Geographic-Aware Augmented Reality for VGI

Paolo Fogliaroni[a, b]∗, Bartosz Mazurkiewicz[a], Markus Kattenbeck[a], Ioannis Giannopoulos[a]

[a] *Research Group Geoinformation, Vienna University of Technology, Austria, [firstname.lastname]@geo.tuwien.ac.at*
[b] *ESRI R&D Center Vienna, Austria, pfogliaroni@esri.com*

* Corresponding Author

Abstract: Volunteered Geographic Information (VGI) has been a constantly growing field over the last decade, but the utilised technologies (i.e., mobile phones) are not able to exploit the full potential concerning effort and accuracy of registering geographic data. This paper introduces the *GeoAR Glasses*, a novel technology enabling the use of Geographic-Aware Augmented Reality for Mobile Geographic Information Systems (Mobile GIS) and Location-Based Services (LBS). The potentials of the *GeoAR Glasses* with respect to current mobile mapping applications is shown by means of an in-situ study (N=42) comparing two different modes of collecting VGI data. For the comparison we take into account the accuracy of the mapped data points and the time needed to complete the mapping. The results show that the *GeoAR Glasses* outperform the mobile application concerning both positional accuracy and completion time.

Keywords: volunteered geographic information, augmented reality, geoAR, in-situ study

1. Introduction

Recently, several studies have indicated that volunteered geographic information (VGI, Goodchild (2007) is of reasonable accuracy (Fan et al., 2014, Neis et al., 2012, Haklay, 2010). Those studies, however, do not analyse the correlation between accuracy level and the approach used to contribute the data. In particular, the positional accuracy of the contributed features might be poor in those cases in which the mapping is done in-situ using a mobile GIS application and, at the same time, the features to be mapped are located in an area which is not physically accessible to the mapper. The mapping of natural features in urban flowerbeds is one substantial example for such a scenario in which objects (i.e. trees, bushes, manhole covers, lampposts etc.) may not be reachable physically because it is prohibited to enter the area they are located in—e.g., because entering the flowerbed might damage vegetation. In these situations, using smartphone applications, may lead to inaccurate results because users are required to estimate the position of objects in relation to (*i*) their own location, (*ii*) the area in which the objects are located in and, possibly, (*iii*) other features than the objects of interest which might also be located in the same area.

In this paper we introduce the Geographic-aware Augmented Reality Glasses (*GeoAR Glasses*) which are suitable to overcome these issues. The *GeoAR Glasses* are an AR headset technology designed to work in outdoor environments. They allow for both, handling geographic data (i.e., providing a holographic representation of geographically referenced data–see Section 3.1) and supporting user interaction with spatial entities (i.e., enabling the user to virtually interact with physical entities–see Section 3.2).

To assess the potential of this novel technology, we set up a between-subjects user study comparing the well-known smartphone mapping approach (condition *smartphone*) to our novel *GeoAR Glasses* mapping approach (condition *GeoAR Glasses*). Participants were asked to map a fixed configuration of ten street cones placed in a delimited area that they were not allowed to enter, thereby enforcing mapping from a distance. We compared two parameters: The accuracy of mapped locations and the time needed to complete the task. The results show that the *GeoAR Glasses* outperform the *smartphone* mapping approach concerning both positional accuracy and completion time. While, in this work, we present the novel *GeoAR Glasses* in a context of point data collection, it has to be noted that the technology lends itself to many other applications. For example, it could be used as a mobile GIS by surveying companies, or be employed in a tourism context to display geographically referenced information about historical monuments.

2. Related work

2.1 VGI Quality

Spatial information collected by volunteers is commonly referred to as Volunteered Geographic Information (VGI) and can come in different forms, such as text, images, maps (Goodchild, 2007). OpenStreetMap (OSM)[1] is probably the most popular and most important example of VGI source, with more than 5 million contributors as of mid-2019. Other examples include Wikimapia[2] and Flickr[3], among others.

Generally speaking, there are different quality measures for geographic data: positional, geometric and semantic accuracy, completeness and reputation (Fogliaroni et al., 2018), among others (Senaratne et al., 2017). As the technology presented in this work deals with positional accuracy we will focus on this quality measure. In order to assess the positional accuracy of OSM data, several studies compared it to official datasets. In (Haklay, 2010), OSM data is compared against Ordnance Survey data. The authors conclude that OSM can be of reasonable accuracy as

---

[1] https://www.openstreetmap.org
[2] https://wikimapia.org/
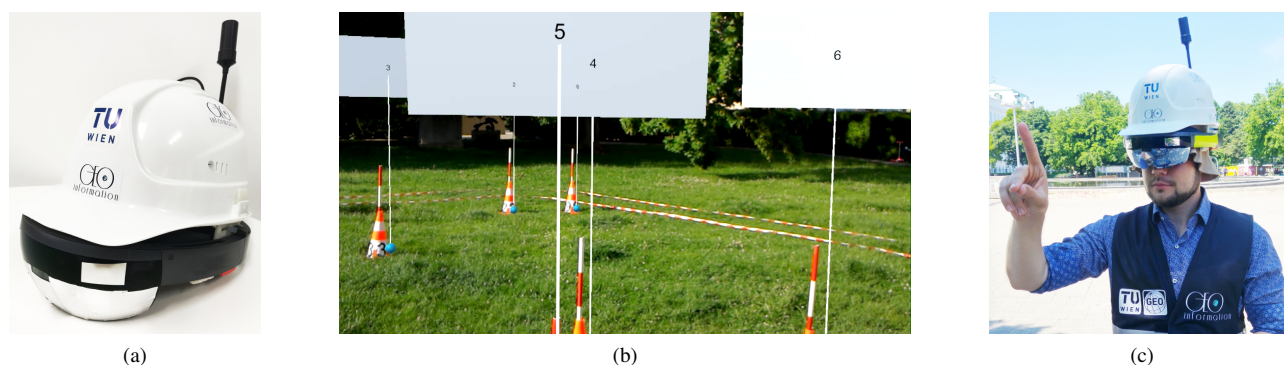[3] https://www.flickr.com/

