluations of user experience of augmented reality applications, such as in analyzing the visual behavior of subjects looking at paintings inside museums [5] or in comparing the usability of AR applications to map-based versions to get directions in an outdoor mall [6]. However, to our knowledge, there are no studies describing the integration of eye tracking in user evaluation of mobile augmented reality mobile applications with historical images for urban outdoor cultural heritage.

The purpose of this paper is to answer the following research questions:

Q1: What are the insights that usability testing with eye tracking brought to the user experience of the Spotlight Timisoara AR application?

Q2: What are the challenges of integrating eye tracking into the usability evaluation of augmented reality applications for urban cultural heritage?

In order to answer these research questions, we performed a usability evaluation with semi-structured interviews, observations, think-aloud protocol, SUS questionnaire, and Product Reaction Cards, together with eye tracking. The study took place in June 2022, outside, in one of the historical neighborhoods of Timisoara, with six participants. The usability testing sessions were run by a moderator and an observer, and the eye tracking was performed with a Pupil Core pair of glasses.

In Section 2, we present the theoretical background for our research. In Sections 3 and 4, we describe the materials and methodology for running the usability testing with eye tracking. Sections 5 and 6 present the results of the experiment and a discussion of the results, respectively. Section 7 concludes the paper.

2. Theoretical Background

The current section briefly describes the main components of the problem space, both individually and in combination.

2.1. Usability Testing

One of the significant aspects of the quality of mobile applications is usability, which evaluates how simple an interface is to use and how satisfied users are with that use. The application must satisfy the human needs of users in order to offer a pleasant experience [7].

User evaluation can be performed through interviews, observations, the think-aloud protocol, the SUS questionnaire, or Product Reaction Cards.

The think-aloud protocol is a usability evaluation method that engages participants in speaking out loud their immediate reactions to using the tested application. Researchers record the comments and analyze them after the tests [8].

The System Usability Scale (SUS) questionnaire [9] is an instrument used to test the usability of commercial products. The SUS questionnaire is a 10-statement survey and records the user's agreement or disagreement with 10 aspects of the product that is evaluated. The final score can be explained through an adjective rating scale ("Worst Imaginable", "Awful", "Poor", "OK", "Good", "Excellent", "Best Imaginable").

Product Reaction Cards are another method used for the evaluation of software products. The method can be applied in a physical way, using physical cards to be picked by the participants, or in a virtual way, with the help of a Word/Excel file. This method was initially created by Microsoft, and its aim was to determine how desirable a product is [10].

Usability testing can be enhanced with eye tracking, where the main focus is on monitoring participants' eyes while they execute particular activities.

2.2. Eye Tracking

When we talk about how the eyes move when we look at particular objects, we can say that a series of movements is divided into several categories, the most important being the saccades, fixations, and smooth pursuit movements. The very fast movements that the eyes make when we read or look at particularly more complex objects make the foveal region reorient to a new location, this way producing jerky movements called saccades [11]. Another curious feature identified by Burr et al. is that humans are meant to see nothing during saccades, saying that “we are entirely blind” [12] at that time. Because the pace of a saccade is so fast, vision is suppressed, making saccades less significant in eye-tracking research than fixations. The term fixation was designed to describe the period between two saccades when visual information processing is found to be stable. When our eyes take a break to scan particular objects, the foveal vision is held centrally in one place, and the visual system takes over detailed information on what was viewed [13].

Eye tracking has become a widespread technology for understanding human behavior, the method being used both for research and for commercial purposes. This method allows us to measure and discover different durations and flashing intervals of the gaze, to find out what element was observed, or not, by the tested subjects, and to evaluate the pupil’s reaction when presented with different stimuli.

As previously discussed, most researchers believe that traditional usability testing techniques should always accompany eye tracking; therefore, J. Wang et al. [14] investigated the relationship between data received from eye-tracking data and traditional test data. Participant eye movements, such as saccades and fixation, were recorded simultaneously using specific software, and with the help of metrics collected after the tests, a close connection was discovered between the values of the eye-tracking data and the data collected from the traditional test methods.

Another study, by Hong-Fa Ho, explored how eye-tracking technology can be applied to product design and tried to explain the importance of eye movement throughout this process [15]. The findings mentioned come from the visual tracking of users’ behavior, but also from the data collected by the eye-tracking devices, the analysis leading to the coding of seven types of regions of interest based on attention. Ho and Lu [16] also noted that the size of the pupil could be evaluated to measure the level of emotional interest when a participant interacts with a product.

2.3. Augmented Reality

Although the concept of augmented reality appeared around the 1960s, it has only recently started to be known by a broader range of people and to be used to improve communication between real and virtual contexts. In the context of our research, D. Han et al. [17] proposed investigating and testing the requirements of tourists when it comes to developing augmented reality tourism applications in urban culture. Similarly, R. Safitri et al. [18] have built a mobile application with AR and researched how augmented reality can help provide information about tourism to people living in or visiting Indonesia. At the end of the research, they recognized the added value of augmented reality in the mobile application when the users who tested the application mentioned that it is an excellent way to promote tourist destinations.

Augmented reality technology proves to be helpful in various fields of study and offers a new approach to collecting and viewing complex data in the virtual environment.

