

Location Based Services for Human Self-Localization

Gerhard Navratil*, Manuel Schmitzer, Ioannis Giannopoulos

Research Group Geoinformation, Vienna University of Technology, Austria, [firstname.lastname]@geo.tuwien.ac.at

* Corresponding Author

Abstract: Human self-localisation is an important part of everyday life. In order to determine one's own position and orientation, the allocentric representation, usually in the form of a map, has to be aligned with one's own egocentric representation of the real world. This requires objects (anchor points) that are present in both representations. We present two novel approaches that aim to simplify the process of alignment and thus the self-localisation. The Viewshed approach is based on visibility analysis and the Image Recognition approach identifies objects and highlights them on the map. On the basis of an empirical experiment with 30 participants in the city of Vienna, Austria, the two approaches were compared with each other as well as with a standard approach using a 2D map representation. The goal is to assess and compare aspects like efficiency, user experience, and cognitive workload. Results show that the Image Recognition method provided the best support and was also most popular among users. The Viewshed method performed well below expectations.

Keywords: self-localization, viewshed, image recognition, user tests

1. Introduction

Wayfinding is an important part of everyday life. It describes the process necessary to decide on a route from one's own location to the destination and then follow it (Gollege, 1999, p. 6). When using maps it is necessary to determine one's own position and orientation both in the real world and on the map, e.g., after exiting public transportation. This determination is also called self-localization (Kiefer et al., 2014). Subsequent wayfinding steps will require external assistance if self-localization is not carried out correctly. Meilinger et al. (2007) describe self-localization as the process that takes place when one is disoriented. Disorientation means absence of knowledge of where one is in relation to one's own inner representation of the environment or in relation to an existing map. In order to regain orientation, objects of reality are compared with the map content. New technologies offer a multitude of possibilities to support users in self-localization. This raises the question of suitable methods in terms of simplicity and efficiency.

Depending on the situation, and the complexity of the environment (Giannopoulos et al., 2014b), various strategies can be applied in self-localization or orientation. For example, a verbal approach can be used in which the situation is described by linguistic means. Another method is the mental rotation. Users of a map deal with two representations. The map itself represents a fixed representation of the spatial environment. It is called an allocentric representation and is independent of the user's perspective. The perspective of the user, on the other hand, is an egocentric representation, which is flexible and describes reality as the user observes it (Klatzky, 1998). In order to carry out self-localization, it is necessary to harmonize these two representations. They are cognitively transformed until they agree. This transformation is called mental rotation (Gunzelmann et al., 2006). During mental rotation, distance and relative position of the objects in relation to the users play a

major role (Iachini and Logie, 2003; Gunzelmann and Anderson, 2004; Klippel and Winter, 2006). Gunzelmann and Anderson (2004) showed that the error rate and the time required to perform the orientation task increases with the angle between the orientations of egocentric and allocentric representation increases. The number of visible objects also affects the time required. If more objects are present, the situation is more complex and leads to a slowing down of the orientation process. Also complex terrain or large distances can distort the egocentric perspective. However, the examples picked in this paper avoided these problems.

One way to simplify self-localization for users is to support them during mental rotation. Objects can be used as anchor points in this process. Easily recognizable objects, also called landmarks, are preferential. Landmarks stand out from their surroundings due to certain properties (Raubal and Winter, 2002). If a landmark is visible in both representations, i.e. in the map and in the egocentric representation, it can serve as an obvious anchor point and thus significantly facilitate self-localization.

Peebles et al. (2007) discuss cognitive factors that influence the performance of a drop-off orientation task (the users are in an unknown environment and have no prior spatial information about the situation). In this case, it is important how landmarks used for the orientation are represented on the map. For example, an object can be a landmark in reality due to its height or roof shape, but not in a planimetric 2D map. Their experiments showed that the use of 2D landmarks is much more reliable, but people still often prefer 3D landmarks for the orientation task. Landmarks can have both a positive and a negative effect on self-localization. A suitable landmark can provide a simple anchor point for mental rotation and greatly simplify the orientation task. On the other hand, unsuitable landmarks that are only present in one of the two representations can act as a distraction (Kiefer et al., 2014). They slow down users or can even lead them to false assump-

tions. The choice of suitable landmarks and anchor points thus plays an important role for human self-localization.

In the context of this work two new approaches are presented which should support mental rotation and thus the self-localization: a Viewshed and an Image Recognition approach. A viewshed analysis splits the space into visible and non-visible areas. This information is used to identify visible landmarks and help the user to choose suitable anchor points. The second approach is based on image recognition. The position of a building captured by the camera is marked on the map after recognizing it. This determines the orientation and offers the users an anchor point. Both approaches were implemented as a prototype in a mobile application. The Basic method corresponding to the classical use of a digital map was implemented for comparison. In an experiment with 30 participants in a real environment, the three methods were tested on efficiency, cognitive load, and user experience.

2. Supporting Human Self-Localization

Human self-localization must align the egocentric and the allocentric representation. Successful completion of this step is a precondition for the determination of the user's position and orientation. At the moment, the most common method to support this task is the visualization of the approximated user position. The position is captured as part of the context and affects the visualization on the map. Since this is a widespread approach, it is considered standard and is therefore used in all three methods discussed below (Basic, Viewshed, and Image Recognition).

