Spatially Localised Immersive Contemporary and Historic Photo Presentation on Mobile Devices in Augmented Reality

Loris Sauter
loris.sauter@unibas.ch
University of Basel
Basel, Switzerland

Tim Bachmann
t.bachmann@unibas.ch
University of Basel
Basel, Switzerland

Luca Rossetto
rossetto@ifi.uzh.ch
University of Zurich
Zurich, Switzerland

Heiko Schuldt
heiko.schuldt@unibas.ch
University of Basel
Basel, Switzerland

ABSTRACT

These days, taking a photo is the most common way of capturing a moment. Some of these photos captured in the moment are never to be seen again. Others are almost immediately shared with the world. Yet, the context of the captured moment can only be shared to a limited extent. The continuous improvement of mobile devices has not only led to higher resolution cameras and, thus, visually more appealing pictures but also to a broader and more precise range of accompanying sensor metadata. Positional and bearing information can provide context for photos and is thus an integral aspect of the captured moment. However, it is commonly only used to sort photos by time and possibly group by place. Such more precise sensor metadata, combined with the increased computing power of mobile devices, can enable more and more powerful Augmented Reality (AR) capabilities, especially for communicating the context of a captured photo. Users can thereby witness the captured moment in its real location and also experience its spatial contextualization. With the help of a suitable data augmentation, such context-preserving presentation can be extended even to non-digitally born content, including historical images. This offers new immersive ways to experience the cultural history of one's current location. In this paper, we present an approach for location-based image presentation in AR on mobile devices. With this approach, users can experience captured moments in their physical context. We demonstrate the power of this approach based on a prototype implementation and evaluate it in a user study.

CCS CONCEPTS

• **Information systems → Multimedia and multimodal retrieval**; **Search interfaces**; *Location based services.*

KEYWORDS

multimedia retrieval, spatial-retrieval, augmented reality, cultural heritage, mobile application

ACM Reference Format:
Loris Sauter, Tim Bachmann, Luca Rossetto, and Heiko Schuldt. 2023. Spatially Localised Immersive Contemporary and Historic Photo Presentation on Mobile Devices in Augmented Reality. In *Proceedings of the 5th Workshop on the analySis, Understanding and proMotion of heritAge Contents (SUMAC '23), November 2, 2023, Ottawa, ON, Canada.* ACM, New York, NY, USA, 7 pages. https://doi.org/10.1145/3607542.3617358

SUMAC '23, November 2, 2023, Ottawa, ON, Canada

ACM ISBN 979-8-4007-0279-2/23/11.
https://doi.org/10.1145/3607542.3617358

1 INTRODUCTION

One of the most common ways to show and share an experienced moment is by capturing said moment as a photo. Studies show that revisiting captured photos has a positive influence on general well-being [5, 11]. With the omnipresence of smartphones, such photos can be taken whenever and wherever deemed fit. Consider, for example, scenarios such as the following. A once-in-a-century snowfall prompts a mother to take a picture of her heavily snow-covered garden to eventually show her kids. In a foreign city, a tourist takes a selfie at an interesting plaza to preserve the memory of the visit. One of the commonalities of these scenarios is the fact that the experienced moment is out of the ordinary and, as such, deemed worthwhile to persist in the form of a photo, some even to be shared with others. The lack of a certain local context upon showing the photos to others is present in all three scenarios: It being the stark difference between when the area of the festival is back to normal, the true extent of how much snow is in the garden, or the immediate surroundings of the selfie; In all three scenarios, a more location-based context would be beneficial in order to show and share the experience.

Smartphones already capture more data for photos and store them as metadata, mostly information on the capturing device, such as optical information, but also the timestamp and GPS location in the form of EXIF metadata [1]. One way of using this spatial information could be by presenting these images on a map, leveraging the geo-coordinates that are part of the EXIF metadata format. However, we argue that while this approach is feasible, there is a loss of information caused by the reduction of dimensionality, as these maps usually are two-dimensional. In particular, since the perspective of ground-level shot photos is perpendicular to what a map typically represents.

The technology of Augmented Reality (AR) has been continuously improved in recent years, and most smartphones support the usage of AR or even actively provide the means to leverage the technology [14]. Since smartphones were already used to capture the images, the very same devices might also be able to present the photos using AR, given some mapping of real-world coordinates to an AR space. This way, a smartphone's camera feed can be leveraged and augmented with these images. When locally anchored, reproducing the original perspective upon capturing is achievable. Thus, being at the very location of the snapshot provides a local context in a way that would otherwise hardly be possible.