2.4. Usability Testing with Eye Tracking for Augmented Reality

Usability tests are becoming necessary and even mandatory when the application’s success and proper functioning are desired. From traditional methods of testing usability for discovering the functional and non-functional components of the applications, technology has advanced to the point where, in recent years, testing usability through eye-tracking has been developed and applied supplementary more and more to the current studies. Researchers can use special eye-tracking equipment to watch participants’ eye movements

throughout various tasks while studying aspects such as comprehending patterns, social interaction techniques, and the cognitive processes that drive certain people's behavior [19]. This process helps us discover elements that attract the participants' attention, the points with the biggest areas of interest, and also the points that go unnoticed. Moreover, the technology has great potential to support usability testing and achieve effective results in improving design and user experience.

To our knowledge, there are no studies on the challenges of employing eye-tracking technology to test the usability of mobile augmented reality mobile applications with historical images for urban cultural heritage. Some studies cover the problem only partially. For example, researchers in [5] have used eye tracking to determine how museum visitors look at paintings, in order to improve how augmented reality applications are designed for these specific cases. In [6], the authors conducted an eye-tracking study to compare the usability of the Yelp app in its two forms: augmented reality and map-based.

3. Materials

3.1. The Spotlight Timisoara AR App

As mentioned in the introductory section, the multiplatform Spotlight Heritage Timisoara is a digital cultural initiative of the Politehnica University of Timisoara, built by the eLearning Center and the Multimedia Center, in partnership with the Banat National Museum, part of the Timisoara 2023 European Capital of Culture program [2]. The project allows users to discover a variety of landmarks from the cultural heritage of Timisoara and to read the personal stories of the residents about the communities and neighborhoods from the old days. The architecture behind the multiplatform Spotlight Heritage Timisoara has been designed in such a way that it allows multiple usage scenarios: web, mobile, touchscreen, AR, and VR.

As part of the mobile version, two mobile applications have been created: the main Spotlight Timisoara application and the augmented reality add-on called Spotlight Timisoara AR. Both are available for Android and for Apple. In terms of augmented reality functions, the main application offers an AR view of the surroundings, as an alternative to the map-based view, while the augmented reality add-on allows the recognition of landmarks [20] and shows their overlap in real time with images and information on how the landmark looked in the past [21]. The AR add-on is started directly from the main application, but it exists in the app stores as a separate application because of technical reasons. To use the AR add-on, users must be positioned in front of the tourist attraction about which they want to find information and scan it through the mobile device camera. As a result, the pictures and details of the scanned building are displayed in the application interface (Figure 1).

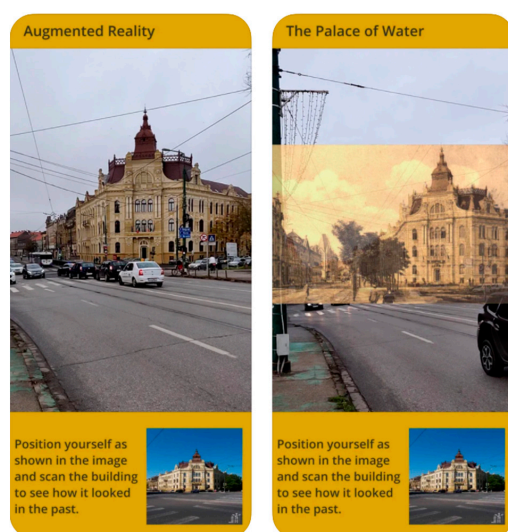


Figure 1. Screenshots from the Spotlight Timisoara AR app.

3.2. Pupil Core Eye Tracker

Pupil Core is an eye-tracking device built in the form of glasses that contains two cameras with an infrared spectrum positioned inwards, towards the eyes, and a universal scene camera orientated in front, applied on the frame. Glasses can record and provide accurate data on eye movements and eye gaze, regardless of head position, when connected to a laptop/PC via a high-speed USB 2.0 [22].

The glasses do not have a complicated structure on the hardware side, and the design minimizes their weight and durability (Figure 2). The eye camera can operate with a sampling frequency of up to 200 Hz and a resolution of 192×192 px, while the world camera has several frequencies, its maximum reaching 120 HZ and 480p [23]. Eye cameras are well thought out; the implementation of the “dark pupil” detection mechanism is a vital process in capturing the eyes.



Figure 2. Participant using the Pupil Core eye-tracking device.

In addition to all hardware components built into the device, the Pupil Core glasses also need a power supply for data management and transfer, so they are used together with a laptop/PC and the open-source platform provided by the developers (Figure 3). The software can be accessed and downloaded free of charge and consists of two essential parts: the Pupil Capture program used to record eye movements and the Pupil Player for playing and analyzing the recorded data [22]. Pupil Capture is a program that interprets video signals received from the three existing cameras, namely the world camera and the two eye cameras. The program has the functionality to detect the pupil, follow gaze, and indicate time markers on the surface where movement is observed [24].

Pupil Player has the role of playing videos with data recorded through Pupil Capture. It is the primary tool used to view and analyze data, and as in the cases mentioned above, it has an interface called Player Window. The system also includes an algorithm that uses a model-based approach. It is called Pye3D and implements a 3D mathematical eye model for capturing kinematics and eye optics. Eyeball position estimates are defined by binocular cameras in the 3D coordinate system and are based on frame-by-frame measurements of gaze and pupil size [25].

Another essential process for proper device operation is the calibration of the cameras; the data collected during calibration are used to correlate the scene camera with the ocular cameras [26]. Although there are different ways to perform the calibration process, the basic principle does not change. Users must follow a specific point on the device screens or in the real world according to the markers defined. The way markers are presented on the screen is called choreography [24] and is highly dependent on the calibration method (Figure 4).