Figure 1. The GeoAR technology: (a) The *GeoAR Glasses* combine a Microsoft HoloLens and a PPM 10XX GNSS receiver . The front glass of the HoloLens and its environmental cameras have been covered with a light-filtering film to counter the excessive brightness of sunlight; (b) Field view during an experiment; every mapped point feature is marked by a blue sphere; (c) a participant wearing the *GeoAR Glasses*.

they found an average positional error of 6 metres. Similarly, OSM building polygons have been compared to official vector data of Milan Municipality, Italy (Brovelli et al., 2016). The analysis shows a systematic translation of 0.4 metres and a mean positional error of 0.8 metres. These two examples and the work in (Ribeiro and Fonte, 2015) provide evidence that the data quality of OSM varies (e.g. urban vs rural areas). While these studies, among others, indicate positional accuracy issues in OSM data, they do not take into account whether the order of magnitude in positional error is related to the used mapping technique.

2.2 OSM Data Collection Methods

The most common ways of contributing new data to OSM are *armchair mapping*, *bulk imports*, and *in-situ surveys* (Mooney and Minghini, 2017). Armchair mapping refers mainly, but is not limited to, digitising geographic features based on satellite imagery of areas that are frequently not known to the contributing person (OpenStreetMap, 2019b). This approach comes at the cost of neglecting knowledge about the local environment; this knowledge, however, may be highly important to interpret a feature correctly. Additionally, armchair mapping may yield incomplete or inaccurate results in those cases where some features are fully or partially covered by others, e.g. buildings or trees covering features at the ground level.

Bulk imports refer to cases where proprietary data are released for use in OSM and huge sets of features are uploaded in a single batch into the database. One notable example resulting in a large increase of data available in OSM is the 2008 bulk import of the US Census TIGER/Line (Zielstra et al., 2013).

In-situ surveys which can come in different shapes, e.g. on the spot editing, collecting points and tracks via Global Navigation Satellite System (GNSS) devices, photo or video mapping, and field papers (OpenStreetMap, 2018). Postprocessing is required for all but the first of these options. Several smartphone applications enable direct editing of OSM (OpenStreetMap, 2019a) data. Users can either map their own position or place entities anywhere else on the map using a typical point-and-click interaction.

2.3 Outdoor Augmented Reality

Early approaches to outdoor AR rely on GNSS for localisation and inertial and magnetic sensors for orientation and heading. Examples include the Touring Machine (Feiner et al., 1997), the mobile augmented reality system (MARS) presented in (Höllerer et al., 1999), and the work presented by Azuma et al. (Azuma et al., 1999) and Baillot and colleagues (Baillot et al., 2001). The results of these early approaches were promising but outlined two main problems: First, using magnetic sensors yields unreliable results due to electromagnetic distortions; second, the update frequency (typically $1Hz$) of the GNSS receiver is too low to offer a seamless user experience.

Subsequent work has focused on overcoming both issues. The most promising approaches rely on vision-based localisation. For example, a model-based hybrid system was presented which used a 3D model of the environment to improve the tracking of edges based on camera images (Reitmayr and Drummond, 2006). A second example is the work done by Li and colleagues who used AR to visualise underground infrastructure (Li et al., 2018) using a Google Tango Phab 2 (Marder-Eppstein, 2016). This system makes heavy use of area description files (ADF) of the environment that have to be collected in a preliminary offline step in which the Tango device is used to scan the environment. In the online phase the currently scanned environment is compared against the corresponding ADF in the database to compute a seven-parameter coordinate transformation to map the geographic coordinates of the underground infrastructure onto the world space of the Tango device. The virtual objects representing the underground infrastructure are created in the Tango coordinate reference system and the localisation capabilities of the Tango device are leveraged to keep the virtual objects registered to the real environment as the user moves. Our *GeoAR Glasses* applies a similar approach but does not require to preliminary scan the location of interest.

3. Technology and Methodology

3.1 *GeoAR Glasses* Technology

Figure 1a shows our *GeoAR Glasses*, a *Geographic-Aware* Augmented Reality headset. The main hardware components are a Microsoft HoloLens (v1) and a PPM 10XX GNSS receiver. The HoloLens is a mobile AR headset which has been designed for indoor environments. It allows for positioning virtual objects in a local Cartesian reference system

whose origin is located at the physical position occupied by the HoloLens at start-up. Its local coordinate reference system (CRS) is left-handed with x-, y-, and z-axis pointing to the right, upwards, and forward, respectively. The PPM 10XX GNSS receiver is equipped with a GPS+GLONASS L1/L2 Helix antenna and mounts a Trimble MB-TWO board. With optimal satellite coverage, this device provides a positional accuracy of $1.5m$ without correction data, $0.8m$ with EGNOS augmentation, and $0.008m + 1ppm$ with RTK correction data and resolved ambiguities.

We enabled a Microsoft HoloLens to work in geographic space by applying techniques derived from the fields of land surveying and adjustment computations. During a short calibration phase, which is done by the experimenter, we compute a seven-parameter similarity transformation (Ghilani, 2017, Section 18.7) by collecting four calibration points (aka, "pass-points"). The points are collected in both the geographic and the local reference system. For the former we use the RTK-corrected data provided by the PPM 10XX GNSS receiver, with correction data obtained by the EPOSA[4] virtual RTK web service; the latter are provided by the HoloLens. The calibration points are used to compute the transformation parameters directly on the HoloLens according to the formulae given by Ghilani (Ghilani, 2017, Section 18.7). The computed parameters allow for spatial transformations from geographic space into local space and vice-versa.

