A comparative study of tracking methods for a guided walking city tour in outdoor spaces for tourists through AR on smartphones.

En jämförande studie av spårningsmetoder för en utomhusapplikation för guidade stadsresor genom AR på mobiltelefoner.

LISA SCHMITZ
Sammanfattning

En jämförande studie av spårningsmetoder för en utomhusapplikation för guidade stadsresor genom AR på mobiltelefoner.

A comparative study of tracking methods for a guided walking city tour in outdoor spaces for tourists through AR on smartphones.

Lisa Schmitz
KTH Royal Institute of Technology
Stockholm, Sweden
lschmitz@kth.se

ABSTRACT
Recent advancements in mobile phone technology have allowed mobile augmented reality (MAR) to become feasible. Today’s mobile phones have enough computing power to display augmented reality content and new frameworks make MAR development more accessible. It is no surprise that one of the most popular areas of applications are city tours as this has been a target field since the early days of augmented reality (AR) [8]. Without altering the appearance of the city, virtual content can be placed to bring hidden information, such as the city’s history, closer to tourists. The most common choice of the tracking method for this type of application is location-based tracking. Relying only on the GPS signal and sensors like the accelerometer and the gyroscope, the position of the phone is tracked. The location of the digital content in the real world is given by geospatial coordinates. Unfortunately, the accuracy of the sensors is insufficient for accurate placement. Furthermore, the technology’s main advantage over other techniques, such as marker-based tracking, is that the application does not require any change in the city environment. In contrast to that, the other leading technique, marker-based tracking, is a computer vision technology that requires visual clues to work. Marker images would have to be placed in the city for the marker-based tracking technology to function. However, location-based tracking can cause erratic behaviour of the virtual objects, which decreases the quality of the experience. This paper compares location-based and marker-based tracking to show the user experience strengths and weaknesses of both methods to provide design guidelines for choosing the most suitable tracking technology when developing an outdoor walking application. In order to uncover the strengths and weaknesses, one experimental prototypes for each tracking technology has been developed. The analysis of the results of a controlled user study highlights the comparative strengths and weaknesses of each technology, location-based and marker-based tracking. The measured user experience differences demonstrate that for scenes where AR application designers and city officials are leading to incorporate visual markers, visual-based tracking will outperform location-based tracking.

Keywords
Augmented Reality, AR, Location-based tracking, Marker-based tracking

1. INTRODUCTION
Augmented Reality (AR) is the term to describe the enhancement of the real world with digital content [3]. Milgram et al. [16] describe a virtuality continuum that ranges from Real environment to Virtual environment. On this scale, augmented reality is placed between both ends, but closer to the Real environment than virtual reality, for instance, as shown in figure 1.

![Virtuality Continuum after Milgram.](image)

The AR technology creates possibilities for new experiences in an already existing environment. In an optimal scenario, the digital object would meld together with the real world without the user being able to make a distinction between both. We are still far away from this ideal image, but the technology has evolved significantly in the last decades. The year 1965 is marked as the beginning of AR research [7]. This was followed by research in mobile AR systems (MARS) and a first prototype of such a system was developed in 1997 [8]. This system called Touring Machine was able to augment an outdoor campus walk with additional information like text labels in buildings. Since then, MARS has developed significantly. Today, a mobile phone provides the necessary computing power to run AR application. Furthermore, the battery life on today’s smartphones provide sufficient power to run AR applications long enough to accomplish the most common tasks in AR environment, such as information lookup and model visualization. With an increase in the popularity of AR and virtual reality (VR) technologies in the last years, increasingly more applications use this new technology. A large part of those applications are targeted at tourists and mobile city walking applications are one of them. The choice of tracking technology for those applications has mostly been location-based tracking [8, 9, 10, 11, 12]. In the last years machine learning has led to great advances in computer vision and the hardware in even the smallest pocket computers like mobile phones is capable of running such technology.

Furthermore, new frameworks like the Argon browser [14] are leading the way for a faster and platform-independent development of MAR applications. Unfortunately, the tech-
nology is not at a point yet where the requirements of an ultimate AR experience can be achieved. Solutions for realistic simulations of physical interactions have to be developed to accomplish a seamless transition between virtual and real world content. One obstacle that has to be overcome to create an ideal AR experience on mobile phones is the sensors. They are inaccurate and therefore cause erratic behavior of the digital content. A prior study from 2017 [13] evaluated a user-centered design approach of a MAR application using location-based tracking. It showed that the erratic behaviour, which was caused by the sensors, confused the users and decreased their enjoyment of the application. A lot of resetting and patience on the user’s side was necessary until a good experience was reported. This illustrates the problems and the possibilities that MAR bears at the moment: the mobile phones on the market are capable of creating a good AR experience, but the sensor accuracy interferes and turns it into a tedious interaction. Therefore, this study questions the choice of location-based tracking for MARS. Marker-based tracking is a stable alternative and might be a good choice as a tracking method not only for indoor but also for outdoor applications whenever the smartphone sensors are inaccurate.