Figure 3. Laptop used as power supply and data management and transfer during the testing session with the Pupil Core eye-tracking device.



Figure 4. Markers placed around the smartphone display to calibrate the Pupil Core eye-tracking device.

4. Methodology

The current section describes the profile of the users who participated in the study and the evaluation procedure.

4.1. Participants

One of the aims of the Spotlight Heritage Timisoara project is to transform the plain passion for technology into a passion for culture, art, and heritage, with the help of technology. In this sense, our target group for the Spotlight Timisoara AR application is young people, addicted to technology, with basic knowledge of the city of Timisoara and wanting to know more about its history and heritage. We applied a screener to recruit participants, and, finally, six persons were selected, based on their availability to participate in an in-person moderated usability testing session with the use of eye tracking and the fact that they have not used the Spotlight Heritage application before. Their profile is described in Table 1.

Table 1. Demographic information for participants.

Participant	Age Group	Gender	Occupation	Domain	Smartphone Proficiency
P1	19–25	F	Student	Psychology	Advanced
P2	19–25	F	Student	Psychology	Advanced
P3	26–35	M	Employed	Sales	Advanced
P4	26–35	F	Freelancer	Advertising	Advanced
P5	26–35	M	Student	IT	Intermediate
P6	19–25	M	Employed	IT	Advanced

4.2. Procedure

The tests were carried out in the open air, in the proximity of one of the landmarks described in the Spotlight Heritage project, namely The Palace of the Southern Region Casino (Figure 5). The old photo was initially taken from a position where today there is a tram station, so the tests had to be carried out while standing on the tram platform.

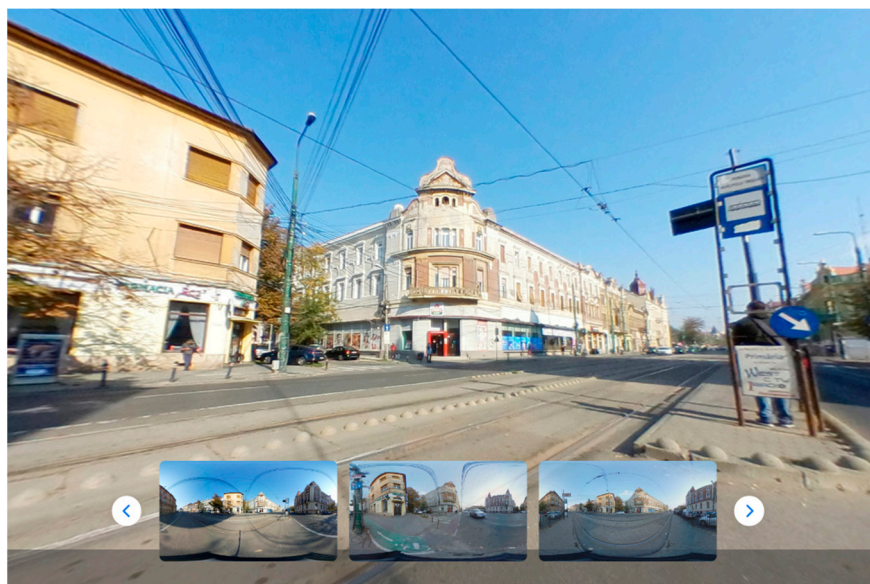


Figure 5. Panorama view of the tested location—The Palace of the Southern Region Casino.

The study started with the reception of the participants one by one, the moderator of the test being the one who waited and led them to the spot where the action took place. They then received an overview of what was going to happen in the experiment and were asked to sign the recording and confidentiality agreements. The moderator explained that only the movement of their eyes and what they were seeing would be recorded; some photos from the tests would be taken if they also agreed.

Participants were asked to answer a round of open, verbal questions about the experience, satisfaction, or frustration they had previously encountered using mobile applications.

In the end, the researchers invited participants to ask all the questions they had before starting the test; after that, participants were advised to try to perform the tasks on their own until completion or until they concluded that the tasks could not be finalized. After making sure that the participants understood everything that was explained to them, the moderator and her assistant started the eye calibration process, which was needed to accurately collect eye movement data as participants interacted with the application.

The usability test was designed in the form of a single scenario as follows: “You want to know more about the culture, monuments, and history of the buildings of the cultural heritage of Timisoara. A friend recommended the Spotlight Timisoara AR app, which can provide helpful information on the history of buildings and monuments in the city, and you have decided to use it.” The scenario was followed by five different tasks.

In the first task (henceforth named task 1, first impression), participants needed to find the Spotlight Timisoara AR app in the virtual store, download it, and explore the main sections, in order to gain a first impression of the app and its interface. Next, they were asked to change the application's language (task 2, change language), followed by the third task that guided them to find the landmark near which they are at that moment on the embedded map and explore more details (task 3, find landmark). The fourth task (task 4, view AR map) required the participants to explore their surroundings in AR mode, and the fifth task (task 5, scan with AR) led them to scan the nearest landmark and see with AR how it looked in the past.