The *GeoAR Glasses* positional accuracy was assessed during several test phases during which we also carried out accuracy measurements. In optimal conditions for the GNSS antenna, we obtained a mean 2D positional accuracy (i.e., the distance on the ground between an actual point and the corresponding geographic coordinates) ranging between $1cm$ and $5cm$ within the calibration area — i.e., the area delimited by the locations of the pass-points. It is important to note that — since the HoloLens positioning system relies on a vision-based technique — the positional accuracy of the calibration points in the local reference system also depends on the structure of the environment. The larger the number of visually distinguishable features (e.g., an urban environment featuring buildings and other visually distinguishable landmarks) the environment offers, the more accurate the coordinates of the pass-points in the local reference system are and, accordingly, the transformation. In these cases, we reach the best accuracy ($1cm$). If the environment, however, offers only a small number of reference features (e.g., a green area where it is difficult to distinguish between two similar trees), we obtain an accuracy of approx. $5cm$.

Outside the calibration area, the accuracy deteriorates linearly at a rate of approx. $5cm$ to $7cm$ for every $5m$ distance from the boundary of the calibration area. The accuracy deterioration rate depends on the size of the calibration area, i.e., the larger the area, the smaller the deterioration. We performed tests with square-shaped calibration areas with side lengths of $5m$ and $10m$ and performed positional accuracy measurements up to a distance of $50m$ from their centroids. A constant accuracy outside the calibration area can be maintained by applying a continuous calibration process—i.e. by collecting further calibration

points as the user walks away from the initial calibration area.

In case the calibration area is not left or the accuracy outside the calibration area is not crucial for the application at hand, the GNSS antenna is no longer needed once the calibration phase is finished. This is a major advantage of the *GeoAR Glasses*: They can be used even in areas which do not feature an optimal GNSS satellite coverage.

3.2   *GeoAR Glasses* Data Collection Approach

The HoloLens features a cursor that is always located at the centre of the user's field of view. If the line of sight that runs through the cursor does not intersect any object, the cursor floats in front of the user at a fixed distance. Otherwise, the cursor sticks to an object's surface, i.e., it is placed at the intersection between the surface and the ray running through the centre of the user's field of view (see Figure 2a). In our experiment, we used the HoloLens to generate a mesh surface of the experimental area that is needed to let the cursor hit the terrain. Note that, alternatively, the mesh can be generated from available digital terrain models. Taken together, collecting a single point feature required to, first, gazing at the point to be mapped and, then, perform an *air-tap* gesture, as shown in the picture sequence in Figure 2.

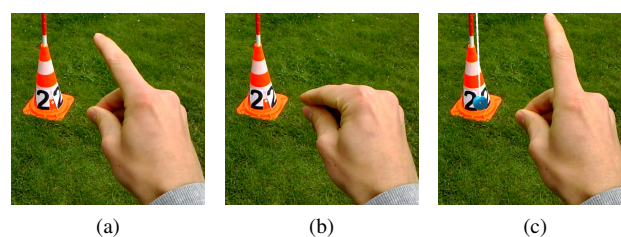(a)   (b)   (c)

Figure 2. Mapping a point using the *GeoAR Glasses* consists of (a) gazing at the point to be mapped and (b, c) perform an air-tap gesture. A cyan sphere of $5cm$ radius appears at the tapped location (c), providing visual feedback of the mapped point.

Once the air-tap gesture was performed, a cyan sphere of radius $5cm$ appeared at the tapped location (see Figure 2c) with a board attached on top of it which displays the number of the mapped point (see Figure 1b). The number was incremented automatically by the application and the board was set $2m$ above the sphere in order to avoid visual clutter. Editing the position of a sphere was possible by setting the cursor on the sphere surface and performing an *air-tap* gesture. In doing so, the sphere enters the edit-mode during which it follows the user's gaze (i.e., the cursor) until the *air-tap* gesture is performed again to confirm the new position.

3.3   Smartphone Data Collection Approach

All control group participants performed the experiment with a OnePlus 5T smartphone (Android 8.1.0) using the application OSMAND (v3.3.8), a smartphone application which implements the established mobile mapping UI. This particular mapping application was chosen for three reasons: First, it comes with export mechanisms for newly added features; second, it does not upload any changes automatically to OSM; third, its user interface is easy to understand and, hence, suitable for beginners.
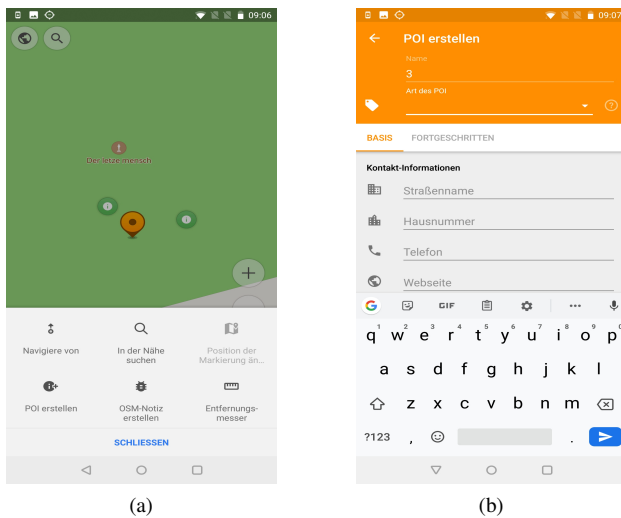
---

[4]http://www.eposa.at

Figure 3. The interface of OSMAND (V3.3.8). (a) Creation of feature 3. (b) Give the feature a name.