With the rise of digitized archive materials, non-digital-born content can also be considered. However, this imagery conventionally is not annotated with such metadata. In some cases, archival metadata might be available alongside such content with either no spatial information or far less fine-grained in the form of a region name. Similarly, for temporal metadata, the resolution might vastly differ from the precise measurements of modern devices. Nevertheless, given any method to add the required spatial and temporal metadata, these historical images can be spatially contextualized as well.

In this paper, we present a method for spatially contextualized image presentation in Augmented Reality on mobile devices. We specify the required metadata to model the location-based recreation of a perspective. Using that, we implement our methodology for the reproduction and spatial contextualization of a captured image. An evaluation carried out on this proof-of-concept implementation of our approach demonstrates the method's usability. Furthermore, we show examples of instances where, given enough metadata, the same principles hold for historical imagery.

The remainder of this paper is structured as follows: after a brief overview of related work in Section 2, we present our approach of spatially localized immersive photo presentation using AR in Section 3. In Section 4, we discuss the architecture and key elements of our prototype implementation and present the findings of our evaluation in Section 5 before finally concluding in Section 6.

2 RELATED WORK

Augmented Reality enables novel interaction with the world and its inhabitants. One of the ways for such interaction is through Location-based Social Networks (LBSNs) [15]. At its core, LBSNs provide the means of connecting to others through physical proximity and geo-tagged social media content [17]. Yue et al. [16] propose such an LBSN for mobile AR with which they enable users to publish geo-referenced content in the form of 3D models and specific interaction modes such as tapping these 3D modes to trigger user-defined effects. Additionally, a leader-follower and reputation mechanism expands upon the sharing of content, which is the foundation of user-user interaction. Their prototype implementation, similarly to ours, targets iOS devices and builds on top of Apple's AR technology. The geo-referenced object placement in AR is established by tracking the distance between the virtual object and smartphone based on GPS coordinates and enhanced with AR motion tracking. A client-server approach is used to facilitate the social interaction components such as sharing.

In the field of urban or landscape planning, AR is also a promising technology. For instance, using AR, planned objects can be virtually positioned and, in doing so, enable users to understand the changes with them. In [9], the authors propose an AR system that renders such planned objects, e.g., wind turbines. Additionally, the authors aim to provide a realistic depth perception of the placed objects. In their approach, this problem is simplified with the argument that particularly since wind turbines are large structures, "*... everything that is not the sky is in front of the turbine, and the sky is behind it.*" [9, Section 5]. Client-side, on the mobile device, they rely for positional information on GPS sensor readouts in addition to the compass for orientation. Server-side, the objects to place are embedded in a Geographic Information System (GIS). In summary, objects then are in the feature layer of the GIS server, and cartography data is in the map layer. Their prototype is based on Unity3D[1] and AR Foundation[2], which in turn provides the functionality manufacturer-provided AR software development kits such as Apple's ARKit[3] or Google's ARCore[4].

Prior to this work, we had built a system with a focus on reproducing historic image perspectives [13]. This predecessor only operated on historical images that were manually annotated with spatial and temporal metadata. The work includes two presentation modes: the first superimposed the historic image onto the camera feed of the smartphone. A slider enables users to control the opacity. Alignment to the real world has to be performed manually by repositioning the smartphone. In contrast to the aforementioned overlay mode, the presented AR mode is a predecessor to the one proposed in this paper. However, due to the historical imagery, there were only three parameters used for the AR placement of the images: latitude, longitude, and bearing (the angle between magnetic north and the image). For the implementation, we used Unity3d, and AR was facilitated through Google's ARCore framework.

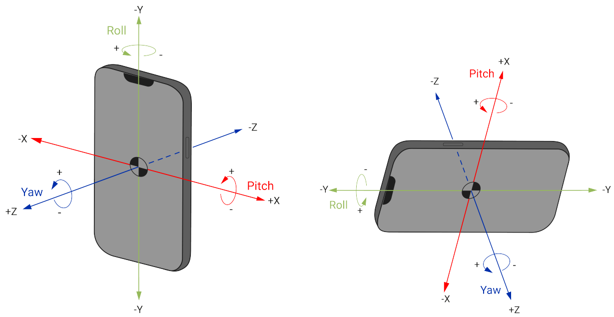
In the domain of cultural heritage, the reproduction of historical perspectives is also a key concern of [8]. The authors propose a system for semi-automatic geolocalisation of historic imagery. Nevertheless, possibly due to a significant amount of areal imagery in their dataset, the authors' presentation mode is not facilitated in AR. Quite the contrary, the authors propose a system for conventional two-dimensional displays that render a three-dimensional world.