While the participants performed the tasks, the moderator and the observer watched and took notes on what the subjects were doing and saying. The actual testing for each participant lasted about 45 min. At the end of the testing session, participants were asked to fill out a post-questionnaire consisting of the System Usability Scale and Product Reaction Cards. In addition, they responded to some open verbal questions regarding the interaction with the application, such as the aspects that the participant liked the most and the least, which was the most surprising feature, what part of the experience frustrated them (if true), how they would describe their interaction with the application, or what suggestions for improvement they have.

5. Results

The current section reports the results of the user testing sessions, the SUS questionnaire, the Product Reaction Cards method, and the eye tracking. While the authors have inserted some comments here, most of the discussion of the results happens in the next section.

The pre-questionnaire revealed that all participants have heard of or tried one or more AR applications, such as games (Pokemon Go), messaging (Snapchat), or clothing try-ons. They enjoyed the technology, but some of them noticed that it still has some glitches here and there.

Regarding their knowledge of the city, the participants declared that they are not sufficiently accustomed to its history and landmarks, but would like to know more about it.

5.1. User Testing Results

Quantitative data from the tests have been analyzed to derive the level of effectiveness, efficiency, satisfaction, and frustration that participants indicated when using the Spotlight Timisoara AR application.

To measure the effectiveness of the application, we analyzed which tasks were successfully completed and which were not. Tasks 1 to 4 were finalized by all participants, while task 5, "scan with AR" was completed successfully by all except P5 and P6, who gave up because scanning the buildings did not produce an effect in the app. The frustration of the user was driven by the fact that the application did not display any message about a failure to scan.

In terms of efficiency, we measured the time it took participants to complete each task (Table 2).

Table 2. Time per task for each participant (in seconds), with averages and standard deviations.

Participants	Task 1	Task 2	Task 3	Task 4	Task 5	Total Time
P1	56	14	67	30	63	230
P2	72	16	75	62	58	283
P3	67	14	79	39	47	246
P4	84	15	24	33	57	213
P5	58	17	38	66	n/a	179
P6	49	14	51	39	n/a	153
AVG	64.33	15	55.66	44.83	56.25	-
STDEV	12.62	1.26	21.83	15.30	6.70	-

Task 2, “change language” was the fastest and easiest task to solve, with three out of six participants recording the same completion time, that is, 14 s. In contrast, the others were only a few seconds apart in completing the requirements. When a user is unfamiliar with the platform, in order to perform specific tasks, they rely on instinct and associate similar actions performed in other applications with the actions they need to do in the current application. This aspect indicates that the elements used to change the language in the Spotlight Timisoara AR application are in the right place; the participants, even if they were not familiar with the application, went instinctively fast through the settings and made the corresponding modifications.

For task 3, “find landmark”, the slightly larger difference between participants in completion time is explained by the fact that some participants were better acquainted with the city and were able to find the nearby landmark faster than others. Future versions of the application will need to implement a “locate me” button to help users find their place on the map much faster.

Qualitative data were collected through the analysis of the participants’ body language, facial expressions, and think-aloud comments, but mainly through the post-questionnaire.

When participants were asked about their general thoughts about the application, all gave positive feedback, even those who did not manage to complete all tasks. Navigating landmarks with the AR map in real time (task 4) attracted most of the participants’ interest.

However, they also expressed their frustrations regarding task 5, “scan with AR”. Five out of six participants said that scanning the building’s facade was very hard, with information displayed for only a few seconds on the screen and not enough time to notice every detail, something that needs to be fixed in the next versions.

5.2. SUS Questionnaire Results

Table 3 lists the results of the SUS questionnaire applied to participants, who were asked to rate each question from 1 (strongly disagree) to 5 (strongly agree), depending on how much they agree with the statements [27] about the Spotlight Timisoara AR application.

Table 3. Individual participant SUS scores.

<i>p</i>	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	Score
P1	2	2	5	1	3	2	5	1	4	2	77.5
P2	4	1	4	1	4	2	4	2	3	4	72.5
P3	2	1	4	1	4	3	5	2	3	2	72.5
P4	4	1	4	1	5	2	5	1	4	2	87.5
P5	4	4	5	2	4	4	5	2	4	2	70
P6	2	1	4	2	4	2	4	2	2	2	67.5

According to the study of A. Bangor et al. [9], the score range for the SUS questionnaire can vary from 0 to 100. Anything less than 50 is considered poor and means that a rethink of the product is needed. The Spotlight Timisoara AR application received a score of 74.58 points, which is rated as Good on the adjective rating scale, which means that users are satisfied with the application, but there is room for improvement.

5.3. Product Reaction Cards Method Results

Table 4 lists the product reaction words chosen by each participant (they were allowed to pick between three and five cards).

Table 4. Product reaction words chosen by each participant.

Participants	1st Card	2nd Card	3rd Card	4th Card	5th Card
P1	Creative	Understandable	Annoying	Interesting	Uncontrollable
P2	Accessible	Creative	Frustrating	Interesting	Useful
P3	Clear	Creative	Inconsistent	-	-
P4	Attractive	Good quality	Exciting	Complex	Flexible
P5	Creative	Effective	Effortless	Interesting	Friendly
P6	Direct	Fragile	Optimistic	-	-

Most of the cards are positive ones (20 out of 26), with Creative and Interesting appearing the most. Some of the negative cards, such as Fragile, Annoying, and Inconsistent, describe the (still) imperfect experience of scanning the surroundings with augmented reality technologies.