As experiments were conducted in German (see section 3.4) the application was used with its German interface (see Figure 3). OSMAND (V3.3.8) provides a classical graphical feedback (blue dot) about position and orientation of the smartphone and participants could pan, zoom and rotate the displayed map. In order to collect a feature, participants were required to, first, perform a long-press on the map at the corresponding location they wanted to create the feature at. They, then, had to click the button *Aktionen* (Actions), then click *POI erstellen* (create POI), provide the feature number as name and then confirm by pressing *enter* twice, in which the second press confirms the pop-up window regarding the blank POI type. A circular marker will then appear on the map at the specified location. In order to edit a previously set marker, users had to tap on the marker to be edited, press *Aktionen* (Actions) and then select *Position der Markierung Ändern* (Change the position of the marker). From this point in time on, the map can be dragged to place the point at a new position. The new location needs then be confirmed by the user by clicking the confirm button.

3.4 Experimental Setup

In order to compare our GeoAR approach to an established way of mapping on smartphones, we asked participants to map ten traffic cones as fast and as accurate as possible in ascending order. The participants were instructed to map the centre of the cone.

Figure 4 shows the spatial layout of the experimental setup used across both experiment conditions; Table 1 reports relative distances between the cones for reproducibility reasons. The order of traffic cones was randomly chosen but fixed to the same sequence for all participants across conditions. The overall experimental area was $10m \times 10m$ in size, whereas the mapping area (represented by the dashed-line polygon in Figure 4) covered $9.1m \times 9.1m$. Participants were allowed to move freely in the area between the drawn-through and the dashed line, which was .9 metres wide, but were neither allowed to enter the area inside the dashed polygon nor to cross the overall experimental area. The geographic locations of the points were randomly chosen as well as the fixed starting point for all
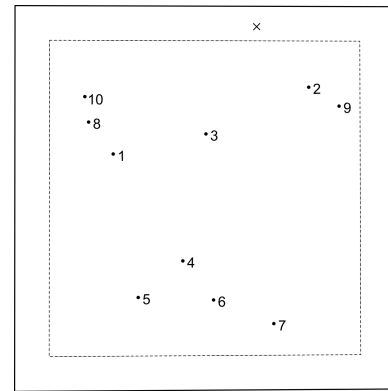

Figure 4. A schematic representation of the spatial layout of the experimental setup. Numbers indicate the randomized order in which all participants were required to map traffic cones. The starting position is denoted by an X. Participants were neither allowed to enter the area inside the dashed polygon nor to cross the overall experimental area (the outer polygon), which was $10m \times 10m$ in size.

trials was (shown as an X in Figure 4). Neither a two-sided Clarke-Evans test ($R = 1.0335$, $p = .8392$) nor a two-sided Conditional Monte Carlo test of complete spatial randomness based on quadrat counts ($\chi^2 = 25.303$, $p = .481$) yielded a significant result for cone positions, i.e., the assumption of spatial randomness of these locations could not be rejected. The cones were placed on grass

	1	2	3	4	5	6	7	8	9	10
1	0	5.38	2.47	3.32	3.79	4.62	6.09	1.05	6.02	1.67
2		0	2.94	5.59	7.05	6.07	6.22	5.81	.93	5.84
3			0	3.36	4.61	4.33	5.25	3.07	3.54	3.30
4				0	1.50	1.29	2.88	4.37	5.73	4.98
5					0	1.97	3.60	4.75	7.23	5.41
6						0	1.69	5.66	6.01	6.27
7							0	7.13	5.91	7.69
8								0	6.54	.67
9									0	6.63
10										0

Table 1. The distance matrix of cone positions; distances are given in metres.

in order to ensure high visibility and a red-white coloured stick was placed in the centre of each cone to ensure that their location was fixed and to give participants a further hint on the location of the actual cone centre on the ground (see Figure 5 for a picture of the experimental area). The reason for using traffic cones was twofold. First, compared to, e.g., real green features such as trees, confounding effects with respect to height and diameter can be controlled as all cones were of equal size and shape. Second, the ground-truth geographic location can be measured using a PPM 10XX GNSS receiver external antenna (no tree coverage), which was inevitable because of the need to recreate the test setup on every of the $N_D = 8$ days of data collection. This means, first, we used the PPM 10XX GNSS receiver to position the cones as close as possible at their randomly chosen location. Using this method we were able to replicate the setup of locations within a range of 6 cm in easting direction and 9 cm in northing direction across days (the slight positional variation across different days is due to inclination of the sticks used to keep the cones in place). Second, to avoid confounding effects from pos-

sibly occurring differences in locations of traffic cones, we based all calculations of differences between mapped and ground truth positions on these daily measurements[5]. In addition to that, for each day, we performed a calibration around the experiment area (see Section 3.1) by collecting the calibration points at fixed spots. Using the best GNSS accuracy level, we ensured a transformation accuracy of less than $5cm$ by collecting the calibration points. Occasionally, however, the HoloLens would shut down during the day due to overheating because of the environmental temperature. In any such cases, we restarted the application and repeated the calibration procedure.

Figure 5. An image of the experimental area.

