Augmented Reality Photo Presentation and Content-based Image Retrieval on Mobile Devices with *AR-Explorer*

Loris Sauter[1][0000-0001-8046-0362], Tim Bachmann[1][0000-0002-3928-4413], Heiko Schuldt[1][0000-0001-9865-6371], and Luca Rossetto[2][0000-0002-5389-9465]

University of Basel, Basel, Switzerland
{firstname}.{lastname}@unibas.ch
t.bachmann@unibas.ch
University of Zurich, Zurich, Switzerland
rossetto@ifi.uzh.ch

Abstract. Mobile devices are increasingly being used not only to take photos but also to display and present them to their users in an easily accessible and attractive way. Especially for spatially referenced objects, Augmented Reality (AR) offers new and innovative ways to show them in their actual, real-world context. In this demo, we present the *AR-Explorer* system, and particularly its use 'in-the-field' with local visual material. A demo video is available on Youtube: https://youtu.be/13f5vL9ZU3I.

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

1 Introduction

Multimedia collections are growing at an accelerating rate, especially due to the ubiquity of smartphones and their improved sensor readings in terms of high image quality, as well as orientation and location sensors. However, the browsing and searching capabilities along the spatio-temporal metadata are not widely available to users. Spatial metadata, in particular, can greatly enhance an image's communicative power. Coupled with the ever-increasing quality of Augmented Reality (AR) technologies, spatio-temporal metadata can be used to enhance the way we revisit images based on an on-the-spot AR presentation.

Based on prior research [11,12], we introduce *AR-Explorer*, a mobile AR-based browser for contemporary and historical images at their real-world location, enriched with content-based search. The system builds on the open-source content-based multimodal multimedia retrieval stack *vitrivr* [10] and is implemented as an iOS application. The major novelty is the ability to capture photos and re-experience them in their real-world location through an AR image presentation combined with content-based search. The presented work builds on preliminary work targeting Android, which did not include contemporary imagery [12]. Especially for cultural heritage, there is existing work that exploits

Fig. 1: Screenshots of the system in use with historical content (home screen on the left, search view on the right). In Basel, Switzerland, the depicted square used to be a parking lot in the 1930s. Nowadays, the old town part of the city is a pedestrian-only zone.

the capabilities of AR [4,6,9,8], see also a survey in [5], while recent work in this domain limits itself to traditional desktop applications [7].

The contribution of the paper is twofold: first, we present the architecture and methods of *AR-Explorer*, and second, we introduce a demonstration (together with an accompanying video) that showcases the capabilities of our system.

The remainder of this paper is structured as follows: in Section 2, we discuss the architecture of our system, and in Section 3, its implementation. The demonstration is presented in Section 4 and Section 6 concludes.

2 *AR-Explorer*

As capturing and presenting photos on smartphones is a well-integrated and well-established process, care must be given to an intuitive user experience when designing additional functionality. Challenges include the look and feel, query modes, and an immersive AR presentation mode. The *AR-Explorer* implementation is capable of capturing photos and annotating them with necessary metadata in order to recreate the exact same perspective in AR at a later time. Additionally, content-based retrieval methods provided by *vitrivr* [10], such as Query-by-Example, Query-by-Location, Query-by-TimePeriod and Query-by-Keyword, are available to users. The latter particularly employs deep learning-based visual text co-embedding to enable enhanced search [13]. On the data production side, a typical workflow is as follows: First, the user captures an image with the in-app camera. Then, the app sends the image and its metadata to the server that extracts content-based retrieval features. Finally, after a short while, the image is available for search. On the consumer side, typical interactions are as follows: First, the user starts the app in the map mode and uses one of the query modes to get results from the back-end. Then, either by using the gallery or the map view, results are presented. Finally, the AR presentation is started by either selecting one specific image or physically moving to such a location and starting the AR view by reacting to the notification.

When using *AR-Explorer*, users have multiple options for expressing their interests. The following provides an overview of the supported query modes which can be used individually or in combination:

Query-by-Location is the possibility to query for relevant content close to the user's current position or to an arbitrary location on a map.