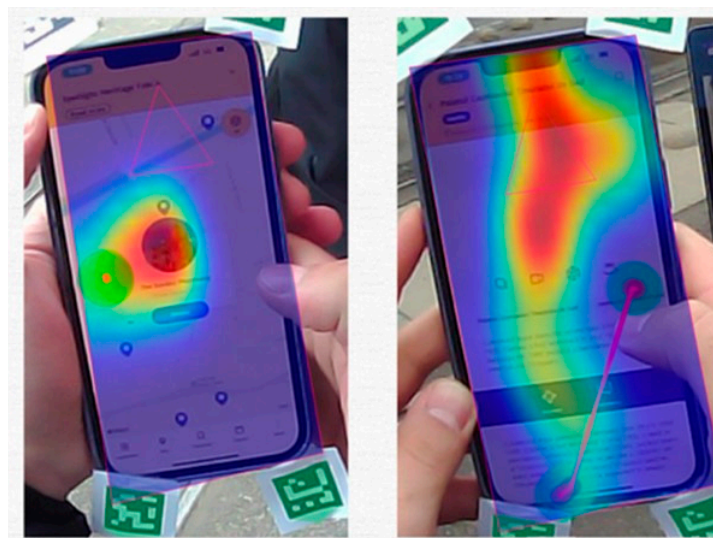
5.4. Eye-Tracking Results

The most important metrics used to analyse the results of the recordings were the number of the fixations registered from specific areas of interest on the application's interface, along with the heatmap, which helped us discover and compare the amount of time users spent looking at some elements while performing the tasks.

With fixations, we can define how much time the participant spent gazing at an element. The longer the fixation takes, the more likely the user is to be having a moment of confusion. Relevant aspects of the fixations include location, duration, dispersion, and confidence. The selected duration range for fixations in this study is between 80 ms and 220 ms, this being the most suitable interval to measure and identify them. Furthermore, the confidence range of the eye tracker varied between 0.89 and 0.96.

For example, in task 4, "view AR map", participants had to find the connection between the standard map view of the area and the augmented reality street view. Five out of six participants localized the AR button very quickly, with a time range of the first fixations between 1.5 and 2 s. That means that the button was well placed in the application's interface, and the users immediately recognized the actions needed to accomplish the task. Furthermore, task 3, "find landmark", and task 5, "scan with AR", which took a long time to solve, also had a higher number of fixations with extended durations.

The eye-tracking heatmap was another tool for extracting data from participants' actions and analyzing them accordingly. In 80% of the situations, participants paid attention to images or call-to-action buttons, the text being just skimmed (Figure 6).

**Figure 6.** Heatmap of the reading pattern in the application.

The scanned building also presented a high level of interest, most of it concentrated only in the upper left part of it. This behavior suggests that participants no longer gave importance to the elements in the surrounding environment, but only to the facade of the building, waiting to see the results. A difference was observed between the gazes of the participants among those who failed to complete the fifth task (scan with AR). They gave more importance to the information button displayed on the top right of the screen, unlike those who succeeded and did not need help (Figure 7).

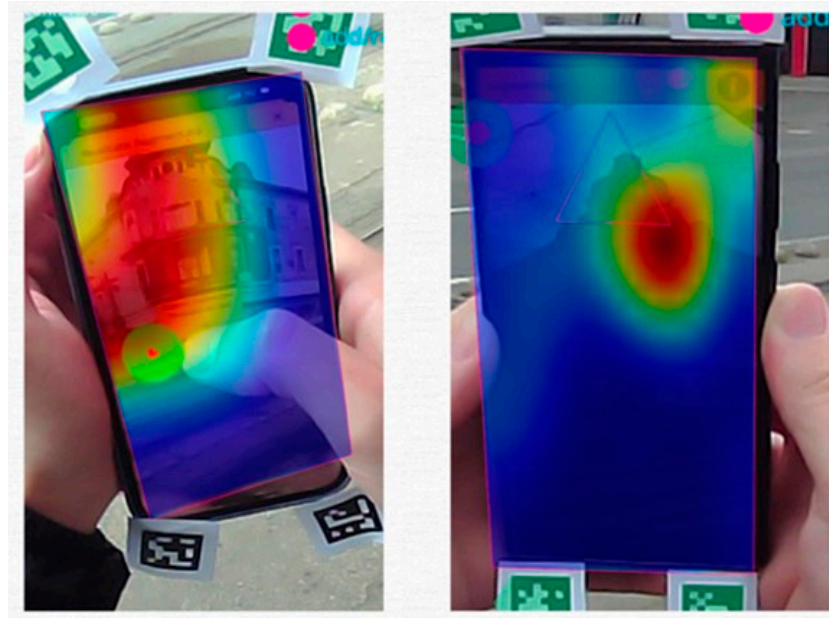


Figure 7. Heatmap from the augmented reality building scanning.

6. Discussions

The current section presents our interpretation of the previously mentioned results and highlights some threats to the validity of the study.

6.1. Interpretation of the Results

The evaluation results indicate that the Spotlight Timisoara AR application was well received by the participants, this being demonstrated by the following: the SUS questionnaire, which indicated a median score of 74.58 points, which corresponds to Good on the SUS adjective rating scale; the Product Reaction Cards method, where most of the chosen adjectives were positive (20 out of 26); the post-questionnaire, which indicates that the participants enjoyed the experience and would use the application again in the future.

However, the participants expressed concerns and even frustration about some aspects of the AR experience. This is made obvious by the participants in their responses to the post-questionnaire and the negative adjectives that some chose during the Product Reaction Cards task.