2.1 Basic Method

The Basic method provides the basic functions of a common mobile map service and extends them by the functions necessary for the experiment. It is used to compare the two new approaches with the current standard of mobile map services. Google Maps was chosen as a platform because it is the best known and most widely used online map service and the Google Maps API already offers most of the functionality necessary for the experiment. Figure 1a shows the visualization of this approach.

2.2 Viewshed Method

The Viewshed method identifies the buildings visible from the user's location. These buildings are highlighted on the map. The concept of viewshed analysis has been used for several years for analyzing the complexity of wayfinding. To date, it has mainly been used to predict the difficulty of the wayfinding task from the geometric configuration of the area (Davies et al., 2006; Peebles et al., 2007; Davies and Peebles, 2010; Giannopoulos et al., 2014a). In the approach presented here, however, viewshed analysis is used to support users during self-localisation.

The classical Viewshed approach is extended slightly for the experiment. The classical approach classifies all points as visible or invisible. This marks the surrounding street areas and facades as visible. This result is correct but may be difficult to interpret for non experts. Thus not the points but the buildings are marked as visible even if this is only true for a small portion of the facade. This allows users to immediately recognize which buildings represent anchor points and are therefore suitable for self-localization. Figure 1b shows the visibility of buildings in one of the test locations.

2.3 Image Recognition Method

The Image Recognition method does not primarily aim at a cognitive facilitation of the task. Rather, the orientation determination is to be automated. Orientation is easiest to be defined by two points, a starting point and a destination point. Assuming that the first step of self-localization is the determination of one's own position, only the destination point needs to be determined. Image recognition is used for this step. The observed object is displayed on the map after recognizing it on a picture taken by the camera of the mobile device, similar to Google's Visual Positioning Service (Reinhardt, 2019). In combination with the user's position, the orientation is determined. The user now only has to process the displayed information cognitively. This approach is similar to the GAIN-LBS presented in (Anagnostopoulos et al., 2017), where the direction of the user's gaze is determined by eye-tracking.

In the form used here, the first task is to determine the target point by image recognition. Then the user's position is determined. The approximate position, determined by a positioning service, is used to visualize an approximated orientation. The user knows on which object she has pointed the camera, thus she knows the position of the object in the egocentric representation. In addition, the object is displayed on the map, so that she also knows the position in the allocentric representation. Thus, she automatically has an anchor point for the alignment of the two systems, helping to determine her own position.

The Vuforia SDK¹ for augmented reality was utilized to enable image recognition. It is a local app that compares the images received from the built-in camera to images in a locally stored database. As soon as one of the images is recognized, the information of the shown building is used to help the user orient herself by showing an arrow from the current position to the building (see Figure 1c).

3. Experiment

An experiment was conducted to test the applicability of the three methods and identify strengths and weaknesses. The field tests were done between August 9, 2018 and September 5, 2018 between 9AM and 7PM. Similar weather situations were chosen to exclude this parameter.

3.1 Participants

The sample consisted of 30 participants (18 male). The average age of the participants was 24.5 ± 2.7 years. Information about the participants was collected through questionnaires. The Santa Barbara Sense-of-Direction Scale (Hegarty et al., 2002) was used to determine spatial skills. The average score was 4.7 ± 1.0 . 21 of the participants were holding a university degree and the remaining ones stated a secondary school degree as their highest level of education. All participants stated that they used digital maps at least sporadically. Only one participant indicated familiarity with the test area but she was not familiar with the exact test locations. Thus the data of this person could be used in the evaluation.