1.1 Research question

What are the user experience differences between location-based tracking and marker-based tracking when used to display virtual content on a smartphone for an AR city walking tour as measured by quantitative data from time spent to solve interaction tasks, the changes of roll, pitch and yaw of the camera while solving the tasks, a questionnaire with Likert scale questions and qualitative data from observations and answers to a semi-structured interview?

2. BACKGROUND AND THEORY

This section provides background knowledge of the tracking methods that are most commonly used for developing AR applications. It also describes the history of MARS development and recent work in the field of mobile outdoor walking tours. Building from the state-of-the-art, the motivation for this study is illustrated in more detail.

2.1 AR Tracking Technologies

Placing virtual content in the real world environment and making it appear to be a physical part of it is one of the key requirements any AR system should fulfill. Therefore, it is necessary to have knowledge about the viewer’s pose in relation to the digital content. To gather this information, tracking technologies are used to create an abstract model of the environment. Three main categories of tracking methods can be noted today which are 1) vision-based tracking; 2) sensor-based tracking; and 3) a combined solution called hybrid-based tracking [19]. The following paragraphs will examine the first two technologies in greater detail. As hybrid-tracking does not introduce new concepts that are not already part of image-based or sensor-based tracking, the study does not focus on this technology.

2.1.1 Image-based Tracking

Image-based tracking relies on computer vision algorithms that can interpret data collected by optical sensors such that information about the view pose and the virtual content is gained. The three types of sensors that are used in image-based tracking are infrared, visible light and 3D structure sensors [4]. Infrared sensors have mostly been used during the early stage of research and are not part of the standard equipment of a mobile phone. However, the camera stream acts as a visible light sensor. The camera data can be captured and used as input data for computer vision algorithms to track the position of the virtual content. Tracking techniques that use visible light data are further categorized into marker-based tracking, natural feature detection and model-based tracking. Markers, sometimes called fiducials, are feature-rich images that have to be added to the physical environment to calculate the position and orientation of the camera. Markers can come, for example, in the form of a QR code. In contrast to this method, natural feature detection does not require a modification of the environment. More computational intensive algorithms are required to detect corners and edges in the camera stream. The other tracking technology that uses image data captured by a camera is model-based tracking. This method tracks real world objects using 3D structures as a basis. Understanding 3D motion with a monoscopic camera has not yet proven successful, which is why this field is dominated by 3D structure sensors that use dedicated hardware instead. A recent example for a tracking technology that uses 3D structure sensors is Google Tango1. It is an AR framework that uses data collected by a depth-sensor to accomplish tracking. Unfortunately, as of June 2017, only two available mobile phones are equipped with the necessary hardware to make use of the technology1. Therefore, this cannot be considered as an appropriate choice to make AR accessible to the general public. In order to do so, marker-based tracking is the best choice at the moment (June 2017).

2.1.2 Sensor-based Tracking

Sensor-based tracking through magnetic, optical, inertial, acoustic and mechanical sensors is used to track the position of the device and the virtual content [19]. This tracking method is commonly used in MARS. Inertial tracking can be conducted by sensors like magnetometers, accelerometers and gyroscopes. They use gravity to determine the position and orientation and are therefore popular for outdoor applications. However, they are very sensitive to disturbances and they drift over time, which results in incorrect data. This often causes erratic behaviour of the virtual content [4]. AR systems on phones collect data through those inertial sensors and determine the user’s geospatial position.

2.2 Related Work

The work by Feiner et al. [8] marks a starting point for the research in MARS for outdoor environments. As an experimental MARS prototype, the Touring Machine provided virtual text labels on buildings to guide the user through the environment. The system had to be worn as a backpack and showed the digital annotations on a see-through head-worn 3D display. Feiner et al. chose location-based tracking using GPS and magnetometer/inclinometer to determine the position of the user and to place the labels correctly in the real world.

Later research [10] showed the potential of a MARS as a multimedia documentary walking tour application. The system

1https://get.google.com/tango/
was used to present the history of the Columbia University campus. The content provided to the user consisted of images, audio, videos, hyperlinks, 3D models and 360° panorama images. This aspect is still surprisingly relevant over 15 years later, since most modern mobile AR tourist guides offer a subset of this content selection [11]. Since that time, hardware developed significantly and, fortunately, it is no longer necessary to carry around an 18 kilogram construction on the back to have an immersive AR experience. Mobile phones are indeed very convenient and powerful platforms to create AR applications that are accessible to everyone. Since it has been shown that AR can increase the learning experience [6], it is an effective medium to bring the history of a city closer to tourists. This is one reason why many commercial, but also scientific projects, have been developed in this area in the last years [5]. As one of the first AR applications for mobile phones, the Wikitude browser [1] should be mentioned. Using GPS and Wikipedia entries, it overlaid information for certain points of interest on the real-time camera view of an Android phone and therefore, had the fundamental function of almost every guided walking tour application already implemented.