The general good opinion about the application, despite its frustrating parts, seems like a paradox. This could be explained by the fact that the pre-questionnaire revealed that the participants already had some experience with using AR applications and were aware that it is still an imperfect technology.

One major usability problem is the fact that the map does not have a “locate me” feature, which would help users know where they are at any given moment and, consequently, what landmark they are able to scan there. This is a real hindrance for locals who moved to the city a short time ago and even more so for tourists who are visiting the city for the first time.

Another major usability issue is the lack of a “system status” in the application (the first of Jakob Nielsen’s ten heuristics [28]) when scanning the landmark with AR. Participants

complained that, in the beginning, they were not sure if the application started scanning the landmark or not; after a long time, when no results showed up, they were not sure if it was taking a long time to process the captured image or if it was not able to detect anything. Some of them continuously tapped on the screen, to no avail, thinking that this might make the algorithm work better (as when they tap to focus the camera in the usual scenarios of taking a picture).

The eye-tracking method revealed that participants briefly scanned the Information button in the top right corner, during the confusing moments of nothing happening on the screen, but were not enticed enough to tap on it for additional information or help. The next versions of the application should implement a subtle animation on the button to cause the users to become more curious about tapping on it. Another solution would be for the application to display a brief onboarding tutorial immediately after starting the AR scanning process.

The glitchy sensation of the AR experience was also due to a moderate usability problem, which consisted of the fact that, sometimes, the old photo of the building appeared very briefly and disappeared each time the smartphone was moved a bit. This happened because of the very tight dependence of the computer vision algorithm on the image captured by the camera. To avoid such glitches, the application should leave the old photo on the screen for more time, until the scanned building completely disappears from the camera.

In addition, some participants expected the 3D registration to be more accurate during the AR experience [29]. This did not happen because the old photo of the building was not perfectly overlaid on the actual landmark as captured by the smartphone camera. The application allows the users to pinch-to-zoom and move the old photo on the screen in order to make it 1:1 with the actual building. However, these gestures are described only on the Information screen, which the participants did not access.

The eye-tracking method pointed out another issue with the AR experience. Reviewing the recorded videos afterward revealed that participants were almost completely focused on what happened on the screen and paid very little attention to what was happening in their surroundings. This could potentially have a negative effect on their safety since to be able to properly scan the buildings from the angle that best displays the old photo on top of the actual landmark, users need to stand in crowded places, on sidewalks, or even on tram platforms (as in the studied use case). The application should warn the user, when it detects ample movement, to be aware of the surroundings.

This study also highlighted some challenges of integrating eye tracking into user evaluation of mobile augmented reality applications for urban outdoor cultural heritage.

First, special considerations needed to be taken during eye tracking because of the need to keep glasses connected to the laptop at all times. This requires an additional person to carry the laptop and reduces the total available time for testing due to limited battery capacity.

In addition, four QR codes need to be permanently attached to the corners of the smartphone to maintain the calibration. This might impede users in their natural manipulation of the smartphone. A contribution to this was also brought about by the position in which participants needed to hold the smartphone in order for the eye tracking system to record their eye movements in the optimal mode. The participants had to hold the phone at face level almost all the time, which was not comfortable.

Consequently, these aspects create an inconvenience to the users which might alter the way they use the smartphone in usual circumstances, thus altering the results of the evaluation. This disadvantage can be alleviated by using other hardware that does not require calibration with QR codes and a permanent connection to a laptop.

Second, testing outdoors requires good weather conditions. Very bright sunlight or wind and rain (as happened briefly during the testing of the Spotlight Heritage AR app) severely impede the user testing process, impeding it even more when eye-tracking equipment is used. This is not only because of the complexity of the hardware that needs to

be protected from bad weather, but also because inappropriate weather distorts the image of the landmark to be recognized.

Third, as opposed to performing eye tracking on usual smartphone apps, in the case of AR experiences, other objects or people can frequently occlude the vision (as happened with moving trams in our case study, due to the fact that the tram platform was the optimal point from which the building could be scanned in order to properly overlay the old photo). This makes the participants pause or focus their attention on other places, thus possibly altering the results of the testing data.

In conclusion, while eye tracking can reveal useful insights for testing AR applications for outdoor urban cultural heritage, researchers must carefully take into account the overhead it adds to the user evaluation process.

6.2. Threats to the Validity of the Study

We acknowledge the existence of potential threats to the validity of our study.

First, there is a selection threat to the validity of the moderated usability testing sessions, as participants were not chosen randomly. Instead, convenience-based sampling was used; i.e., the authors approached people in their personal or professional circles who were in the target group of the Spotlight Timisoara AR application. They selected, on a “first-come, first-served” approach, those who expressed their availability to participate in an in-person moderated usability testing session, with the use of eye tracking, and who had not previously used the Spotlight Timisoara AR application.

Second, the number of participants—six—employed in this study is rather small (although we employed more than five, which is considered by Jakob Nielsen to be the optimal number of participants in a single iteration of a usability test [30]). This is because the aim of the research was not to determine (almost) all the usability problems, but to gain some deeper insights into the usability of the application and to derive some big challenges in performing usability testing with eye tracking on such applications.

7. Conclusions

Digitalization of cultural heritage has attracted the interest of art lovers and experts as a viable means of preserving, restoring, and promoting cultural heritage. Among current technologies, augmented reality is increasingly popular as a concept for integrating and mixing digital aspects into consumers’ physical worlds.