We employed a between-subjects design using two different conditions: In condition *GeoAR Glasses*, participants were asked to map the position of ten traffic cones as fast and as accurate as possible using the *GeoAR Glasses*; in condition *smartphone*, participants were assigned the same task using the OSMAND (V3.3.8) application. Participants assigned to the *smartphone* condition had to do one additional step to complete the same task: assign each mapped cone a number (see Figure 3). As we expected to find a large effect we calculated $N = 42$ as a suitable sample size for comparing independent groups, based on the significance level $\alpha = .05$, power $1 - \beta = .8$, ratio of group sizes $N_2/N_1 = 1$ and effect size $d = .8$ (Cohen, 1988) using the tool GPower (Faul et al., 2007). Participants were acquired in-situ, i.e. next to the experimental area which was located in a park. They were not reimbursed immediately for participation but could take part in a lottery using a Samsung Galaxy Tab 10.1 as a possible price. A prerequisite for participation was that participants did not contribute to OpenStreetMap on a regular basis. All experiments were conducted in German language. We ran experiments successively, i.e. we completed all *smartphone* trials first, and, once finished, continued to run *GeoAR Glasses* experiments. Across conditions participants were subject to the same procedure: They, first, gave their informed consent to participate in the experiment. Next, they filled in a demographic data survey also comprising questions about their prior engagement in mapping tasks, usage of smartphones and their experience with augmented and/or virtual reality devices. The task description was then given to participants in writing. The task description explicitly asked participants to map the geographic position of the corresponding centre for each traffic cone as accurate and

as fast as they could. After having read the task description, participants were subjected to a short training phase for the mapping task in order to familiarise them with the task and their assigned condition. The training area was located next to the experimental area.

All *GeoAR Glasses* participants were taught about the airtap gesture and the functionalities of the application and helped to correctly wear the glasses, making sure that the cursor was visible. Throughout the training phase, the participant was asked to map 3 manhole covers next to the experiment area using the method described above (see Section 3.2). The experimenter was able to check the results of the participants with a log console on a tablet, thereby guiding participants through the training process ensuring that participants familiarised themselves with all functionalities. In particular, participants were explicitly asked to use the edit mode.

In order to ensure comparability, *smartphone* participants were required to map the three manhole covers which were also used for *GeoAR Glasses* participants. The experimenter explained the required functionalities of the mapping application in a step-wise manner and, thereby, has been mapping the first manhole cover. Subsequently, participants were asked to map the remaining 2 manhole covers. Again, particular attention was paid to the edit mode functionality.

The training phase lasted until all questions by participants were settled. Participants were then guided to the fixed start location. At this point in time, the task was again explained to them in detail. Participants of condition *smartphone* were then informed that the statue called *Der letzte Mensch* (The last man), which is still present in the OpenStreetMap map, does no longer exist. Only a small, concrete rectangle which is located next to the starting position remained of this statue. Having, again, answered possible questions, participants engaged in the actual experimental task of mapping ten traffic cones as fast and as accurate as possible in their given order. Having finished the task, participants filled in several post-task questionnaires; one of these was the German language *Fragebogen räumliche Strategien* (Münzer and Hölscher, 2011), a self-report measure of spatial abilities.

4. Results

Due to space limitations, we focus on two lines of results in comparing mapping results using our new *GeoAR Glasses* approach to the OSMAND (V3.3.8) approach. We first provide empirical evidence for an increased absolute positional accuracy when the *GeoAR Glasses* approach is used. Second, we report on a large-sized time difference between the two approaches[6].

4.1 Descriptive Statistics

Overall, $N = 42$ persons participated within $N_D = 8$ days of data collection during May 2019 and early June 2019[7].

---

[5]Due to technical reasons, ground truth measurements were not recorded on two days; we used median imputation to account for these missing datasets.

[6]Data analysis was conducted using GNU R v3.6.0 and its packages dplyr v0.8.1 (Wickham et al., 2019), sp v1.3.1 (Bivand et al., 2013), ggplot2 v3.1.1 (Wickham, 2016), raster v2.9.5 (Hijmans, 2019), coin v1.3.0 (Hothorn et al., 2008), tidyr v0.8.3 (Wickham and Henry, 2019), bootES v1.2 (Gerlanc and Kirby, 2015).

[7]A rather large number of days was needed to finish data collection due to the fact that participants were acquired in-situ as well as due to weather conditions (heat) in this period of time.
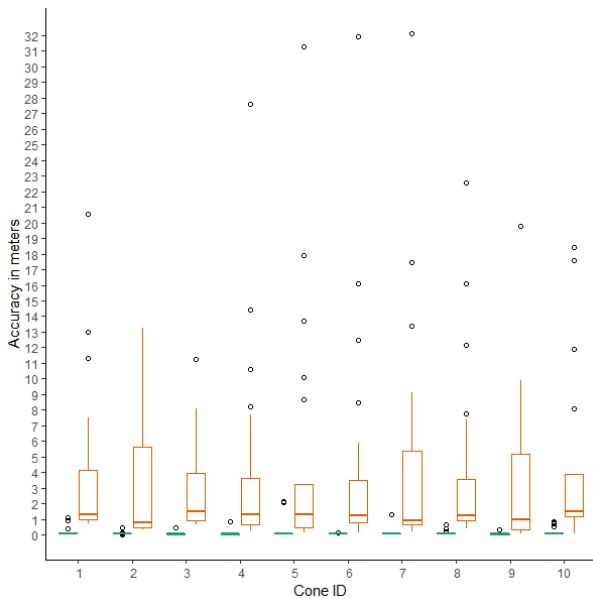
Figure 6. Boxplots of deviations between actual geographic locations of traffic cones and the positions set by participants stratified by condition (green: *GeoAR Glasses*, orange: Smartphone). All accuracy values are given in metres.