A more recent application is CityViewAR developed by Lee et al. [12]. This mobile application has been created in response to the earthquake from 2011, which damaged great parts of the city of Christchurch. It uses location-based data provided by smartphone sensors to display content in form of images, text, panorama pictures and 3D models related to the sights that have been damaged by the earthquake. To solve the reliability problem of geospatial data, certain view points have been chosen to guarantee a reliable calibration of the smartphone sensors and with it, a more stable presentation of the 3D models. This is only one example that shows the downsides of location-based tracking. Choosing the tracking method, which is most suitable for a task, is an important factor when it comes to developing a guided city walking tour. Smartphone sensors are not accurate enough for a correct determination of the spatial position. Not only can we register an increasing amount of AR applications but also one of AR toolkits and frameworks [5]. Easy and free to use frameworks like Vuforia 2 and Argon AR [14], that have been released in the last years, make the development of AR application accessible and allow rapid prototyping. The Argon browser application has been developed with the goal of creating an AR environment that works as a platform for various virtual content. With the help of the Argon JavaScript framework, AR web applications can be developed and are able to be viewed in the Argon browser. Furthermore, the application supports image-based tracking with Vuforia as well as location-based tracking for mobile phones. This is why I chose Argon JavaScript as a particularly suited framework for the development of the prototype for this study. In fact, the Argon framework has been popular for AR research in presenting cultural heritage [17]. For instance, Midtown Buzz [9] lets the user explore the Midtown neighborhood showing 3D models of future buildings, marking key locations of the town’s history and providing access to videos relevant to certain locations. A more narrow research field in the area of MARS are tracking methods. Most papers, that have been published in the last years, present innovative tracking technologies [5]. However, very little has been done recently to compare and to evaluate the capability of tracking methods to display content under various external conditions that appear in an outdoor walking tours.

In an early study, Meyer et al. compare mechanical, optical, magnetic and acoustic implementations of position-tracking [15]. The results of the study, together with a catalogue of applications that make use of particular tracking methods, the paper works as a guideline for designing applications that need to use position tracking.

More recent research has been conducted to classify VR tracking methods and rate certain properties like accuracy, stability and range [4]. A study specifically set in the research field of AR covers the desired properties of AR tracking methods [2]. On the downside, the universality of the found requirements reduces its value for research today. A more focused study is necessary to equip developers with guidelines for the choice of a suitable tracking method.

3. MOTIVATION

AR has been a fascinating technology since the very beginning and has always driven the development of innovative applications. Unfortunately, many design-focused studies develop a complete prototype, for instance city walking tours, to evaluate the usability instead of focusing on testing a single important feature like the tracking method. Given the nature of a city walking tour, location-based tracking is the first choice of a tracking method as it doesn’t require modifications of the real world environment. This tracking method works well with virtual content that is not placed in the physical environment like a panorama view of a site. Even though this is often announced as an AR experience, such content cannot be classified as an AR application. Therefore, this study aims to compare location-based tracking and image-based tracking as being used in an AR application to display only virtual content that acts as part of the environment. A critical examination of both tracking methods is conducted in order to create guidelines for designing AR walking tour applications. AR has undoubtedly potential and mobile phones work as a driving power to further develop and research this technology. Nevertheless, it is easy to get carried away by fantastic ideas of possible applications without focusing on developing a thorough fundamental understanding of the possibilities and limitations that the technology has. Since user friendly MARS are still a novelty, more research is needed to gain intermediate knowledge and insight into the components of such a system instead of evaluating a complete system. Even when technologies evolve and new tracking methods arise, the knowledge about the tested tracking methods with certain properties has a persistent value [18]. The necessity of this can be seen by results of a recent study [13]. The study comes to the conclusion that using a mobile application to show a virtual object in the real world does only achieve a SUS score of 72. This means that, even with the improvements to counter flaws like the sensor inaccuracy, the application showed need for revision and was only barely acceptable as an actual mobile application. Since tricks to improve the usability have already been implemented, it seems advisable to take a step back and evaluate whether the choice of tracking technology should be revised. Even though there is a lack of
papers that focus on usability of tracking methods, the high amount of papers that have been published in the area of AR tracking methods\cite{5} in general indicates the need for improvement of the current available methods. By answering the research question how location-based and marker-based tracking compare to each other, a groundwork for further research in MAR development is laid.

4. METHOD

In order to evaluate the research question, two experimental prototypes have been developed, one using location-based tracking and one using marker-based tracking. The prototypes were used in a user study to collect quantitative and qualitative data and gain information about the difficulty of performing certain tasks. Those tasks were designed to cover situations and interactions that would be likely to be part of a mobile city walking tour.