In this study, we evaluated the usability of the Spotlight Timisoara AR application, part of the Spotlight Heritage Timisoara project, a digital storytelling platform for the city of Timisoara (Romania), European Capital of Culture in 2023. The AR app allows users to scan the historical landmarks of Timisoara and show how their facades looked in the past.

To gain a deeper understanding of the user experience of the application, we conducted a usability assessment with eye tracking and more traditional methods, such as semi-structured interviews, observations, think-aloud protocol, SUS questionnaire, and Product Reaction Cards. The study, which was conducted outdoors, in a historical neighborhood in Timisoara, in June 2022, had six participants who were in the target group of the application. The usability tests were conducted by a moderator and an observer, and the eye tracking was performed with a Pupil Core pair of glasses.

Pupil Core is an eye-tracking pair of glasses that features two cameras with an infrared spectrum oriented to the eyes and a universal scene camera oriented to the front. When connected to a laptop or PC, the glasses can record and provide accurate data on eye movements and gaze regardless of head position.

The evaluation results indicated that the application was well received by the participants, despite some flaws in the AR experience. The main usability problems consisted of the absence of a “locate me” feature to help users orient themselves on the map, the lack of a “system status” during the AR experience, the hidden user tutorial, the glitchiness of the AR experience due to the rapid change in the information on the screen and the bad 3D

registration of the old photo of the facade, and the absence of a way to warn users to be constantly aware of their surroundings.

Since the scientific literature lacks a study describing the integration of eye tracking in the user evaluation of mobile augmented reality applications for urban outdoor cultural heritage, our aim was also to determine the main challenges of such an integration.

We determined that such challenges consist of complex calibration and dependence on connectivity to a laptop, which might impede the natural usage of the smartphone, inappropriate weather for using the eye tracking hardware and for scanning with AR, and occluding objects which disturb the natural flow of the evaluation.

We concluded that while eye tracking can provide useful insights when testing mobile augmented reality applications in urban cultural heritage, researchers must carefully consider the cost of adding it to the user evaluation process.

In the end, we acknowledged as possible limitations of the study the reduced number of participants and the convenience-based sampling method of selecting them, but we argued that they were representative of the target group of the application. As future work, we intend to solve the identified usability problems and run another round of usability tests with eye tracking, this time extending the number and diversity of participants, in order to obtain as many benefits as possible [31].

Author Contributions: Conceptualization, S.V.; data curation, D.S., S.V. and O.R.; funding acquisition, S.V. and D.A.; methodology, D.S, S.V. and O.R.; project administration, S.V.; software, D.S.; supervision, S.V.; validation, O.R. and D.A.; visualization, D.S. and O.R.; writing—original draft, D.S. and S.V.; writing—review and editing, O.R. and S.V. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Politehnica University of Timisoara under grant number 10162/11.06.2021.

Institutional Review Board Statement: The study was conducted according to the Ethical Regulations and Guidelines of the Politehnica University of Timisoara, Romania (http://upt.ro/Informatii_etica-si-deontologie_164_ro.html (accessed on 30 May 2022)).

Informed Consent Statement: GDPR data protection information was provided and respected, and informed consent was obtained from all subjects involved in the study.

Data Availability Statement: The data presented in this study are available on request from the corresponding author.

Acknowledgments: We gratefully acknowledge the contributions of all participants in our experiments, as well as those involved in the Spotlight Heritage Timisoara cultural project.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Boboc, R.G.; Băutu, E.; Gîrbacia, F.; Popovici, N.; Popovici, D.-M. Augmented Reality in Cultural Heritage: An Overview of the Last Decade of Applications. *Appl. Sci.* **2022**, *12*, 9859. [[CrossRef](#)]
2. Tiriteu, D.; Vert, S. Usability Testing of Mobile Augmented Applications for Cultural Heritage—A Systematic Literature Review. In Proceedings of the RoChi-International Conference on Human-Computer Interaction, Online, 22–23 October 2020; pp. 137–144. [[CrossRef](#)]
3. Vert, S.; Andone, D.; Ternauciuc, A.; Mihaescu, V.; Rotaru, O.; Mocofan, M.; Orhei, C.; Vasiiu, R. User Evaluation of a Multi-Platform Digital Storytelling Concept for Cultural Heritage. *Mathematics* **2021**, *9*, 2678. [[CrossRef](#)]
4. Goldberg, J.H.; Wichansky, A.M. Eye tracking in usability evaluation: A practitioner's guide. In *The Mind's Eye*; Elsevier: Amsterdam, The Netherlands, 2003; pp. 493–516.
5. Naspetti, S.; Pierdicca, R.; Mandolesi, S.; Paolanti, M.; Frontoni, E.; Zanolli, R. Automatic Analysis of Eye-Tracking Data for Augmented Reality Applications: A Prospective Outlook. In *Augmented Reality, Virtual Reality, and Computer Graphics*; Springer: Cham, Switzerland, 2016; pp. 217–230. [[CrossRef](#)]
6. Josephson, S.; Myers, M. Augmented Reality Through the Lens of Eye Tracking. *Vis. Commun. Q.* **2019**, *26*, 208–222. [[CrossRef](#)]
7. Hussain, A.; Mkpojiogu, E.; Musa, J.; Mortada, S. A User Experience Evaluation of Amazon Kindle Mobile Application. *AIP Conf. Proc.* **1891**, 2017, 020060. [[CrossRef](#)]