¹<https://developer.vuforia.com/downloads/sdk>

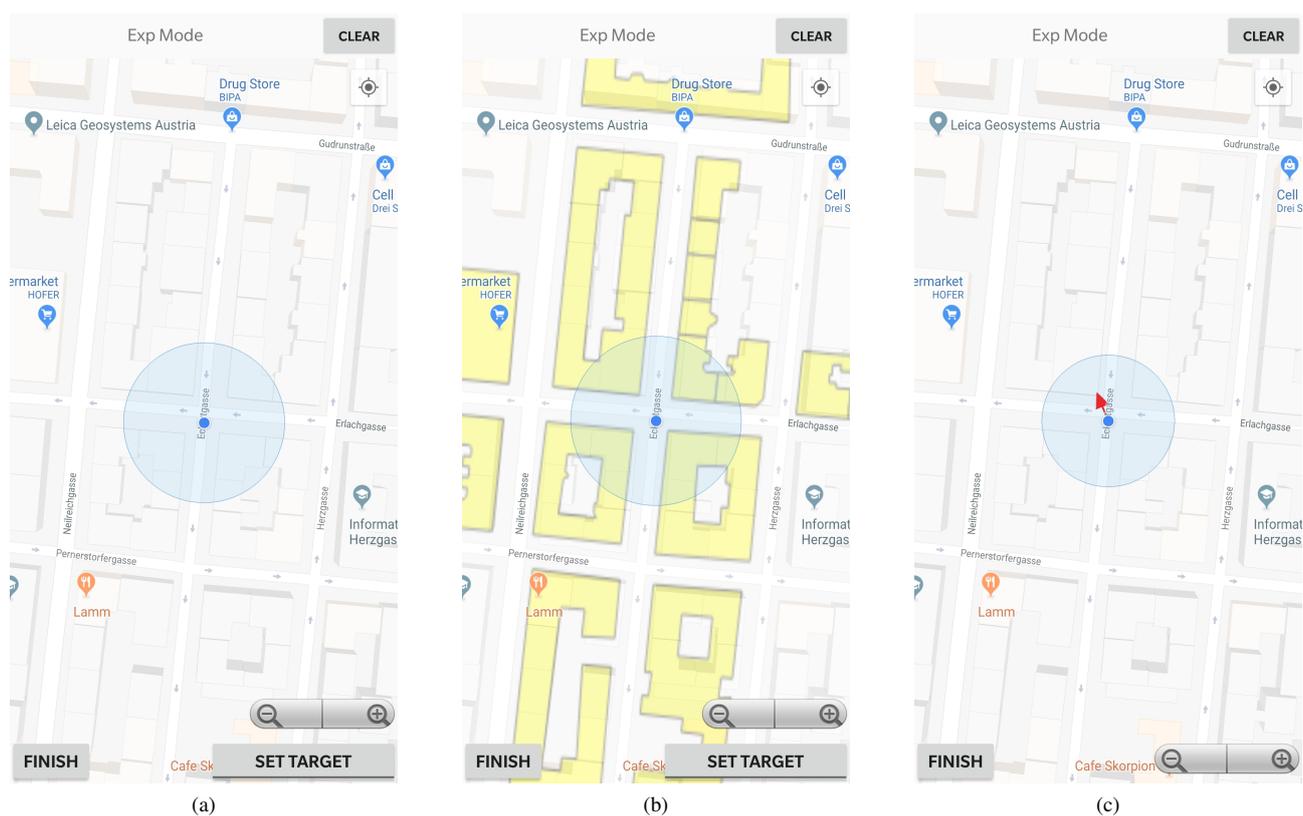


Figure 1. The three approaches: (a) Map view of the Basic method; (b) Map view of the Viewshed method; (c) Map view of the Image Recognition method. Background for all approaches is Google Maps.

3.2 Setup

The empirical experiment was performed in a real environment. All self-localization tasks were performed with the same mobile device, a OnePlus 5T smartphone. The mobile phone was equipped with an application supporting the three different methods. The experiment covered a total of four locations, which are described in more detail below.

3.3 Design

A within-subject design was used to compare the three methods. The first of the four locations served as a training location where all three methods were demonstrated and tested by the participant. The remaining three locations were used as test locations. Each participant had to use each method exactly once in one of the test locations. Since the test locations had different spatial characteristics, counterbalancing was necessary. The order of the methods was permuted such that each method was applied 10 times at each location by 10 different participants. In addition to the three self-localization tasks, each individual experiment also included a test on Judgement of Relative Direction (JRD). This is an orientation task in which the participant imagines standing at a certain point and looking in a given direction. Relative to this orientation, the participant should then indicate the direction to a previously learned object (Shelton and McNamara, 2001; Waller and Hodgson, 2006). The duration of a single experiment was slightly less than one hour and the participants were only tested individually to avoid mutual interference. The performed task was a combination of position determination and determination of spatial orientation,

whereby the participant was supported by one of the three orientation methods. Self-localization requires two locations, the current position and the position of a given target (i.e., the corner of a building). The first two methods automatically set the mark for the current position and the participant only had to add the second location while the Image Recognition method calculated the second location and the participant had to identify his own position. The task was completed as soon as both location were specified and the finish button was pressed. The following data were measured during self-localization:

- the time required to perform the self-localization task;
- the number of clicks during the self-localization task including map clicks and button clicks;
- the position of the starting point input;
- the position of the destination input.

The experiment was designed in such a way that the self-localization at the various locations was based on drop-off experiments. That means that a participant is in an unknown situation and does not have spatial information about the location until the time of the experiment (Thompson et al. 1990, Peebles, Davies and Mora 2007). The participant should therefore arrive at the test site with as little spatial information as possible. Thus, the routes between the different locations played an important role and had to be planned in the design phase. Several measures were taken to limit the information gained during navigation. The participants were asked to look around as little as possible between the self-localization tasks and to focus their attention on the ground. On the other hand, the

tramway was used between the training area and the first test location in order to further restrict the acquisition of knowledge. Originally, it was planned to use at least one tramway stop between the test locations. During the pilot experiments, however, it turned out that the required time had a negative influence on the motivation of the participants. Therefore, the design for the final experiment avoided the use of tramway to keep the duration of the experiment below one hour. However, each route had a small detour with at least two changes of direction and the participants never approached the JRD target directly when arriving at a location.