4.1 Choice of Tracking Technologies

As shown in section 2.1, there is a large variety of tracking technologies available, but not all are equally suited to run on a mobile phone. The hardware on the phone has to support the tracking technology such that it can be considered a reasonable choice for a mobile AR application. The choice of method for the sensor-based tracking technology is straightforward as the sensors, that are part of the majority of mobile phones, are GPS, accelerometer and gyroscope. This is why location-based tracking with inertial sensors was one of the chosen tracking methods to be evaluated. As for image-based tracking methods, infrared and 3D structure sensors have not been considered since they are not part of the standard equipment of a mobile phone. I decided to use marker-based tracking as a representative for visible light tracking technologies as a second tracking technology. Natural features as well as 3D structure tracking are computationally heavier than marker-based tracking. To guarantee a good performance, marker-based tracking was the preferable technology.

4.2 Prototypes

In order to compare location-based tracking with marker-based tracking, two prototypes have been developed that each use one of the tracking technologies to display three different instances of virtual content in the physical environment. As the main purpose is to compare the two tracking technologies as applied to mobile outdoor walking tours, I designed the prototypes with this goal in mind. With the help of historical pictures, that have kindly been provided by the Spårvägsmuseet \footnote{http://sparvagsmuseet.sl.se/} in Stockholm, virtual content has been developed. The content has been taken out of scenes that are shown in the historical pictures.

4.3 Content Design

One of Azuma’s requirements \cite{3} for an AR applications is the following:

\begin{equation*}
\text{It is registered in 3D.}
\end{equation*}

This statement refers to the quality of the virtual content being part of the physical environment, which was an important aspect of the prototype design. Many applications that are categorized as AR do not augment the reality, in so far as that the content does not often become part of the environment. For instance, some studies about MAR applications for city tours name panorama views as the users’ favourite content \cite{12}. This does not strictly count as AR and does not convey the spirit of the typical AR experience. Therefore, content that was examined in this study uses only the real world as a background.

4.3.1 3D Object

Placing a 3D object in the real world is maybe the most classic instance of virtual content expected from an AR application. It is highly dependent on the environment since it has to line up well with the surroundings to achieve a good user experience. Challenges that come along with this are, among others, occlusion handling. I designed a 3D model of a tram stop sign with a schedule. An image of the model is shown in figure 2.

\begin{figure}[h]
\centering
\includegraphics[width=0.2\textwidth]{image2}
\caption{Picture of the 3D object.}
\end{figure}

4.3.2 Animated graffiti

Inspired by silhouette graffiti, a marker has been designed that shows a scene at Tegelbacken in the time from 1877 to 1907. I only used printed markers for this study, but it can be imagined that the graffiti could also be shown on dedicated walls in the city. The image-based tracking prototype uses the background as the marker and the tram is added as virtual content. In contrast to that, the location-based tracking prototype also displays the background as virtual content. An image of the animated graffiti can be seen in figure 3.

4.3.3 Timeportal

A timeportal classifies as 3D content and is mostly independent of the environment. The concept behind it is to build a window into another time to visualize historical scenes. This kind of portal can appear almost anywhere in the physical environment without looking out of place. For the study, a small scene has been designed that allows to have a glance at the location during another time period. An image of the timeportal is shown in figure 4.

4.4 User study

The prototypes, that have been described in 4.2, were used in a user study to evaluate how location-based and marker-based tracking for a mobile AR application compare. I instructed the participants of the study to perform tasks with
each prototype and handed them a questionnaire afterwards to evaluate the prototypes.

4.4.1 Experimental environment and participants

A formal study has been conducted in Tegelbacken in central Stockholm. Tegelbacken has been a station of trams, so called spårvagn, from 1877 to 1967. The area in which the AR content has been placed covers about 80m and is shown in figure 5.

In order to provide a safe study environment, the study did not take place at the exact historical site. Instead, the active part of the study has been conducted at an area only for pedestrian next to the original place of the tram stop. I made this decision in order to avoid any possible collisions with vehicles.

From the 21st of April to the 3rd of May, 12 participants took part in the study between 10am and 6pm. The participants were able-bodied in order to participate in a study that requires walking around outdoors. It was also mandatory that the participants were familiar with operating smartphones. The weather conditions at the times the study was conducted varied from sunny to cloudy but never rainy. The participants’ previous experience with AR ranged from 1 (I never used it before) to 5 (I use it almost everyday). Figure 6 shows the detailed distribution of the prior experience with AR. The participant group consisted of 11 computer science which is the cause for the relatively great experience with AR. Nevertheless, I tried to find participants with low and high degrees of experience alike. The average age of the participants was 29.33 and the median age was 26.
The goal of the study was to find out how location-based and marker-based tracking compare when used in a mobile city walking tour application. The two tracking technologies were the two main conditions and they were tested in between subject. I randomized the order of the subjects to counterbalance learning effects. The participants were divided beforehand into group I and II, without the participants' knowledge. Depending on the group, condition A - which would be tested first - was either location-based tracking for group I or marker-based tracking for group II. Accordingly condition B was marker-based tracking for group I and location-based tracking for group II.