Other work, specifically targeting applications in cultural heritage whose focus is related to ours, is a browser-based mobile application primarily projecting textures of historical photos on simple 3D objects of buildings [10]. The authors emphasize their choice of a mobile web-browser-based application in order to mitigate reservations against installing an application. Furthermore, the proposed application renders the 3D world based on open data information enhanced with their custom image-to-texture pipeline using the mobile device's location and orientation. While our work shares the application to ultimately present at-location historic buildings, we argue that in the days of dedicated mobile apps for almost everything, the hardware benefits of a native app outweigh the limitations suggested by the authors. Additionally, we use AR technologies to project the photos at their real-world location directly into the camera feed for a very immersive experience.

3 SPATIAL PERSPECTIVE RECREATION

Our approach covers two vital aspects of spatially localized immersive photo perspective recreation: the capturing and the presentation. Within the former aspect, we are concerned with the minimal set of sensor readouts of a mobile device's sensors required for proper localization. Since we aim to reproduce the very same perspective using AR in a 3D world, we have multiple degrees of freedom to cover. The latter aspect takes into account the position of the user and how this sensor-based metadata is used to recreate said perspective.

[1]https://unity.com/
[2]https://unity.com/unity/features/arfoundation
[3]https://developer.apple.com/augmented-reality/
[4]https://developers.google.com/ar



(a) Portrait **(b) Landscape Left**

Figure 1: Orientation axes on mobile devices. For both modes, the yaw angle is relevant as it is perpendicular to the screen and, thus, the photo to be taken. Since Y- and X-axis switch upon entering landscape mode, it is necessary to also be aware of the orientation of the device.

3.1 Capturing

Mobile devices provide GPS readouts in order to provide spatial information in the form of latitude and longitude $\langle \varphi, \lambda \rangle$ tuples. Besides GPS information, most modern devices also automatically enrich imagery metadata with temporal information. For our purpose, there are more degrees of freedom, and hence, we require more parameters; Besides the location, we also care about how the photo is rotated along the axis perpendicular to the surface. One way of measuring this orientation is by using the *bearing*, the angle that represents the orientation towards magnetic North, where zero degrees is pointing towards North. We do not consider the drift of magnetic North over time since the introduced error is expected to be in the same range as the measurement noise of the sensors used.

The current three-tuple setup of latitude, longitude, and bearing provides essential information yet assumes the user holds the device perfectly upright and in portrait mode while taking a picture. Since this is not always the case, we additionally capture the yaw angle (see Figure 1) and the orientation of the smartphone.

The final metadata 5-tuple we store is as follows:

$$
\begin{aligned}
m = &\langle \varphi, \lambda, \alpha, \gamma, o \rangle \text{ with} \qquad (1)\\
&\varphi: \text{latitude (GPS sensor) in degrees } \varphi \in [0, 360)\,,\\
&\lambda: \text{longitude (GPS sensor) in degrees } \lambda \in [0, 360)\,,\\
&\alpha: \text{bearing (compass sensor) in degrees } \alpha \in [0, 360)\,,\\
&\gamma: \text{yaw (gyroscope sensor) in degrees } \gamma \in [0, 360)\,,\\
&o: \text{orientation (gyroscope sensor) with}\\
&\qquad o \in \{\texttt{LANDSCAPE RIGHT}, \texttt{LANDSCAPE LEFT},\\
&\qquad\quad \texttt{PORTRAIT}, \texttt{PORTRAIT UPSIDE DOWN}\}
\end{aligned}
$$

3.2 Presentation

In order to accurately recreate the perspective of a photo using Augmented Reality, there are several aspects to take care of. First and foremost, an understanding of the surroundings has to be established. This is granted by, for instance, visual odometry in order for AR to work. In our approach, we rely on smartphone-manufacturer-provided AR technology. Since the objective is the recreation of a specific perspective at its real-world location, its spatial context must be known. Two parameters of the 5-tuple metadata m cover location information in the form of latitude and longitude, which enables us to relate this position with the device's current position. In order to do so, we rely on the *haversine distance* [6], also known as great-circle distance. Due to the assumption of a perfect sphere, which Earth is not, the haversine distance has an average error of 0.5% on the surface. This is, however, negligible in our case due to the fact that our approach relies on a GPS signal, which in particular can be imprecise in urban areas. Equation (2) shows the haversine distance, where φ_1, λ_1 and φ_2, λ_2 are the latitude and longitude of the device and photo, respectively.

$$
d = 2r \cdot \arcsin\left(\sqrt{\sin^2\left(\frac{\varphi_2 - \varphi_1}{2}\right) + \cos(\varphi_1) \cdot \cos(\varphi_2) \cdot \sin^2\left(\frac{\lambda_2 - \lambda_1}{2}\right)}\right) \tag{2}
$$

Additionally, the difference in bearing between the current position and the to-be-presented photo must be calculated as well in order to have the angle between the two points, see Equation (3) with φ_1, λ_1 and φ_2, λ_2 again being the latitude and longitude of both the current position and photo.