3.4 Test Areas

All locations are situated in one district of Vienna separated by a walking distance of 10 minutes. The distance was necessary in order to avoid any influence between the individual self-localisation tasks. Since most of the participants live in or near Vienna, prominent landmarks should not be visible from the test locations. The presence of suitable targets for Image Recognition was also important in the choice of locations. Finally, the locations should also cover various types of spatial configurations. The goal was to achieve the best possible external validity. The three test locations have different spatial characteristics, and this should have an effect on the self-localization task.

The training location was also used as a meeting point and should be easy to reach. Since it was used for explanations and training only, minimizing pedestrian or car traffic was not required. The exit of the station of an underground line was selected.

The first test location is in a narrow street. The target building has a U-shape with the opening to the position of the participant. Thus the building could be recognized as a landmark and serve as an anchor point for the self-localization. Furthermore, the position of the participant at this location was not directly at the intersection, but in one of the streets. This makes the geometry of the target building more recognizable and influences the results of the Viewshed method, which marked significantly more houses on the east-west axis than on the north-south axis. This provides a simple first indication of the spatial orientation. A bridge between two buildings across the road was used as the JRD target.

The second test location was in a narrow street without outstanding buildings. Therefore, there was no obvious anchor point. The only available landmark was a retail store at a nearby intersection. This can be identified by the logo at the entrance and was also represented on the map. The store also served as a JRD destination.

The final test location was situated next to a park, which served as a landmark and anchor point. It was also the JRD target for Test Site C.

3.5 Procedure

Each experiment started at the training location. The participant was greeted and provided information in written form. This included a short thematic introduction, an overview of the parameters measured during the experiment, the general procedure of the experiment, and information on data protection and the handling of personal data. Subsequently,

the participant had to fill in questionnaires concerning personal information as well as a self-assessment of spatial abilities. The Santa Barbara Sense of Direction Scale, SB-SODS (Hegarty et al., 2002) was used to assess the participant's spatial abilities. The demographics questionnaire included questions concerning age, gender, highest level of education, color weaknesses, local knowledge, and experience with digital maps. The participant was then introduced to the mobile application and could test all three methods without a time restriction. After this, the training phase ended and the experiment was started.

The participant was then guided to the three test locations. The supervisor asked the participant to focus on the ground. At each test location, the participant had to perform the self-localization task using the method specified by the supervisor. For this, the participant was guided to a predefined point. The mobile phone was handed over on site with the application already running. Before starting the self-localization task, the required steps were briefly repeated. Then the participant was shown the target object and was asked to execute the self-localisation task. After completion, the participant was given further two questionnaires to complete: the UEQ questionnaire (Schrepp, 2018) to assess the user experience (UX) and the Raw Nasa TLX questionnaire (Hart, 2006) to measure the cognitive workload of the participant. This process of self-localization and subsequent completion of the two questionnaires was carried out at each of the test locations, i.e. three times in total.

After finishing the task at the third test area, the participant was guided out of the area and asked to perform the JRD test for all three test sites. The test was not announced since it was supposed to show how well the participant got to know his environment depending on the method used (the results of the test are omitted in this paper since they were not conclusive). The JRD test concluded the experiment and the participant was guided back to the starting point if necessary.

4. Results

The data obtained during the experiment were statistically evaluated in order to be able to answer the research questions. A Shapiro-Wilk test was first performed on all dependent variables to determine whether the measured values were normally distributed. A parametric test was performed, if this was the case. Otherwise, a non-parametric test was used. If no explicit test is stated, it is the results of a Wilcoxon test. The general significance threshold was assumed to be $\alpha = 0.05$. Strictly speaking, the Bonferroni correction had to be applied, where in the present case the p-value has to be divided by 3. This results in $\alpha = .017$. In the following a signal threshold of $\alpha = .05$ is used, but in borderline cases we refer to the Bonferroni correction. The focus in the following is on tests that produced significant results.

4.1 Completion Time

First, we focused on analyzing the total completion time for each approach. The measured values of 26 of the 30 participants were taken into account. Four samples contained outliers in at least one of the methods and therefore excluded from evaluation. The resulting averages and their

method	Average [s]	N	StdDev. [s]
B	29.4	28	9.6
V	30.4	28	14.9
IR	25.6	30	12.3
Total	28.4	86	12.5

Table 1. Completion time for self-localisation: Base (B), Viewshed (V), and Image recognition (IR) method

standard deviation are shown in Table 1. The results for the base and the Image Recognition method were normally distributed, so a T-test with paired samples could be performed. This showed that the Image Recognition method was significantly faster ($p < .012$; $t(25) = 2.704$). The measured times for the Viewshed method were not normally distributed and thus the Wilcoxon test had to be used for the comparisons with the Viewshed method. No significant difference was found between the base method and the Viewshed method ($p = .770$, $Z = -.292$). A comparison of Viewshed and Image Recognition showed that the Image Recognition method was significantly faster ($p < .043$, $Z = -2.019$). However, considering the more strict Bonferroni correction, this would not be a significant result.

4.2 Error Rates

With regard to error rates, no inferential analyses could be carried out, as the users only had to localise themselves once with each method. Their position entries were either correct or incorrect. This classification was done manually based on semantics. A position on the correct side of the street and on the correct side of a building was classified as correct. The input was classified as incorrect if any of these two conditions were violated. The resulting error rates of the three methods were compared in a descriptive manner. With the Basic method, 23 out of 30 participants localized themselves correctly (error rate of 0.23). The Viewshed method led to 26 correct self-localizations (error rate of 0.13) and using the Image Recognition method, all 30 participants successfully completed the task (error rate of 0.00).