<table>
<thead>
<tr>
<th>Group</th>
<th>Condition A</th>
<th>Condition B</th>
</tr>
</thead>
<tbody>
<tr>
<td>II</td>
<td>location-based</td>
<td>marker-based</td>
</tr>
<tr>
<td></td>
<td>tracking</td>
<td>tracking</td>
</tr>
<tr>
<td></td>
<td>marker-based</td>
<td>location-based</td>
</tr>
</tbody>
</table>

Table 1: Conditions A and B in group I and II

The study started with the instructions for the participants. The instructions the participants received contained information about the history of Tegelbacken, an explanation of the purpose of the study which was to compare marker-based and location-based tracking, a short overview of the three tasks the participants where about to perform and a reminder to be aware of the surroundings while using the prototype applications. This part of the study took place in a nearby coffeehouse. Subsequently, I guided the participants to the location of the active part of the study. At the location, I handed them the prototypes to test condition A and B. Condition A was tested first, followed by condition B. I gave the participants a questionnaire after testing each condition. The questionnaire instructions asked the participants to express their agreement with the statements in the questionnaire on a Likert scale ranging from 1 strongly disagree to 5 strongly agree. The questions were designed to gain quantitative data about the difficulty of the tasks. In both prototypes the users were shown the three different types of virtual content: the 3D object, the timeportal and the animated graffiti. The virtual content was presented one at a time in a fixed order. The participants had to perform three different tasks with each virtual content, depending on the tracking technology. The following were the tasks that had to be performed:

1. **Location-based tracking**: find the virtual content
   - Marker-based tracking: find the marker that displays the virtual content

2. **Location- and marker-based tracking**: interact with the virtual content by either rotating or repositioning a part of it

3. **Location- and marker-based tracking**: move around so that the red 3D cuboid does not block the view of the virtual content anymore, then take a screenshot

I intended the tasks to simulate conditions that occur during a city walking tour. The task design also takes the main requirements that Takuma defined for AR [3] into account: the interaction with the virtual content and the content being 3D. The second tasks covers the interaction with the virtual objects and the third tasks addresses the three-dimensionality of the objects. For this purpose, a red cuboid object was placed in front of the virtual object. The participants had to change the angle they were looking at the virtual object in a way that the red cuboid would not cover object. The third task was also designed to simulate the walking part of the walking city tour by forcing the participants to move around and change their position. In a city walking tour it can be expected that users move frequently. This is why the stability of the object position is an important feature.

While the participants performed the second and the third task, the roll, pitch and yaw of the camera was logged, as well as the time needed to perform the task. The normalized camera direction vector representing the roll, pitch and yaw, was logged every 60 frames with a frame rate of 30 fps. The change of the camera direction will show how much the users had to change the position of the device to accomplish a task. Higher changes of certain angles of the camera show that the viewer changed the camera position rapidly. This will most likely occur in situations where the viewer has problems to find a correct position to look at the virtual object. However, the roll, pitch and yaw was not logged for the first task since the study has been performed in a limited area and the search for the virtual objects cannot reasonably be compared. Since the main focus of the study lies on evaluating the tracking and therefore the display of the virtual content, task 2 and 3 are more suitable and sufficient to evaluate this aspect. After the participants had completed the tasks and filled out the corresponding questionnaire, I conducted a semi-structured interview to gain a better insight into the participant’s impression of the prototypes. The interview marked the end of the study.

The study tests were conducted on an Apple iPhone 6 with 1GB RAM, a 1.4 GHz dual-core 64-bit CPU and a ARMv8-A “Typhoon” and a PowerVR Series 6 GX6450 (quad-core) GPU.

5. RESULTS

This section presents the results that were gathered during the user study.

5.1 Interaction Difficulties

I constructed the study so that the results would give a better insight into the advantages and disadvantages of marker-based and location-based tracking in an outdoor environment. The difference in time needed to solve a task and the difference in the change of roll, pitch and yaw of the phone camera drew a clear image of the difficulty of rotating and translating parts of a virtual object and of moving around it. Overall, people spent more time to solve those two tasks with the prototype using location-based tracking than with the one using marker-based tracking (see figure 14c). This is also reflected by the results of the survey I conducted after testing each tracking condition, as shown in figure 8 and figure 9. Moreover, 11 of the 12 participants agree or strongly agree that it is easy to rotate / translate and move around a virtual object when using marker-based tracking.
tracking while only 9 agree or strongly agree when using location-based tracking.

I noted that all participants walked around systematically to accomplish the third task, that forced them to find a position where they could see the virtual objects without being occluded. A 3D object of a red box acted as a virtual occluder. Figure 7 shows example screenshots taken by participants from a position where the object was fully visible. All participants finished the tasks for the marker-based tracking and only two participants could not finish the third task for location-based tracking. Participant P1 gave up after trying to find a good position for around four minutes for each of the three objects. This participant later stated in the interview: it was difficult to find a position where the red box was not in front of the objects. I almost got it sometimes. Then I thought: just a little more and then it [the virtual object] jumped. P3 only struggled finding a good position for one of the three virtual objects. The reason she could not accomplish the task was a different one than for P1. While performing the task she said: I think I should go over there but I don’t want to go in that direction. and she pointed at a location where a group of people was sitting and was apparently drinking alcohol. This was the only time participants reacted directly to non-involved people.