On average, participants (20 females and 22 males) were 33.67 years of age ($MD = 31$ years, $range = 19 - 72$ years). The average task duration across conditions was 7.1 minutes ($MD = 5.9$ minutes, $range = 1.27 - 19.48$ minutes). The majority (15 out of 21 participants) of *GeoAR Glasses* participants (9 females, 12 males) indicated no or only very slight experience with augmented reality, whereas the majority of the control group participants (11 females, 10 males) indicated high or very high experience with smartphones (14 out of 21 participants). As the *smartphone* condition particularly requires self-localization skills, we administered a self-report sense of direction scale (FRS, (Münzer and Hölscher, 2011)) to all participants of both groups. A Wilcoxon signed-ranks test, however, did not indicate a significant difference of self-report sense of direction between *smartphone* ($MD = 93$) and *GeoAR Glasses* ($MD = 88$) users ($Z = .39$, $p = .70$, $r = .06$). This result indicates that no bias between the two groups has been induced by the participants' sense of direction[8].

4.2 Accuracy of absolute positions

It is important to note, that the PPM 10XX GNSS receiver was connected to the smartphone used by participants to collect the data; thereby, we ensure a high accuracy of GPS-positioning at approx. 2cm. As mentioned above (see section 3.2), the antenna was used to calibrate the HoloLens application; there was, however, no need to use the antenna during the experiment due to our spatial transformation functionality. The deviations between mapped positions and ground-truth geographic locations of traffic cones were calculated on a daily basis for both conditions. In doing so, we avoid biases which might result from slight differences in the setup between days (see Section 3.4). Figure 6 presents the deviations between the locations mapped by participants and the ground-truth locations of traffic cones for group *GeoAR Glasses* and *smartphone* for each traffic cone. All Tukey boxplots (McGill et al., 1978, Wickham, 2016) for group *smartphone* (shown in orange) are heavily right-skewed across cones and the inter-quartile range of deviations is not equal across cones. Contrastingly, boxplots show a very low spread and a very high accuracy across both, participants and cones for the *GeoAR Glasses* condition (shown in green).

Based on a bootstrapping procedure (Efron, 1979) using $B = 10,000$ resamples we find 95%-CIs for differences in means between the two groups as well as the effect size based on Hedges' g (Hedges, 1981). Bias-corrected and accelerated confidence intervals (Efron, 1987) are used because these are known to show only very small coverage errors for means – regardless the normality of the population distribution (Wang, 2001); in accordance with the advice given in (Kelley, 2005) Hedges' g was used as an effect size, which is also presented in Table 2.

Cone	M_S	M_H	Δ_M	95%-$CI_{\Delta M}$	g	95%-CI_g
1	3.92	0.18	3.75	[2.14; 6.76]	1.01	[0.71; 1.30]
2	3.20	0.07	3.13	[1.72; 5.21]	1.10	[0.77; 1.50]
3	2.82	0.08	2.74	[1.80; 4.37]	1.28	[0.99; 1.62]
4	4.09	0.10	3.99	[2.02; 8.15]	0.83	[0.54; 1.10]
5	4.68	0.27	4.42	[2.02; 9.10]	0.78	[0.49; 1.07]
6	4.30	0.06	4.24	[2.16; 9.31]	0.78	[0.51; 1.04]
7	4.85	0.13	4.72	[2.29; 10.04]	0.82	[0.53; 1.11]
8	4.11	0.13	3.99	[2.13; 7.44]	0.93	[0.67; 1.23]
9	3.29	0.06	3.24	[1.73; 6.13]	0.92	[0.61; 1.24]
10	4.13	0.20	3.94	[2.13; 7.02]	0.99	[0.73; 1.30]

Table 2. Differences (Δ_M) in means between the *GeoAR Glasses* (M_H) and *smartphone* (M_S) condition per cone. The confidence intervals reported for these differences are bias corrected and accelerated and were found for each cone separately ($B = 10000$ resamples); similarly, column g reports Hedge's g as an effect size and the corresponding BCa 95-% CIs.

These results indicate a significant between-groups difference in accuracy for each cone; participants assigned to the *GeoAR Glasses* condition yield more accurate results and the effects are very large according to the classification by (Cohen, 1988). In general, the positional accuracy achieved using the *GeoAR Glasses* approach ranges between $6cm$ and $27cm$. This is in sharp contrast to the results achieved by the control group which shows a range in positional accuracy between $2.82m$ and $4.85m$. The least difference ($2.74m$) between the two conditions was achieved for cone 3, whereas the largest difference ($4.72m$) was found for cone 7.

4.3 Time

In case of the *GeoAR Glasses* condition, overall task time was automatically logged by the application during trials. OSMAND (V3.3.8), however, is not able to log time. We, therefore, measured task time based on the screencasts recorded, as these provide a good indicator when trials start and finish. A potential bias, however, may result from the time elapsed when the smartphone was given to participants by the experimenter and vice-versa. Therefore, we subtracted 30 seconds from the experiment time derived from screencasts in order to take these actions into account.

---

[8]We plan to further evaluate the impact of spatial abilities in general in a future study.

Table 3 provides summary statistics for overall task time for both conditions. *GeoAR Glasses* users outperform smartphone users with respect to several aspects.

Cond.	MIN	$x_{.25}$	MEAN	MD	$x_{.75}$	MAX	IQR	SD
GeoAR Glasses	76	152.5	246.1	184.5	309.5	598	157	157.99
Smartphone	221	409.8	612.5	617	795	1159	385.2	257.28

Table 3. Summary statistics for overall task completion time per condition. All values are given in seconds.

The minimum completion time for the *GeoAR Glasses* condition was almost a third of the time needed by the fastest *smartphone* condition participant. The median for the *smartphone* condition was more than three times as large as the median for the *GeoAR Glasses* condition. The slowest participant in condition *GeoAR Glasses* (598 seconds) was still faster than the upper half of the *smartphone* condition participants (617 seconds). On average, task completion time for *GeoAR Glasses* condition participants was only 25 seconds more than the fastest *smartphone* participant. Moreover, the lower half of *GeoAR Glasses* condition participants was faster than the fastest *smartphone* condition participant. Vice versa, completing the task took the slowest *smartphone* condition participant almost twice the time of the slowest *GeoAR Glasses* condition participant.