$$
\begin{aligned}
\beta = \arctan^2\Big(&\sin(\varphi_2 - \varphi_1) \cdot \cos(\varphi_2), \qquad (3)\\
&\cos(\varphi_1) \cdot \sin(\varphi_2) - \sin(\varphi_1) \cdot \cos(\varphi_2) \cdot \cos(\varphi_2 - \varphi_1)\Big)
\end{aligned}
$$

At this point, the angle and distance are known, and we construct a local coordinate system to which we transform. The benefit of this local coordinate system is that its origin is at the user's location and simplifies the positioning in AR. We synchronize the origin with this local coordinate system with the GPS position of the device and, thus, re-calculate the positioning upon (notable) real-world movement.

Using the metadata m (see Equation (1)), we construct the transformation matrix as shown in Equation (4), ultimately resulting in the positional quaternion. Note that the distance d is negative due to the origin of the local coordinate system being at the user's position. Finally, we use constants to correct for the smartphone orientation, as shown in Table 1.

$$
T = \begin{pmatrix} \cos(\alpha) & \sin(\alpha) & 0 & 0 \\ -\sin(\alpha) & \cos(\alpha) & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}^{-1} \cdot \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & -d & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \tag{4}
$$

Table 1: The constant correction per smartphone orientation.

Device Orientation	Bearing Correction	Flip Image Axes
`Portrait`	0°	No
`Portrait Inverse`	0°	No
`Landscape Left`	-90°	Yes
`Landscape Right`	90°	Yes

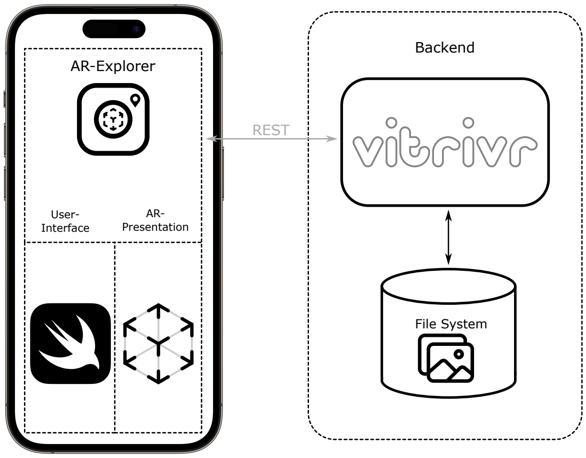


Figure 2: The high-level architecture diagram of our proposed system. REST communication is used between the server, an extension of vitrivr, and our iOS client. The server's filesystem is facilitated for cross-device image sharing.

4 IMPLEMENTATION

We implemented our approach for location-based image presentation in AR using a client-server architecture in order to share images as well as the required metadata. While our very first prototype in this domain was called *GoFind!* [13], we label this work as *AR-Explorer* in order to distinguish the different prototypes. However, *GoFind!* has evolved to an umbrella term for various prototypes in this domain.[5] The server is an extension of the open-source content-based multimodal multimedia retrieval stack vitrivr [7, 12]. The communication to the back-end is facilitated through a REST API, of which we leverage the spatio-temporal query support [3]. The high-level architecture of our system is shown in Figure 2

Apart from being our content-based image retrieval engine, the server also serves for cross-device sharing of imagery and corresponding metadata. For this purpose, we differentiate between *public images*, i.e., images that were shared within our system. Images created within our clients provide the necessary metadata upon sharing and local images. Additionally, we employ a very small collection of historic images of Basel. Spatial and temporal metadata was provided by the collection donor – a local association that curates historic landscape imagery in a 'back then' versus 'now' fashion. These historic images were marked as public and thus queryable by clients.

We started our proof-of-concept implementation in 2022, targeting devices with iOS 15 (and was later updated to also support iOS 16). Apple's ARKit, the AR framework, then recently released as version 5, including support for the LiDAR sensors of iPhones, and improved and faster plane detection. Originally released alongside of iOS 11 in 2011, ARKit enabled developers to facilitate AR applications for iOS devices. One of the features of ARKit is *location anchors*. In selected cities, there are provided anchors of geo-coordinates in the AR coordinate system, which enable an easy-to-use way to anchor AR content at a real-world location. However, since this feature is only available in selected cities, we cannot rely on it in a general-purpose approach. Instead, we have implemented our own mapping method as presented in Section 3. Furthermore, in order to promote shared moments, we employ the iOS-provided notification system to inform users of AR photos in close proximity. Besides notifications AR directions, based on Apple Maps,[6] are provided, given an image is selected. This way, users are able to find public photos within the system without local navigation knowledge. In Figure 3, we show key views within the implementation, specifically designed to be as intuitive as possible. From 3a to 3d, the screenshots represent one possible flow of interactions;