5.2 Stability Issues

One unintended incident occurred while a participant was using the location-based tracking prototype. P4 found himself standing inside the virtual timeportal object and was confused about what had happened. It can therefore be noted that the inaccuracy of the mobile phone sensors makes it difficult to place a virtual object at a certain location and thus it can happen that users find themselves standing inside the virtual object. This can be very confusing and certainly does not contribute to a well-designed AR experience.

(a) Responses to the statement: It was easy to position and rotate the virtual objects.

(b) Responses to the statement: It was easy to find a position where the red box was not occluding the virtual objects.

Figure 8: Difficulty rating for task 2 [rotate / translate] and task 3 [move] using marker-based tracking.

(a) Responses to the statement: It was easy to position and rotate the virtual objects.

(b) Responses to the statement: It was easy to find a position where the red box was not occluding the virtual objects.

Figure 9: Difficulty rating for task 2 [rotate / translate] and task 3 [move] using location-based tracking.

Figure 10: Average responses to the statement: It was easy to solve the task.

5.3 Experienced Realism

Furthermore, 6 of the 12 participants agree that the virtual objects displayed with the marker-based tracking feel like a part of the real world (see figure 8), while only 4 agree to this when asked about the location-based tracking prototype, as can be seen in figure 9. A reason for this could be
the erratic behaviour of the virtual objects that 7 participants noticed in the prototype with location-based tracking (see figure 12). In contrast to that, only 2 participants noticed such behaviour when using the marker-based tracking (see figure 11).

(a) Responses to the statement: The virtual objects felt like a part of the real world.

(b) Responses to the statement: When I moved around, the virtual objects changed their position unexpectedly.

Figure 11: Experienced realism of the virtual objects with marker-based tracking.

Figure 12: Experienced realism of the virtual objects with location-based tracking.

5.4 User Experience of the Virtual Content

Since every virtual object presented in the test had different properties, not every object showed the same results when tracked either by location or by a marker. It cannot be assumed that every type of virtual content is experienced similarly with each tracking method, the results depending on the virtual content are presented for completeness. The participants’ favourite virtual object for both tracking methods is the same one, the 3D schedule. The reason that was most often given for making this choice was for both location-based and marker-based tracking that the object was the most realistic one. Statements supporting this as the main reason were:

It is most realistic. - P2

It was the closest to reality. - P7

It was positioned in a way that felt more grounded than the others. - P8

[...] it fit best into the environment. - P9

It feels more in the real world. - P10

However, the number of participants, who stated to appreciate the 3D model the most, varied for location-based and marker-based tracking. Of the participants, 75% reported that this is the most appreciated virtual object in the location-based tracking prototype, when only 50% of the participants stated the same for marker-based tracking. Figure 17 gives a detailed overview of the distribution of the votes for the most appreciated virtual object.

In this context it is interesting to look at the amount of time spent per task and the change of roll, pitch and yaw of the camera during each task for each of the virtual objects (see figure 14). In line with the results shown in figure 15 and in figure 16, it is clear that overall more time was spent on each task when using location-based tracking. Only the 3D schedule object constitutes an exception. The rotation and translation of parts of the virtual objects while using marker-based tracking took more time than while using location-based tracking. Likewise, the changes in the yaw of the camera are significantly higher for the tasks performed with marker-based tracking. This is the rotation around the vertical axis of the phone. This shows that the participants had a problem to find the right angle to look at the virtual objects. As expected, the roll and pitch however showed little to no significantly different changes. This is because rotation around those axes would not help to solve the tasks the participants had to perform. In particular task 3 required a change of angle around the vertical axis in order to avoid the occluder.

Figure 15 also shows that the change of yaw was signifi-
Task 2 (rotate / translate): Change of roll, pitch and yaw.

Task 3 (move): Change of roll, pitch and yaw.

Time needed to solve task 2 (rotate / translate) and task 3 (move).

Figure 14: Change of camera direction per task and time needed to solve each task.

Significantly higher for the timeportal during task 2 than for any of the other objects. The reason for this is most likely the design of the timeportal. Since the scene can only properly be seen when looking straight at it, it was more difficult for the participants to inspect the object properly while using location-based tracking. The inaccuracy of the GPS sensor turned this into a task that was exceptionally challenging. In contrast to that, the graffiti and the 3D schedule were still sufficiently visible from a larger variety of angles.

Participants who mentioned the 3D schedule object as the most appreciated one for the location-based tracking but not for the marker-based tracking stated the following reasons:

I liked the city hall in the background and the tram was closer to me. Felt more real in space.
- Timeportal, P3

Placed against a wall, the timeportal makes sense, and easier to rotate until you see the train with markers.
- Timeportal, P4

Graffiti, task 2 (rotate / translate): Change of roll, pitch and yaw.