4.3 User Experience

It is not only important whether a new method is efficient, but also whether users find the use of the method positive. Attractiveness describes the overall impression of the users. Perspicuity, efficiency, and dependability denote pragmatic, goal-directed quality aspects. Stimulation and novelty are hedonic, not goal-directed quality aspects. The UEQ questionnaire was used to investigate how well the various methods were received by the participants. Figure 2 shows the mean values of the different UEQ scales for the three methods. The Image Recognition method scored best on all scales except dependability. Compared to the Basic method, the results in the scales efficiency, stimulation and novelty were significant. In comparison with the Viewshed method, the results in the scales attractiveness, efficiency, stimulation, and novelty were significant. The Viewshed method, although significantly better than the Basic method in terms of novelty and stimulation, was inferior on the four other scales. After the Bonferroni correction, stimulation is not a significant result. In order to gain a better understanding of the quality of the methods, the results were also compared with established products. There is a UEQ benchmark for this, which comprises 246 product

evaluations from a total of 9905 test users (Schrepp, 2018). Figure 3 shows how the methods presented here compare with this benchmark. Both new methods were above average in all categories and the Image Recognition method in four of the six scales was even in the range of the best 10%.

5. Discussion

The results show that Image Recognition performed significantly better than the other methods in many important areas. In the Viewshed method, hardly any significant improvements could be found compared to the Basic method. If one considers the time required to perform the self-localization, the participants were significantly faster with the support of the Image Recognition method than with the other two methods. However, no significant difference could be found between the Basic and the Viewshed methods.

It is also interesting to note that the standard deviation of time is by far the highest for the Viewshed method and the lowest for the Basic method. The low standard deviation of the Basic method may result from the high degree of familiarity with digital maps. All participants had experience with digital maps and thus knew how to apply the Basic method. The larger standard deviations of the other two methods suggest that not all participants were able to exploit the potential of these methods. This was confirmed by verbal feedback from the participants as some participants expressed that they did not know how to use the information presented by the system. This assumption was consistent with the UEQ results, where the Viewshed method scored worst on four of the six scales.

A between-subjects analysis concerning the three test areas only produced significant results for the first location between the Basic and Image Recognition method. At the second location, the Image Recognition method was comparable to the Basic method and at the last location it even scored the worst. Of course, it must be noted that there were only a maximum of ten measured values per method and per location. With such a small sample size, not too much should be interpreted. However, it is peculiar that the Image Recognition method performed well at the first location and fell behind the other methods at the third location. A comparison of location complexity before the test led to the assumption that the second location would be most complicated since no distinct feature (house shape or park) was available. However, verbal feedback from the participants showed that the retail store was used several times as an anchor point, whereas the geometry of the building at the first location was not used (or at least not mentioned) by anyone. Therefore, in retrospect, it is assumed that the first location was the most difficult for self-localization. This gives reason to assume that with the increasing difficulty of the situation for self-localization, the Image Recognition method is gaining in importance compared to the other methods. This method provides a reliable anchor regardless of the complexity of the situation.

The results on error rates require some discussion since a certain pattern emerges. Image Recognition clearly scores best with the Viewshed method as second best and the Basic method trailing. However, without inferential statistics, conclusions need to be viewed with caution. Interestingly,

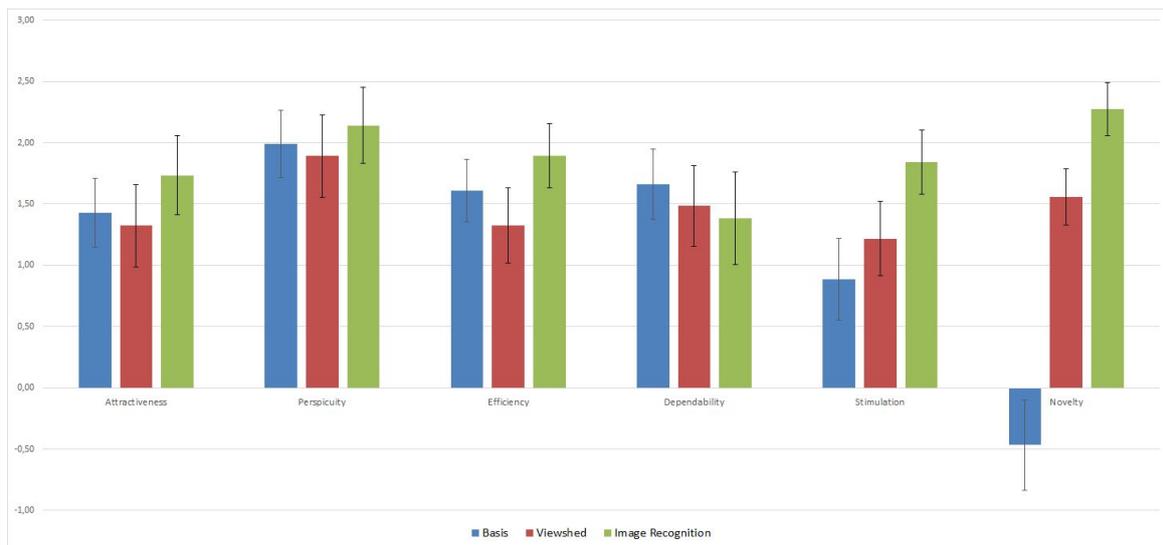


Figure 2. Average of UEQ analysis: Attractiveness, perspicuity, efficiency, dependability, stimulation, novelty for Basis (blue), Viewshed (red), and Image Recognition (green)