(a) On the *home screen*, the map is presented to the user with previously captured images. The buttons in the right-hand side menu bar enable (in order) the user to change the map type, jump back to the current location on the map, change to the gallery view, upload locally captured images to the server, search through public images on the server and the photo capture mode.
(b) With the *search view*, users can either search by date range or location, including a range filter. The collection to be searched within is a combination of one's own photos, public imagery that had to be previously imported into the server, or photos from other users that they marked as public.
(c) Once one particular photo is selected, an AR directions mode is enabled, and the *navigation view* superimposes a route to the target. This mode uses Apple Maps data to get the routing.
(d) In the *AR view*, the selected photo is placed in AR using our approach and presented to the user.

Besides the views shown in Figure 3, a camera mode with the default look and feel for iOS devices is incorporated into the prototype. Upon capturing an image, all necessary data is collected. Specifically, GPS, compass, and gyroscope sensor readouts are stored alongside the image locally and uploaded to the server if the user marked the image as public. Additionally, a traditional gallery view is implemented in order to get a quick overview that is not map-based of the photos available.

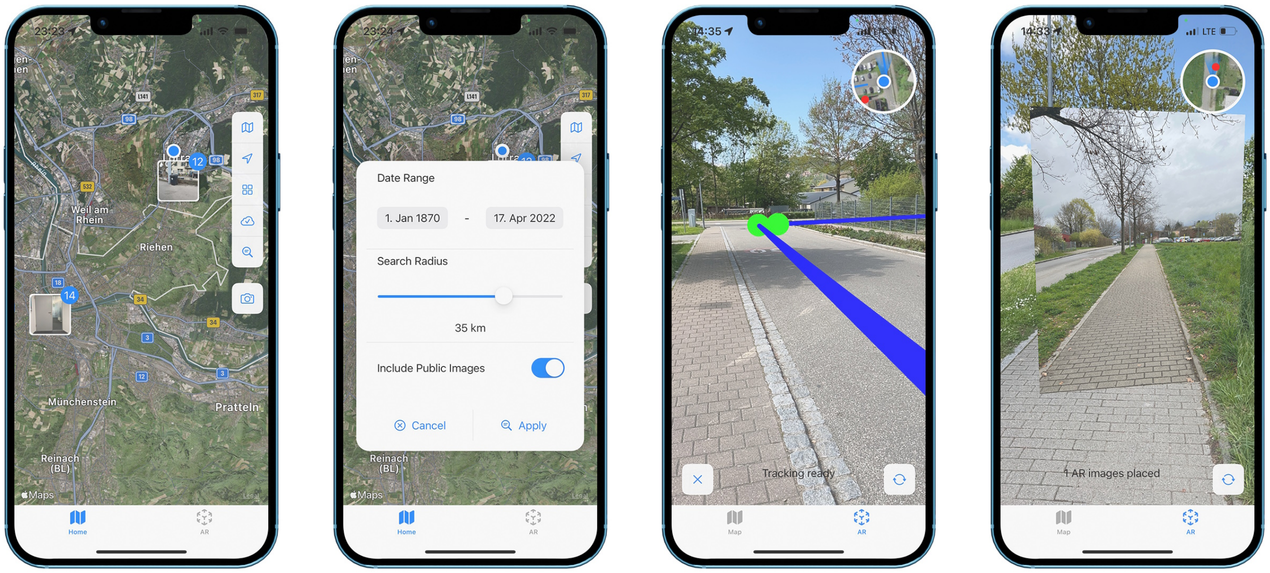
Revisiting our initial example scenarios, our proposed approach enables the presentation of the heavy snowfall in the garden as well as the spatial contextualization of photos taken during a visit to a foreign city. Furthermore, the prototype implementation, enriched with AR-supported navigation, particularly for foreign territories, users are directed to AR photos.

5 USER STUDY

To evaluate our approach, we created a prototype app as described in Section 4. The app targets iOS 15. Hence, supported devices include iPhones from model 6 and iPad models from iPad Air 2015 onward. A first evaluation round consisting of 20 participants was conducted from May to June 2022, and a second one with 8 participants was conducted in late April 2023. For both evaluations, predominately students from the Department of Mathematics and Computer Science at the University of Basel were recruited and

[5] Such as https://github.com/sauterl/gofind, https://github.com/sauterl/gofind-android and this work https://github.com/timbachmann/AR-Explorer

[6] https://www.apple.com/maps/



(a) Home Screen **(b) Search View** **(c) AR Navigation** **(d) AR View**

Figure 3: Screenshots of the proof of concept implementation.