Timeportal, task 2 (rotate / translate): Change of roll, pitch and yaw.

Schedule, task 2 (rotate / translate): Change of roll, pitch and yaw.

Figure 15: Change of camera direction per task and time needed to solve task 2 (rotate / translate) per object.

It felt most realistic.
- Graffiti, P7

Stronger interactive feeling with the real world.
- Graffiti, P10
Graffiti, task 3 (move): Change of roll, pitch and yaw.

Timeportal, task 3 (move): Change of roll, pitch and yaw.

Schedule, task 3 (move): Change of roll, pitch and yaw.

Figure 16: Change of camera direction per task and time needed to solve task 3 (move) per object.

P7 also stated in the semi-structured interview that interacting with the graffiti felt magical, like the newspapers in Harry Potter. And even though P6 did not choose the graffiti as the most appreciated virtual object, he stated afterwards in the interview: Now that I think about it more, I actually also appreciated the graffiti a lot. I see how it is cool to interact with the graffiti. Apart from that, one participant said the following while changing the position of the tram in the graffiti object using marker-based tracking: The tram isn’t really there, is it? and looked away from the mobile phone and at the marker. Moreover, even though the 3D schedule was the most appreciated object for the marker-based tracking, the graffiti and the timeportal together were appreciated by 50% of the participants as well. Therefore, the conclusion can be drawn that the realism of the object - which is dependent on how well an object fits into an environment - plays an important role when choosing a tracking method. The graffiti and the timeportal make sense to be placed against a wall and that was only accurately achieved with the marker-based tracking in this study. In contrast to that, the 3D schedule post was better placed further away from the wall, standing freely on the street. This assumption is supported by the following statement made during the semi-structured interview:

In general, the performance [of marker-based tracking] was better. But I definitely see the purpose for location-based tracking, so that the post can stand freely. But things like the portal, that make sense to be on a wall, are better with marker-based tracking. - P4 (uses AR almost everyday)
5.5 Safety Issues

Other aspects that have been pointed out during the interview were, that even though location-based tracking feels more realistic, this tracking technology requires a strict focus on the phone and 3 participants stated that they felt worried to bump into things. This is less likely to happen with marker-based tracking since it is not necessary to walk around as much as is needed for the location-based tracking.

Moreover, 3 participants mentioned the unrealistic positioning with the location-based tracking. However, still 7 participants reported that it was fun to look at the objects placed without markers and that they liked the exploration part.

Table 2 summarizes the findings of the study and shows the strengths and weaknesses of marker-based and location-based tracking that the study results point out.

<table>
<thead>
<tr>
<th></th>
<th>Location-based</th>
<th>Marker-based</th>
</tr>
</thead>
<tbody>
<tr>
<td>Interaction</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>Exploration</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>Realism</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Stability</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>Safety</td>
<td>-</td>
<td>+</td>
</tr>
</tbody>
</table>

Table 2: Strengths and weaknesses of location-based and marker-based tracking

6. DISCUSSION

This section discusses the results of the study and finishes with criticism of the study.

The study could show weaknesses and strengths of marker-based and location-based tracking under conditions that are likely to occur on a walking city tour, like the interaction with the virtual objects and walking around to look at the virtual objects from different angles. The inaccuracy of the GPS sensor led to erratic and unexpected movement of objects as well as to difficulties at interacting and inspecting the virtual object. However, the feeling of novelty and exploration of the location-based tracking technology outweighs the one of the marker-based tracking. Nevertheless, the weaknesses of location-based tracking are not easy to overlook. Not only does the erratic movement confuse people, in particular if they have little or no prior experience with AR, but also does it exacerbate to view the virtual objects from different angles. This limits the amount of suitable objects immensely. Careful design and placement of the objects can counterbalance this negative effect. The virtual objects would then have to be placed in a safe area that allows walking around without getting into danger. A more interesting alternative could be provided with marker-based tracking instead. Since the markers used for the study were set up in a very limited area and were obviously not part of the scenery, it was easy to find the markers. Though, it took a possible feeling of exploration away from the marker-based tracking part. Participants who stated that they preferred location-based tracking because it was more fun to run around and look for the objects, might have changed their opinion if the marker-based tracking approach had provided them with a similar aspect. Fortunately, location-based tracking is one possible method that can guide the user to the markers’ locations. Moreover, it appears to be a good approach to augment only parts of the physical environment with virtual content instead of the entire physical space. Several participants expressed astonishment when interacting with the augmented graffiti image. Some even said they preferred to look at entire scenes instead of a plain object. Those objects also bear more possibilities of interaction. It has been noted that, even though it was pointed out to all participants in the instructions that the slider on the user interface should be used to manipulate the virtual object, some instinctively touched the part of the virtual object that they were supposed to manipulate on the screen. For example, participants touched the tram in the graffiti image and tried to move it by dragging it around. This behaviour creates possibilities for different kind of interaction with objects in a scene. By enabling the user to explore a scene and discover various ways of interaction with a virtual object, this object can become more aesthetically pleasing. It also seems to be a good guideline to find markers that are not placed against a wall to display other virtual content then augmented images and portals or windows to another scene when using marker-based tracking.