8. Eccles, D.; Arsal, G. The think aloud method: What is it and how do I use it? *Qual. Res. Sport Exerc. Health* **2017**, *9*, 514–531. [[CrossRef](#)]
9. Bangor, A.; Kortum, P.; Miller, J. Determining what individual SUS scores mean: Adding an adjective rating scale. *J. Usability Stud.* **2009**, *4*, 114–123.
10. Product Reaction Card—An Overview | ScienceDirect Topics. Available online: <https://www.sciencedirect.com/topics/computer-science/product-reaction-card> (accessed on 22 November 2022).
11. Hutton, S. Eye Tracking Terminology—Eye Movements. *SR Research Website*. 2 July 2020. Available online: <https://www.sr-research.com/eye-tracking-blog/background/eye-tracking-terminology-eye-movements/> (accessed on 10 August 2022).
12. Burr, D.C.; Morrone, M.C.; Ross, J. Selective suppression of the magnocellular visual pathway during saccadic eye movements. *Nature* **1994**, *371*, 511–513. [[CrossRef](#)] [[PubMed](#)]
13. Learn about the Different Types of Eye Movement—Tobii Pro. 6 August 2015. Available online: <https://www.tobii.com/learn-and-support/learn/eye-tracking-essentials/types-of-eye-movements/> (accessed on 10 August 2022).
14. Wang, J.; Antonenko, P.; Celepkolu, M.; Jimenez, Y.; Fieldman, E.; Fieldman, A. Exploring Relationships Between Eye Tracking and Traditional Usability Testing Data. *Int. J. Hum.-Comput. Interact.* **2019**, *35*, 483–494. [[CrossRef](#)]
15. Ho, H.-F. The effects of controlling visual attention to handbags for women in online shops: Evidence from eye movements. *Comput. Hum. Behav.* **2014**, *30*, 146–152. [[CrossRef](#)]
16. Ho, C.-H.; Lu, Y.-N. Can pupil size be measured to assess design products? *Int. J. Ind. Ergon.* **2014**, *44*, 436–441. [[CrossRef](#)]
17. Han, D.-I.; Jung, T.; Gibson, A. Dublin AR: Implementing Augmented Reality (AR) in Tourism. In *Information and Communication Technologies in Tourism 2014*; Springer: Cham, Switzerland, 2014; pp. 511–523. [[CrossRef](#)]
18. Safitri, R.; Yusra, D.; Hermawan, D.; Ripmiatin, E.; Pradani, W. *Mobile Tourism Application Using Augmented Reality*; IEEE: Denpasar, Indonesia, 2017; p. 6. [[CrossRef](#)]
19. Tobii. What Is Eye Tracking? How Do Eye Trackers Work?—Tobii Pro. 17 January 2018. Available online: <https://www.tobii.com/learn-and-support/get-started/what-is-eye-tracking> (accessed on 10 August 2022).
20. Orhei, C.; Vert, S.; Mocofan, M.; VasIU, R. TMBuD: A Dataset for Urban Scene Building Detection. In *Information and Software Technologies*; Springer: Cham, Switzerland, 2021; pp. 251–262. [[CrossRef](#)]
21. Vert, S. Spotlight Timisoara AR—Mobile Augmented Reality with Historical Images for Cultural Heritage. In *Digital Culture in Education, Science and Technology*; IAFes Publications: Vienna, Austria, 2021; p. 186.
22. Pupil Labs. Core—Getting Started. Available online: <https://docs.pupil-labs.com/core/> (accessed on 11 August 2022).
23. Pupil Core—Eye Tracking Platform Technical Specifications—Pupil Labs. Available online: <https://pupil-labs.com/products/core/tech-specs/> (accessed on 11 August 2022).
24. Pupil Labs. Core—Pupil Capture. Available online: <https://docs.pupil-labs.com/core/software/pupil-capture/> (accessed on 11 August 2022).
25. Pupil Labs. Core—pye3d Pupil Detection. Available online: <https://docs.pupil-labs.com/developer/core/pye3d/> (accessed on 11 August 2022).
26. Kassner, M.; Patera, W.; Bulling, A. Pupil: An Open Source Platform for Pervasive Eye Tracking and Mobile Gaze-based Interaction. In *Proceedings of the UbiComp 2014—2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, Seattle, WA, USA, 13–17 September 2014. [[CrossRef](#)]
27. Lewis, J.R. The System Usability Scale: Past, Present, and Future. *Int. J. Hum.-Comput. Interact.* **2018**, *34*, 577–590. [[CrossRef](#)]
28. Nielsen, J.; Nielsen Norman Group. 10 Usability Heuristics for User Interface Design. Available online: <https://www.nngroup.com/articles/ten-usability-heuristics/> (accessed on 4 September 2022).
29. Azuma, R.T. A survey of augmented reality. *Presence* **1997**, *6*, 355–385. [[CrossRef](#)]
30. Nielsen Norman Group. World Leaders in Research-Based User Experience. How Many Test Users in a Usability Study? Available online: <https://www.nngroup.com/articles/how-many-test-users/> (accessed on 6 June 2022).
31. Faulkner, L. Beyond the five-user assumption: Benefits of increased sample sizes in usability testing. *Behav. Res. Methods Instrum. Comput.* **2003**, *35*, 379–383. [[CrossRef](#)] [[PubMed](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.