Query-by-Time-Period is the ability to select all content that was created within a certain time interval.

Query-by-Example makes use of the content-based retrieval capabilities of the back-end by enabling the user to query using a picture taken by the mobile phone.

Query-by-Text makes use of deep-learning backed visual-text co-embedding [13] in order to textually search within the images.

3 Architecture

AR-Explorer consists of two parts: The iOS application as user interface and the back-end, which is an extension of the open-source content-based multimodal multimedia retrieval stack vitrivr,[1] as illustrated in the form of an architecture diagram in Figure 2. The communication to the back-end is facilitated through a REST API, of which we leverage the spatio-temporal query support [2]. The user interface consists of multiple views, corresponding to the user interaction shown in Figure 1 and a client for the retrieval tasks. Since iOS is the target platform, the app is written in Swift[2] and the AR functionality is provided by Apple's AR Kit.[3]

[1] https://vitrivr.org
[2] https://www.apple.com/swift/
[3] https://developer.apple.com/augmented-reality/

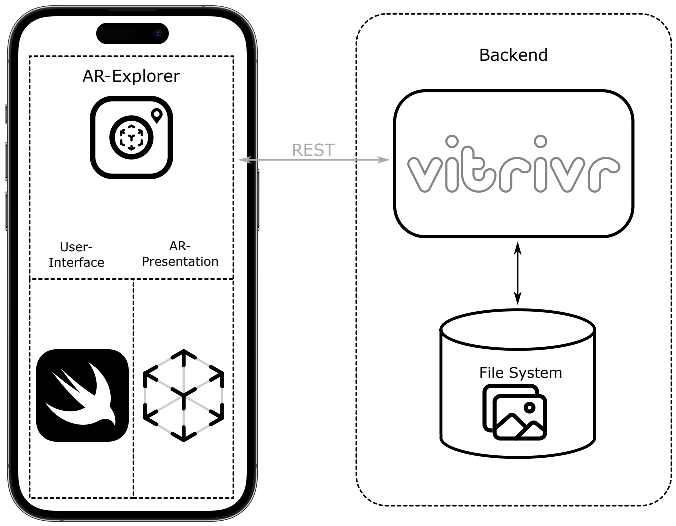

Fig. 2: The architecture diagram of the system. Communication with the backend is facilitated via REST. The iOS app builds on Swift and ARKit.

4 Demonstration

AR-Explorer enables users to re-experience contemporary and historic photo material at their real-world location using AR. Thus, the demonstration scenarios are highly location-dependent. Particularly, historical images covering cultural, architectural, and urban heritage cannot be transferred from one place to another. Hence, the demonstration consists of two parts, an online (local) and an offline part:

Offline: This part is covered by means of the accompanying video, which will also be shown at the demo booth at the conference: Using the aforementioned query capabilities and AR presentation mode, we demonstrate *AR-Explorer*'s application for cultural heritage with images of Basel, the largest city in north-western Switzerland, with a well-preserved medieval old town (and therefore many interesting historic photos – most of which come without any copyright restrictions and thus can be easily and safely shared and re-used). A video showing the application on location is available on Youtube via: https://youtu.be/13f5vL9ZU3I.

Online: During the conference, we will provide opt-in beta testing for iPhone or iPad users who want to try out *AR-Explorer*. We will take conference participants on a physical journey through contemporary (and, if possible, also historic) Amsterdam. Users will be able to experience the city (surroundings of the conference venue) on their own according to the process outlined earlier, from capturing a moment to re-experiencing city history at its real-world location.

5 Evaluation

We conducted an initial user study employing Brooke's System Usability Scale (SUS) [3]. The participants in the user studies during the summers of 2022 and 2023 numbered 20 and 8, respectively, and were primarily associated with the Department of Mathematics and Computer Science at the University of Basel. A range of iOS devices were utilized with minimal restrictions, with the oldest device being an iPhone 6. The System Usability Scale (SUS) is a ten-item survey designed to evaluate the usability of a system in a general manner, resulting in a score. In 2022, the mean SUS score was 77 with a standard deviation of 19, which is considered a *good* rating according to [1]. The more recent user study in 2023 resulted in a mean SUS score of 80 with a standard deviation of 13, which falls into the same *good* category as the 2022 study. We recognize the limited expressiveness of the SUS scores and regard the user studies as a brief assessment of this work.