(a) Contemporary View **(b) AR Historic View**

Figure 4: A side-by-side comparison between the plain contemporary view and the AR view with historical material.

given no time constraints. Since we want to demonstrate the usability of our proof-of-concept, we chose the System Usability Scale (SUS) by Brooke [4]. The ten-item questionnaire uses a five-point Likert scale on each item. The questions alternate between a positively and negatively stated sentiment and should not be used individually. Instead, only the resulting score is to be considered. Additionally to the SUS, we also asked participants in three scenarios, also on a five-point Likert scale, how satisfactory certain elements are. The scale's labels are 1. Disagreement, 2. Slight Disagreement, 3. Neutral, 4. Slight Agreement, and 5. Agreement.

5.1 Evaluation Scenarios

The evaluation covers three scenarios:

Image Perspective Reproduction. Participants are asked to capture a photo using the prototype and subsequently assess in the AR view how satisfactory the AR reproduction is. The main goal of this scenario is to get an impression of how well the proposed image perspective reproduction method is accepted and whether it delivers satisfactory results. We specifically omitted primarily historic perspective reproduction due to the small dataset available.

Walk and Transition. In this scenario, users were tasked to first take a picture, then move several meters and take another photo. Then, users have to walk from the second photo to the first while using the AR view. Our objective with this task is to verify that the transition from one shared moment to another is smooth with respect to AR-positioned objects. In doing so, users are enabled to experience multiple photos in AR mode that are close enough to walk from one to another while still keeping the AR view active. For the user study conducted in 2022, this scenario also included historical images.

AR Navigation. The third scenario's aim is to get a grasp of the usability and practicability of AR navigation powered by Apple Maps. Specifically, navigating to an AR image presentation location as a pedestrian relying on the directions given using AR was

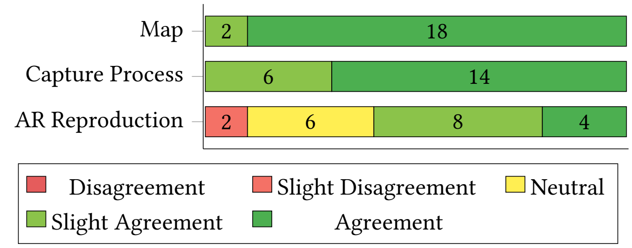


Figure 5: Absolute distribution of feedback regarding scenario one, including three questions and answers reaching from disagree to agree.

evaluated. In order to do so, participants were asked to select an image and follow the projected directions.

5.2 Results

In this section, we present the individual results for the user study from 2022 with 20 participants for each scenario. We close this section by providing the calculated SUS score for both studies.

Image Perspective Reproduction. For scenario one, we asked the users to express their agreement with the following statements on a five-point Likert scale: (i) The capture process was a pleasing experience, (ii) The image was displayed on the map after its capture, and (iii) The image was displayed in AR approximately in the same place it was captured. As depicted in Figure 5, the capturing process, as well as the map view, were positively received. The AR image perspective reproduction, on the other side, was received with slightly less favorable views, as there was neutral and slightly negative feedback.

Walk and Transition. In the second scenario, the three statements participants had to express their agreement to are as follows: (i) The radar (mini-map) functionality is intuitive, (ii) I was notified upon entering a 10m radius around the image, and (iii) The second image appeared in AR view while walking. Shown in Figure 6, mixed to negative results were received specifically to the notification question, and moderately mixed results regarding the smooth transition from one AR image to another one. Nevertheless, the mini-map (originally labeled as radar) was well received.

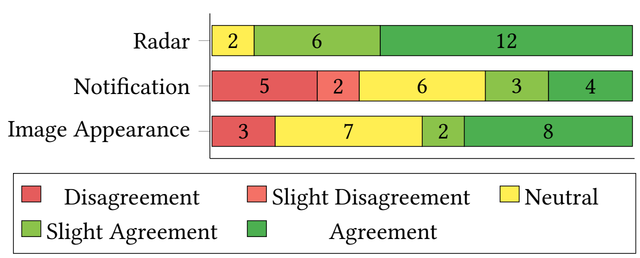


Figure 6: Absolute distribution of feedback regarding scenario two, including three questions and answers reaching from disagree to agree.

AR Navigation. The last of the three scenarios covers AR navigation, when users move toward an AR image using the directions provided by the app. The following two statements were presented to participants: (i) AR view displayed a blue route to the destination, and (ii) The route's heading is approximately correct. The results, given in Figure 7, were generally favorable. A little less so for the general route heading.