Using $B = 10,000$ resamples we, again, constructed bias-corrected and accelerated confidence intervals for contrasts and Hedges' g. The 95%-CI found for the difference in means, $\Delta_M = 366.5$ seconds, was $[234.87; 495.65]$; according to (Cohen, 1988) Hedges' g ($g = 1.68$, $[1.00; 2.43]$ 95%-CI) indicated a large-sized effect, too.

5. Discussion

Generally speaking, the results provide evidence that the GeoAR-approach is superior to the smartphone approach with respect to both positional accuracy and mapping time. The results have seen the least difference between the two conditions for the position of cone 3. This result is an indicator that closeness to other features helps participants: Cone 3 was located close to the position of a no longer existing statue which traces are still visible on the ground. The location of this statue has also been part of the OpenStreetMap map. Participants will likely have relied on this feature. This explanation is also stressed by observing participants during the trials: One major difficulty *smartphone* participants faced with was to align the digital with the real world. These participants did not succeed to either rotate the map mentally or digitally. These difficulties explain the outliers presented in Figure 6 as this resulted in cones being spread all over the green area. The egocentric approach used in condition *GeoAR Glasses* eliminates these issues. Another issue some *smartphone* participants complained about was "the jumping" blue marker which indicated the user position. This was probably due to the large scale of the map, where one step actually was considered as a "jump" on the map and to the relatively slow update rate of the GNSS antenna (1Hz) that does not allow for a smooth transition of the position marker (blue dot).

The boxplots comparing the deviance of positions for *smartphone* provide evidence that a closer look into the impact local spatial layout has is promising. Participants were required to map cones in a predefined order. The results also indicate that spatial configuration may not be as important for positional accuracy as one might expect. Cones 4 to 7 are located next to each other, with cone 7 being closest to the boundary of the green area the cones were positioned in (see Figure 5). The variability of positions for condition *smartphone* is, roughly equal for cones 4 to 6; cone 7, however, has seen an increased variability of cone positions – despite its closeness to the boundary of the green area. Two possible explanations come to mind: Either participants used the layout of 4, 5 and 6 to estimate the position of 7 and, hence, the positional error made earlier yielded an increased impact on 7; or, participants relied on the absolute position of 7 in relation to the graphical representation of the green area, resulting in an increased error because the local spatial configuration of other cones was neglected.

Across cones, the results show a generally high accuracy for condition *GeoAR Glasses* ranging from $6cm$ to $27cm$. This result shows, first, that the GeoAR approach allows for very accurate results in mapping even distant objects, i.e. objects which cannot be walked on. On the other hand, the large mean deviations seen for cones 1, 5 and 10 can be largely explained by user behaviour. Participants did either not notice offsets (participant H13 (cone 10), H38 (cone 1), H20 (cone 5) or did not hit the cone with the HoloLens cursor (participant H20 (cone 10), H13 (cone 5) for cones 10 and 5). Finally, two further outliers regarding cone 10 result from distortions of the spatial mesh (see Section 5.1 for an explanation).

5.1 Limitations

The Microsoft HoloLens v1 has difficulties to create a spatial mesh for cones (and other non-rectangular solid figures). This resulted in some distortions of the mesh in particular in the upper third of a cone. Hitting cones at roughly this height yields a decreased positional accuracy. Participants, however, might not notice this shift, in particular when they do not move during the experiment.

A second limitation applies to the *smartphone* condition: Some participants accidentally turned on the navigation in OSMAND (V3.3.8) during the experiment, which in turn results in an increase of task completion time. This increase is, on the other hand, a result of interaction problems of the application and, therefore, needs not be taken into account during the analysis.

6. Conclusions and Future work

In this paper we introduced the *GeoAR Glasses*, a Microsoft HoloLens modified to operate outdoors and added the capability to handle geographic coordinates. In sharp contrast to other solutions, our *GeoAR Glasses* are able to determine geographic locations without live GNSS data stream once the calibration phase has finished. This paper reported on the results of a between-subjects design study comparing mapping accuracy and time for two conditions. Participants either used a smartphone application called OSMAND (V3.3.8) or our *GeoAR Glasses*. The figures found provide evidence for a large increase in accuracy and a large decrease in time demands using the *GeoAR Glasses*. We are going to follow three paths of future work: We will, first, investigate the impact overall spatial abilities and its subdimensions have on mapping performance, e.g.

we try to gain further insights whether highly allocentric oriented people are faster and more accurate in the *smartphone* condition than highly egocentric-oriented ones. In this context, we will also investigate whether the presence of correctly mapped entities in the *smartphone* condition has an effect on the mapping accuracy. Second, we are going to analyse the impact of the distance covered during the experiment might have on both, positional and configuration accuracy in the *smartphone* condition. Third, we will compare both conditions with respect to task load and user experience.

References

Azuma, R., Hoff, B., Neely, H. and Sarfaty, R., 1999. A motion-stabilized outdoor augmented reality system. In: *Proceedings IEEE Virtual Reality (Cat. No. 99CB36316)*, IEEE, pp. 252–259.

Baillot, Y., Brown, D. and Julier, S., 2001. Authoring of physical models using mobile computers. In: *Proceedings Fifth International Symposium on Wearable Computers*, IEEE, pp. 39–46.

Bivand, R. S., Pebesma, E. and Gomez-Rubio, V., 2013. *Applied spatial data analysis with R, Second edition*. Springer, NY.