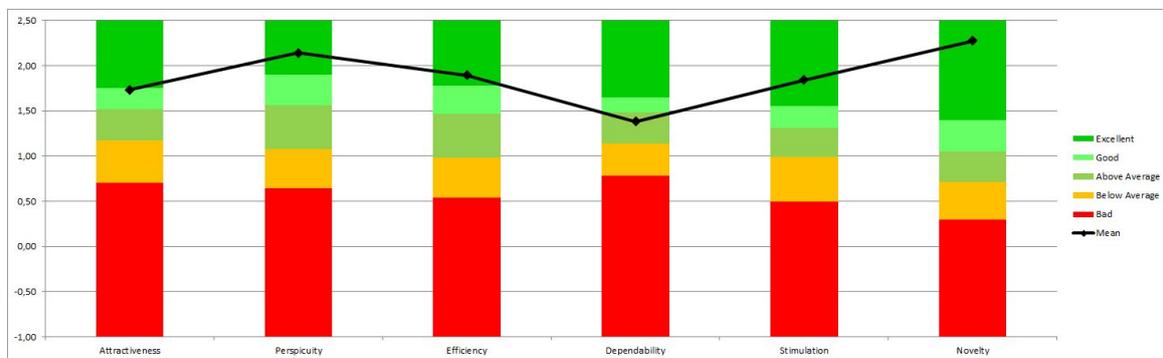


Figure 3. Comparison with the UEQ benchmarks: Basis (diamond), Viewshed (circle), and Image Recognition (square)

only two of the individual tests with a wrong result had one of the positions set incorrectly. In all other cases, either the self-localization was performed correctly or both positions were wrong. In most cases it was a wrong determination of the participant's own position. This would explain why the Image Recognition method, where one of the points is automatically determined, performed best. Furthermore, it can be deduced that the two positions were not determined independently from each other, but that the determination of the second position probably happened relative to the first one. This in turn confirmed that the benefit of the Image Recognition method is not only the determination of the destination, but also the influence of the method on the determination of one's own position. Finally, the Basic method may have suffered from problems in interpreting the map. Several positional entries showed that it was apparently not clear to all test participants how the buildings were represented on the map. In two cases, the sidewalk was confused with the building, resulting in a wrong user position. The Viewshed method, based on marking buildings, makes the separation of these objects much simpler for the users. The result of the Viewshed method could thus be based on better map readability.

In order to get an understanding of the necessary effort and the cognitive load of the users, the number of clicks in the mobile application and the Raw Nasa TLX data were

evaluated. Interestingly, a significant result in favor of the Viewshed method was found for the number of clicks compared to the map-based approach. No significant difference was expected for the number of clicks since the methods share the same user interface. The results might lead to the assumption that the users felt more secure using the Viewshed method and changed the set location less frequently. However, there is hardly any evidence for this in the results of the Nasa TLX. A more likely explanation lies in the additional data shown during the Viewshed method and the associated improved map readability.

The effort associated with the task can also be understood as part of the stress to which the participant was exposed. So it makes sense to include the TLX results into the analysis of the effort of self-localization. Image Recognition is significantly better than the other methods if considering the total load, i.e. the sum of the six TLX scales. This shows that the participants found it easier to use Image Recognition and were less stressed. However, the individual scales should also be considered as there was no significant result for mental demand. Especially here a significant difference would be reasonable, because the determination of the target position is provided by the application and the user gets an additional piece of information for self localization. This should lower the mental demand. A possible explanation why this is not the case would be that Google

Maps is a widespread product and the users are trained in dealing with it and thus with the Basic method. The Image Recognition method differs quite clearly from this process and the users have to adapt to this new approach. The training phase at the beginning of the experiment might have reduced the effect, but could not eliminate it completely.

The UEQ provides information on the user experience. The two new methods, and in four of the six scales also the Basic method, performed above average in the UEQ benchmark. This is a good indicator that the implemented user interface, which was nearly identical for the three methods, was well received by the participants. According to Schrepp (2018), attractiveness is a valence dimension that reflects the value of the product. Perspicuity, efficiency, and dependability indicate the pragmatic quality, i.e. how well the product supports the user in fulfilling the task. Stimulation and novelty are hedonic quality measures that are not goal-oriented. The Basic method clearly has disadvantages in the hedonic scales. This confirms the assumption that self-localization with Google Maps is widely used. It is therefore reasonable that the Basic method did not perform well on these two scales in relation to the benchmark and in comparison to the other two methods.