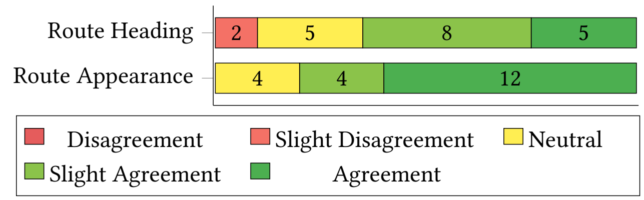


Figure 7: Absolute distribution of feedback regarding scenario three, including two questions and answers reaching from disagree to agree.

Usability. The mean SUS score of the 20 participants is 77 with a standard deviation of 19. According to [2], this results in *acceptable* usability with the adjective *good*. The acceptability ratings' range covers the entire SUS scoring range from 0 to 100, where scores below 50 are considered *not acceptable*, scores between 50 and 70 *marginally acceptable* and scores above 70 as *acceptable*.

The scoring for the far smaller user study of 8 participants results in a mean of 80 and a standard deviation of 13. We additionally asked the participants to label the usability with an adjective, according to [2], which reflects the SUS score (3 times "good" and 5 times "excellent"). The slightly higher mean of 80 still falls into the adjective label of *good* [2]. However, due to the small sample size, this score's expressiveness is limited.

5.3 Discussion

While both the SUS score as well as our questionnaire results are favorable, the robustness, i.e., how well an AR-placed image remains at the same real-world location, shows room for improvement. In particular, there is a tendency towards jumps: every now and so often, the entire AR world is moved by several meters in real-world distance. As a consequence, the longer the AR view is open, the more the AR-anchored elements drift from their original position. A workaround for the prototype was the implementation of an AR refreshment, which re-initializes the AR world based on the current real-world position of the device. All evaluation participants were enabled to use the refreshing feature whenever needed. Similarly, the method is bound to the reliability of the GPS signal and, as such, is not suitable for very dense high-rise areas with weak GPS. However, for Basel, this should not be an issue since no such dense high-rise areas exist.

The notification feature of our prototype to hint at available AR images was perceived comparatively negatively. In our implementation, we rely on a service provided by the operating system, which enables the registration of location triggers for a point and radius. We identify this as a potential cause that only in half of the evaluations the notification worked.

6 CONCLUSION AND OUTLOOK

In this paper, we presented a methodology to leverage spatial metadata from smartphone photos in order to present them in their original spatial context using Augmented Reality. The method includes the capturing of metadata beyond what is typically present for images, i.e., location data, as well as device orientation in the form of five parameters: latitude, longitude, bearing, yaw, and phone orientation. We implemented a proof-of-concept of our approach and conducted a small-scale user study according to the System Usability Scale, which results in a *good* usability. Additionally, we experimentally show that the method is applicable not only to contemporary, digital-born imagery but also to digitized archive content, such as historical city-scape images, manually annotated with the required spatial metadata.

The evaluation results are of limited expressiveness, as the user study was objectively small and rather homogeneous. While we were able to show the feasibility of the approach, there is still room for improvement. Apart from a larger evaluation, the method would benefit from additional means to position the photos in AR. Specifically, in the current prototype, the anchoring is limited and might jump multiple meters over the span of a few minutes. One approach to address this issue could be the employment of computer vision methods and the use of landmark detection on both the photo to be presented in AR and the live camera feed for better alignment and anchoring. The sharing of an experienced moment might be greatly enhanced by expanding upon imagery: video, with or without audio, is also applicable to our proposed approach.

ACKNOWLEDGMENTS

This work was partly funded by the Hasler Foundation in the context of the project City-Stories (Contract no. 17055) and the Swiss National Science Foundation through project "MediaGraph" (contract no. 202125).

REFERENCES

[1] Camera & Imaging Products Association. Established April 2010. Exchangealbe image file format for digital still cameras: Exif Version 2.32. *Exif, Version* Revised 2019 (Established April 2010).

[2] Aaron Bangor, Philip T. Kortum, and James T. Miller. 2008. An Empirical Evaluation of the System Usability Scale. *International Journal of Human-Computer Interaction* 24, 6 (July 2008), 574–594. https://doi.org/10.1080/10447310802205776

[3] Lukas Beck and Heiko Schuldt. 2016. City-Stories: A Spatio-Temporal Mobile Multimedia Search System. In *Multimedia (ISM), 2016 IEEE International Symposium on*. IEEE, 193–196.

[4] John Brooke. 2013. SUS - A quick and dirty usability scale. 57 (2013), 187–191.