As mentioned before, the markers were too easy to spot and the exploration effect, that was very appreciated by most participants, got lost in that part of the study. Choosing existing images that are already part of the scenery and letting the user find those with the help of GPS would have been an alternative. Unfortunately, no such images could be found in a safe pedestrian area close to Tegelbacken. Then again, the study shows that the exploration is an important part of a good AR experience and can make a point to take this into account in future research.

Furthermore, it could have been beneficial to apply the improvements for location-based tracking that Lindström [13] suggested. I implemented only one of the suggested improvements: an arrow to indicate the direction in which the virtual object is placed to help the user find the virtual objects. This seemed essential in order to make it possible for the user to find the virtual objects in an appropriate amount of time. Not making use of this improvement would have only added an unnecessary level of complexity to the task. Finding the markers in a small area was still a lot easier for the users than finding the virtual objects’ geospatial locations with the help of the arrow. The other suggested improvements were the possibility of manually adjusting the altitude of the virtual object and resetting the sensors by shaking the phone. It was decided against using these functions because they were expected to take the focus away from the tasks the user should perform. The task should be as similar as possible for both tracking technologies in order to compare them properly and gain meaningful results. Even though those additions added usability to the location-based tracking, those tasks should be performed by the device and not by the user since this leads away from the original goal of AR: to be unable to distinguish between virtual and real world content. Those improvements were not implemented to get an unspoiled impression of how close the tracking technologies are on that way.

However, a hybrid tracking technology could have been implemented to show the potential of the combined methods. Unfortunately, there was not enough time to develop a feasible prototype that uses this technology. After conducting the study the hybrid approach seems even more like a promising alternative to both location-based and marker-
The study only tested close-by, rather small virtual objects. Tall buildings, that are located further away might have been easier to find and the erratic behaviour might have been less apparent. But since the marker-based tracking technology is meant to rather display objects that are close-by and location-based tracking might place a big object right where the user is standing, this option has been dismissed in favour of better comparability.

7. FURTHER RESEARCH

This study raised the question about how suitable marker-based and location-based tracking methods are for a city walking tour. The results give occasion to question whether location-based tracking is ready to be implemented for a mobile city walking tours. Since the marker-based tracking prototype seemed to be more widely accepted by the participants of the study, the study provides motivation to research the possibility of using markers for a city walking tour. It is easy to imagine that it can be taken advantage of the already existing public spaces. It seems worth to look into the possibility of using, for instance, bus stops as points of interests. There the map of the public transport net - which can be found at almost any bus stop - could act as a marker. In combination with location-based tracking, virtual content that is relevant to the specific location, can be displayed. A hybrid solution could be a good choice and make use of the strengths of both location-based and marker-based tracking. Further research is necessary to determine a good design of a suitable hybrid tracking technology.

Another way of conducting further research would be to find automatic substitutes for the improvements for location-based tracking suggested by [13] and compare the thus improved tracking technology to marker-based tracking. For instance, I noticed progress in the development of alternative sensor resetting methods. PeakVisor \(^4\) is a mobile AR application that acts as a mountain guide for hiking trips. It uses time and the detected position of the sun to calibrate the sensors. The success of this method is not validated, but it seems to be a promising approach.

8. CONCLUSION

The study could show weaknesses and strengths of location-based and image-based tracking for outdoor walking applications by comparing both technologies. Both methods show critical flaws that have to be considered when designing an AR application for outdoors. While the inaccuracy of location-based tracking exacerbates the interaction with the virtual content significantly, marker-based tracking lacks the freedom of placement of the virtual objects. Marker-based tracking depends strongly on the potential marker locations of the environment. Not every marker can be placed in a location that allows to accurately position the virtual content in a way that it fits into the physical environment. However, marker-based tracking shows an advantage in stability of object tracking. Since it displays the objects more stable than location-based tracking, the interaction and the viewing of the the virtual object from different angles is easier to accomplish. The study results are able to provide guidelines of content design and choice of tracking method, but they also indicate that further research is necessary to find a more stable way of presenting virtual AR content. By pointing out the strengths of both tracking technologies, the groundwork for the design of a hybrid method has been created. The inaccuracy of the GPS sensor in today’s mobile phones is not easy to overcome. This study could show the need to think of novel, possibly hybrid, approaches to achieve a pleasant AR experience on mobile devices. As long as mature sensors like a depth sensor are not part of the majority of mobile phones, out of the box thinking is necessary when creating AR applications. Nevertheless, mobile phones should not be completely dismissed as platforms for AR as the accessibility of a new technology is important for its development. The positive reactions of the participants during the study indicate that there is potential to use AR technology successfully for an outdoor walking city application for mobile phones. The study findings can help to guide the design process and achieve a better AR experience.

9. REFERENCES


\(^4\)https://peakvisor.com/