The Basic and the Viewshed method performed similarly in attractiveness, possibly due to the comparable concept. The Image Recognition method was best received by the participants. It stands out from the other methods and is even better than the Viewshed method. Efficiency is the sole pragmatic scale that shows significant differences. This matches the results of the required task time, where Image Recognition performed best. Dependability is the only scale where the Image Recognition method was not the best. This could be due to the situations where the image recognition process took several seconds making the participants feel out of control (Kiefer et al., 2017).

When testing for correlations, it was unexpected that none of the tested dependent variables correlated with users' SBSODS values. Actually, it can be assumed that people with a high SBSODS value have better spatial skills and should therefore perform better in the areas of positioning, time required, and knowledge growth. However, this was not the case. The opposite happened for the correlation with age. No significant correlation was expected here because there were only small age differences between the participants. However, it was shown that with increasing age the self-localization tasks took significantly longer while the error rate dropped. Apparently, older participants tended to look around longer during the experiment and thus produced a better mental map.

The decision to carry out the experiment in a real environment and not in a laboratory plays a decisive role. It is a compromise between internal and external validity. A laboratory provides a controlled environment that can be tailored to the experiment and made as uniform as possible for all participants. This reduces unwanted influences. The disadvantage, however, is the external validity. Tests in real environments tend to produce worse results and participants show a more negative attitude (Duh et al., 2006).

Since a real environment was chosen for the experiment, the external validity could be treated well, but it also led to some limitations. Less data could be collected than would

have been possible in a laboratory experiment. The experiment required that a certain distance between the test locations to achieve independence. This limitation is important for the error rates of self-localization and for the JRD tests. Evaluations were only possible to a certain extent due to the limited number of data points. Furthermore, the internal validity is limited. In a within subject design, only the tested influencing factor should change. However, this is not possible in a real environment. This problem was restricted by instructions, but not eliminated.

The results of the Image Recognition method agree with the assumptions in almost all areas. However, this is not the case with the Viewshed method. In most of the areas tested, it fell well short of expectations and could not stand out significantly from the Basic method. However, this does not necessarily disqualify the theoretical approach, but may have other reasons. The verbal feedback of the participants indicates that some of them did not know what to do with the additional information. A possible explanation would be that the test phase was unsuitable or too short. The test phase was mainly about the correct use of the application. The participants learned how to use all methods, but only performed the actual self-localization with the Basic method. The supporting function of the Viewshed method and how to make use of the additional information had to be found out during the real test. All participants stated to have experience with digital maps. While the use of the information by the Image Recognition method is obvious, the use of the information from the Viewshed method requires different cognitive efforts and may differ the most from the training structure. It may be that most participants did not try to use the additional information, but relied on the basic information from Google Maps. Thus, the Viewshed method was the same as the Basic method, only with better map readability. This assumption fits quite well to the results of the different evaluations.

6. Conclusion and Outlook

Two novel approaches to support human self-localization were presented. Prior studies have shown that the choice of suitable anchor points plays a decisive role for successful self-localization. Users are often distracted by striking objects of reality, although they are not shown on the map and are therefore no suitable anchor points. The presented methods are intended to avoid this problem. With the Viewshed method, the buildings visible from the user location are highlighted in the representation. All other buildings can be excluded as potential anchor points, which significantly reduces the search area and increases the probability of using a suitable object. The Image Recognition method performs a part of the task for the user. The position of the building and an approximated orientation are shown on the map after pointing the camera on it. The information gained in this way also helps determine one's own position showing a clear anchor point.

The two novel methods were compared with each other and also with the Basic method in an experiment with 30 participants in a real environment. The Viewshed method fell well short of expectations. Based on the achieved results, an implementation suitable for everyday use is probably not worthwhile. The method only achieved good results for the number of clicks required and the error rate. Both

are probably based on improved map readability. However, the design of the experiment and the training phase was not optimal for the Viewshed method. This method is the most demanding of the three and it requires the most practice in order to get familiar with. The Image Recognition method, in contrast, showed its suitability for supporting self-localization. The required task time was significantly lower than with the other two methods and the required effort was significantly lower both in terms of clicks and cognitive load. The error rate of self-localization was by far the lowest and the results were also largely outstanding in the area of user experience. The influence of the selected environment on the performance of the approaches has to be addressed in further experiments.

The poor performance of the Viewshed method in this experiment should not eliminate the approach from further considerations. The setup could have prevented that the advantages of the method could be exploited and further investigation applying a different experimental design seem reasonable. The concept of how the Viewshed result can be adapted to facilitate the understanding of the representation is also an interesting research question for future work. The suitability of image recognition for human self-localization, leads to the question of an implementation suitable for everyday use. This requires an extensive database of reference images. An obvious approach to create this would be to use Google Street View images or images from Volunteered Geographic Information platforms. One question is, whether the Image Recognition method is useful for touristic features or if they are used as anchor points anyway. A future study could compare the effect of the Image Recognition method with the Basic method between situations with and without touristic features.