[5] Yu Chen, Gloria Mark, and Sanna Ali. 2016. Promoting Positive Affect through Smartphone Photography. *Psychology of Well-Being* 6, 1 (Dec. 2016), 8. https://doi.org/10.1186/s13612-016-0044-4

[6] Kenneth Gade. 2010. A Non-singular Horizontal Position Representation. *Journal of Navigation* 63, 3 (2010), 395–417. https://doi.org/10.1017/S0373463309990415

[7] Ralph Gasser, Luca Rossetto, and Heiko Schuldt. 2019. Multimodal Multimedia Retrieval with vitrivr. In *Proceedings of the 2019 on International Conference on Multimedia Retrieval, ICMR 2019, Ottawa, ON, Canada, June 10-13, 2019*, Abdulmotaleb El-Saddik, Alberto Del Bimbo, Zhongfei Zhang, Alexander G. Hauptmann, K. Selçuk Candan, Marco Bertini, Lexing Xie, and Xiao-Yong Wei (Eds.). ACM, 391–394. https://doi.org/10.1145/3323873.3326921

[8] Florent Geniet, Valérie Gouet-Brunet, and Mathieu Brédif. 2022. ALEGORIA: Joint Multimodal Search and Spatial Navigation into the Geographic Iconographic Heritage. In *Proceedings of the 30th ACM International Conference on Multimedia*. ACM, Lisboa Portugal, 6982–6984. https://doi.org/10.1145/3503161.3547746

[9] Jan-Erik Kilimann, Denis Heitkamp, and Philipp Lensing. 2019. An Augmented Reality Application for Mobile Visualization of GIS-Referenced Landscape Planning Projects. In *Proceedings of the 17th International Conference on Virtual-Reality Continuum and its Applications in Industry*. ACM, Brisbane QLD Australia, 1–5. https://doi.org/10.1145/3359997.3365712

[10] Sander Muenster, Jonas Bruschke, Ferdinand Maiwald, and Constantin Kleiner. 2021. Software and Content Design of a Browser-based Mobile 4D VR Application to Explore Historical City Architecture. In *Proceedings of the 3rd Workshop on Structuring and Understanding of Multimedia heritAge Contents*. ACM, Virtual Event China, 13–22. https://doi.org/10.1145/3475720.3484442

[11] Akihiro Ogino and Reina Sato. 2022. Effects of Looking Back on Smartphone Photography on Positive Emotions. In *2022 23rd ACIS International Summer Virtual Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing (SNPD-Summer)*. IEEE, Kyoto City, Japan, 148–153. https://doi.org/10.1109/SNPD-Summer57817.2022.00033

[12] Luca Rossetto, Ivan Giangreco, Claudiu Tanase, and Heiko Schuldt. 2016. vitrivr: A Flexible Retrieval Stack Supporting Multiple Query Modes for Searching in Multimedia Collections. In *Proceedings of the 2016 ACM Conference on Multimedia Conference, MM 2016, Amsterdam, The Netherlands, October 15-19, 2016*, Alan Hanjalic, Cees Snoek, Marcel Worring, Dick C. A. Bulterman, Benoit Huet, Aisling Kelliher, Yiannis Kompatsiaris, and Jin Li (Eds.). ACM, 1183–1186. https://doi.org/10.1145/2964284.2973797

[13] Loris Sauter, Luca Rossetto, and Heiko Schuldt. 2018. Exploring Cultural Heritage in Augmented Reality with GoFind!. In *2018 IEEE International Conference on Artificial Intelligence and Virtual Reality (AIVR)*. IEEE, Taichung, Taiwan, 187–188. https://doi.org/10.1109/AIVR.2018.00041

[14] Yushan Siriwardhana, Pawani Porambage, Madhusanka Liyanage, and Mika Ylianttila. 2021. A Survey on Mobile Augmented Reality With 5G Mobile Edge Computing: Architectures, Applications, and Technical Aspects. *IEEE Communications Surveys & Tutorials* 23, 2 (2021), 1160–1192. https://doi.org/10.1109/COMST.2021.3061981

[15] Panagiotis Symeonidis, Dimitrios Ntempos, and Yannis Manolopoulos. 2014. Location-Based Social Networks. In *Recommender Systems for Location-based Social Networks*. Springer New York, New York, NY, 35–48. https://doi.org/10.1007/978-1-4939-0286-6_4 Series Title: SpringerBriefs in Electrical and Computer Engineering.

[16] Yuanwen Yue, Jiaqi Ding, Yuhao Kang, Yueyao Wang, Kunlin Wu, and Teng Fei. 2019. A location-based social network system integrating mobile augmented reality and user generated content. In *Proceedings of the 3rd ACM SIGSPATIAL International Workshop on Location-based Recommendations, Geosocial Networks and Geoadvertising*. ACM, Chicago Illinois, 1–4. https://doi.org/10.1145/3356994.3365507

[17] Yu Zheng and Xiaofang Zhou (Eds.). 2011. *Computing with Spatial Trajectories*. Springer New York, New York, NY. https://doi.org/10.1007/978-1-4614-1629-6