Brovelli, M. A., Minghini, M., Molinari, M. E. and Zamboni, G., 2016. Positional accuracy assessment of the openstreetmap buildings layer through automatic homologous pairs detection: The method and a case study. *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* XLI-B2, pp. 615–620.

Cohen, J., 1988. *Statistical power analysis for the behavioral sciences*. 2 edn, Academic Press, New York.

Efron, B., 1979. Bootstrap Methods: Another Look at the Jackknife. *The Annals of Statistics* 7(1), pp. 1–26.

Efron, B., 1987. Better Bootstrap Confidence Intervals. *Journal of the American Statistical Association* 82(397), pp. 171–185.

Fan, H., Zipf, A., Fu, Q. and Neis, P., 2014. Quality assessment for building footprints data on openstreetmap. *International Journal of Geographical Information Science* 28(4), pp. 700–719.

Faul, F., Erdfelder, E., Lang, A. and Buchner, A., 2007. G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavioral Research Methods* 39(2), pp. 175–191.

Feiner, S., MacIntyre, B., Höllerer, T. and Webster, A., 1997. A touring machine: Prototyping 3d mobile augmented reality systems for exploring the urban environment. *Personal Technologies* 1(4), pp. 208–217.

Fogliaroni, P., D'Antonio, F. and Clementini, E., 2018. Data trustworthiness and user reputation as indicators of vgi quality. *Geo-spatial Information Science* 21(3), pp. 213–233.

Gerlanc, D. and Kirby, K., 2015. bootES: Bootstrap Effect Sizes. R package version 1.2.

Ghilani, C. D., 2017. *Adjustment computations: spatial data analysis*. John Wiley & Sons.

Goodchild, M., 2007. Citizens as sensors: The world of volunteered geography. *GeoJournal* 69(4), pp. 211–221.

Haklay, M., 2010. How good is volunteered geographical information? a comparative study of openstreetmap and ordnance survey datasets. *Environment and Planning B: Planning and Design* 37(4), pp. 682–703.

Hedges, L. V., 1981. Distribution theory for Glass's estimator of effect size and related estimator. *Journal of Educational Statistics* 6(2), pp. 107–128.

Hijmans, R. J., 2019. raster: Geographic Data Analysis and Modeling. R package version 2.9-5.

Höllerer, T., Feiner, S., Terauchi, T., Rashid, G. and Hallaway, D., 1999. Exploring mars: developing indoor and outdoor user interfaces to a mobile augmented reality system. *Computers & Graphics* 23(6), pp. 779–785.

Hothorn, T., Hornik, K., van de Wiel, M. A. and Zeileis, A., 2008. Implementing a class of permutation tests: The coin package. *Journal of Statistical Software* 28(8), pp. 1–23.

Kelley, K., 2005. The effects of nonnormal distributions on confidence intervals around the standardized mean difference: Bootstrap and parametric confidence intervals. *Educational and Psychological Measurement* 65, pp. 51–69.

Li, W., Han, Y., Liu, Y., Zhu, C., Ren, Y., Wang, Y. and Chen, G., 2018. Real-time location-based rendering of urban underground pipelines. *ISPRS International Journal of Geo-Information* 7(1), pp. 32.

Marder-Eppstein, E., 2016. Project tango. In: *ACM SIGGRAPH 2016 Real-Time Live!*, ACM, p. 40.

McGill, R., Tukey, J. W. and Larsen, W. A., 1978. Variations of box plots. *The American Statistician* 32, pp. 12–16.

Mooney, P. and Minghini, M., 2017. A review of openstreetmap data. In: *Mapping and the Citizen Sensor*, Ubiquity Press, pp. 37–59.

Münzer, S. and Hölscher, C., 2011. Entwicklung und Validierung eines Fragebogens zu räumlichen Strategien. 57(3), pp. 111–125.

Neis, P., Zielstra, D. and Zipf, A., 2012. The street network evolution of crowdsourced maps: Openstreetmap in germany 2007–2011. *Future Internet* 4(1), pp. 1–21.

OpenStreetMap, 2018. Pick Your Mapping Technique. https://wiki.openstreetmap.org/wiki/Pick_your_mapping_technique. [Online; accessed 13-June-2019].

OpenStreetMap, 2019a. Android. https://wiki.openstreetmap.org/wiki/Android. [Online; accessed 13-June-2019].

OpenStreetMap, 2019b. Armchair Mapping. https://wiki.openstreetmap.org/wiki/Armchair_mapping. [Online; accessed 13-June-2019].

Reitmayr, G. and Drummond, T., 2006. Going out: robust model-based tracking for outdoor augmented reality. In: *ISMAR*, Vol. 6, pp. 109–118.

Ribeiro, A. and Fonte, C., 2015. A methodology for assessing openstreetmap degree of coverage for purposes of land cover mapping. *ISPRS Annals of Photogrammetry, Remote Sensing & Spatial Information Sciences*.

Senaratne, H., Mobasheri, A., Ali, A. L., Capineri, C. and Haklay, M. M., 2017. A review of volunteered geographic information quality assessment methods. *International Journal of Geographical Information Science* 31(1), pp. 139–167.

Wang, F. K., 2001. Confidence interval for the mean of non-normal data. *Quality and Reliability Engineering International* 17(4), pp. 257–267.

Wickham, H., 2016. *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York.

Wickham, H. and Henry, L., 2019. tidyr: Easily Tidy Data with 'spread()' and 'gather()' Functions. R package version 0.8.3.

Wickham, H., François, R., Henry, L. and Müller, K., 2019. dplyr: A Grammar of Data Manipulation. R package version 0.8.1.

Zielstra, D., Hochmair, H. H. and Neis, P., 2013. Assessing the effect of data imports on the completeness of openstreetmap–au nited s tates case study. *Transactions in GIS* 17(3), pp. 315–334.