Other methods would be to use the compass or eye tracking. Google has been experimenting with the compass for some time now and displays the approximated orientation. If the compass works well, it provides a simple support for self-localization. The use of eye-tracking technology could also be helpful. Image recognition via camera could be replaced by gaze-based interaction. A comparison of the two interaction methods would be interesting. Eye-tracking can also be used for analysis. The cognitive processes play an important role in wayfinding tasks and should be considered. In this paper, the cognitive load was measured using Raw Nasa TLX. However, a more precise analysis using eye-tracking could provide valuable insights.

References

- Anagnostopoulos, V., Havlena, M., Kiefer, P., Giannopoulos, I., Schindler, K. and Raubal, M., 2017. Gaze-Informed location-based services. *Intern. Journal of Geographical Information Science* 31(9), pp. 1770–1797.
- Davies, C. and Peebles, D., 2010. Spaces or Scenes: Map-based Orientation in Urban Environments. *Spatial Cognition & Computation* 10(2-3), pp. 135–156.
- Davies, C., Mora, R. and Peebles, D., 2006. Isovists for Orientation: Can Space Syntax Help us Predict Directional Confusion? In: *Proceedings of the Space Syntax and Spatial Cognition Workshop*, Bremen, pp. 81–92.
- Duh, H. B.-L., Tan, G. C. B. and Chen, V. H.-H., 2006. Proc. of the 8th Conference on Human-Computer Interaction with Mobile Devices and Services. ACM, pp. 181–186.
- Giannopoulos, I., Kiefer, P. and Raubal, M., 2014a. GeoGazemarks: Providing Gaze History for the Orientation on Small Display Maps. In: *Proceedings of the 14th ACM International Conference on Multimodal Interaction*, ACM, New York, pp. 165–172.
- Giannopoulos, I., Kiefer, P., Raubal, M., Richter, K.-F. and Thrash, T., 2014b. Wayfinding decision situations: A conceptual model and evaluation. In: *International Conference on Geographic Information Science*, Springer, pp. 221–234.
- Gollege, R. G., 1999. *Wayfinding Behavior: Cognitive Mapping and other Spatial Processes*. JHU Press.
- Gunzelmann, G. and Anderson, J. R., 2004. Location matters: Why target location impacts performance in orientation tasks. *Memory and Cognition* 34(1), pp. 41–59.
- Gunzelmann, G., Anderson, J. R. and Douglass, S., 2006. Orientation Tasks with Multiple Views of Space: Strategies and Performance. *Spatial Cognition & Computation* 4(3), pp. 207–253.
- Hart, S. G., 2006. Nasa-Task Load Index (NASA-TLX); 20 Years Later. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 50(9), pp. 904–908.
- Hegarty, M., Richardson, A. E., Montello, D. R., Lovelace, K. and Subbiah, I., 2002. Development of a Self-report Measure of Environmental Spatial Ability. *Intelligence* 30(5), pp. 425–447.
- Iachini, T. and Logie, R. H., 2003. The role of perspective in locating position in a real-world, unfamiliar environment. *Applied Cogn. Psychology* 17(6), pp. 715–732.
- Kiefer, P., Giannopoulos, I. and Raubal, M., 2014. Where Am I? Investigating Map Matching During Self-Localization With Mobile Eye Tracking in an Urban Environment. *Transactions in GIS* 18(5), pp. 660–686.
- Kiefer, P., Giannopoulos, I., Athanasios Anagnostopoulos, V., Schöning, J. and Raubal, M., 2017. Controllability matters: The user experience of adaptive maps. *Geoinformatica* 21(3), pp. 619–641.
- Klatzky, R. L., 1998. *Allocentric and Egocentric Spatial Representations: Definitions, Distinctions, and Interconnections*. Springer, Berlin, pp. 117–131.
- Klippel, Alexander, F.-C. and Winter, S., 2006. You-are-here maps in emergencies - the danger of getting lost. *Journal of Spatial Science* 51(1), pp. 117–131.
- Meilinger, T., Hölscher, T., Büchner, S. J. and Brösamle, M., 2007. *How Much Information Do You Need? Schematic Maps in Wayfinding and Self Localisation*. Springer, Berlin, Heidelberg, pp. 380–400.
- Peebles, D., Davies, C. and Mora, R., 2007. Effects of Geometry, Landmarks and Orientation Strategies in the ‘Drop-Off’ Orientation Task. In: *Proceedings of the 8th International Conference on Spatial Information Theory*, COSIT’07, Springer-Verlag, Berlin, Heidelberg, pp. 390–405.
- Raubal, M. and Winter, S., 2002. Enriching Wayfinding Instructions with Local Landmarks. In: *Geographic Information Science*, GIScience’02, Springer-Verlag, Berlin, Heidelberg, pp. 243–259.
- Reinhardt, T., 2019. Using Global Localization to Improve Navigation. *Google AI Blog*.
- Schrepp, M., 2018. *User Experience Questionnaire Handbook*.
- Shelton, A. L. and McNamara, T. P., 2001. Systems of Spatial Reference in Human Memory. *Cognitive Psychology* 43(4), pp. 274–310.
- Waller, D. and Hodgson, E., 2006. Transient and Enduring Spatial Representations Under Disorientation and Self-rotation. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 32(4), pp. 867.