6 Conclusion and Outlook

We have presented *AR-Explorer*, a prototype system consisting of an iOS app providing AR image presentation and capturing capabilities and content-based image search for Query-by-Location, Query-by-Time, Query-by-Example, and Query-by-Text through the well-established content-based multimedia retrieval stack vitrivr. Necessary metadata are captured with the in-app camera functionality, and retrieved images, contemporary or manually annotated historical ones, are presented at their real-world location in AR. Users are directed toward AR image presentation locations with the in-app AR navigation mode that displays routes toward points of interest in AR. When not actively using the app, notifications alert users about possible AR interactions. A preliminary study with users employing the System Usability Scale has yielded encouraging findings, with a SUS score that meets the criteria for being classified as *good*.

Limitations of the current prototype include minor stability issues with the frameworks in use, such as the AR framework causing a slight drift or jumping caused by inaccuracies in GPS localization. One approach to negate this issue would be to employ landmark detection in order to achieve better alignment and positioning of AR images.

Acknowledgements

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. Bangor, A., Kortum, P.T., Miller, J.T.: An Empirical Evaluation of the System Usability Scale. International Journal of Human-Computer Interaction **24**(6), 574–594 (Jul 2008). https://doi.org/10.1080/10447310802205776
2. Beck, L., Schuldt, H.: City-Stories: A Spatio-Temporal Mobile Multimedia Search System. In: Multimedia (ISM), 2016 IEEE International Symposium on. pp. 193–196. IEEE (2016)
3. Brooke, J.: SUS - A quick and dirty usability scale **57**, 187–191 (2013)
4. Cavallo, M., Rhodes, G.A., Forbes, A.G.: Riverwalk: Incorporating Historical Photographs in Public Outdoor Augmented Reality Experiences. In: Int'l Symposium on Mixed and Augmented Reality (IEEE ISMAR). pp. 160–165. Merida, Mexico (Sep 2016). https://doi.org/10.1109/ISMAR-Adjunct.2016.0068
5. Čopič Pucihar, K., Kljun, M.: ART for Art: Augmented Reality Taxonomy for Art and Cultural Heritage. In: V., G. (ed.) Augmented Reality Art, chap. 3. Springer (2018)
6. Fond, A., Berger, M., Simon, G.: Facade Proposals for Urban Augmented Reality. In: Int'l Symposium on Mixed and Augmented Reality (IEEE ISMAR). pp. 32–41. Nantes, France (Oct 2017). https://doi.org/10.1109/ISMAR.2017.20
7. Geniet, F., Gouet-Brunet, V., Brédif, M.: ALEGORIA: Joint Multimodal Search and Spatial Navigation into the Geographic Iconographic Heritage. In: Proceedings of the 30th ACM International Conference on Multimedia. pp. 6982–6984. ACM, Lisboa Portugal (Oct 2022). https://doi.org/10.1145/3503161.3547746
8. Lee, G.A., Dünser, A., Kim, S., Billinghurst, M.: CityViewAR: A Mobile Outdoor AR Application for City Visualization. In: Mixed and Augmented Reality (ISMAR-AMH), 2012 IEEE International Symposium on. pp. 57–64. IEEE (2012)
9. Lochrie, M., Čopič Pucihar, K., Gradinar, A., Coulton, P.: Designing Seamless Mobile Augmented Reality Location based Game Interfaces. In: Int'l Conference on Advances in Mobile Computing & Multimedia. p. 412. ACM (2013)
10. Rossetto, L., Giangreco, I., Tanase, C., Schuldt, H.: vitrivr: A Flexible Retrieval Stack Supporting Multiple Query Modes for Searching in Multimedia Collections. In: Proceedings of the 2016 ACM on Multimedia Conference. pp. 1183–1186. ACM (2016)
11. Sauter, L., Bachmann, T., Rossetto, L., Schuldt, H.: Spatially localised immsercive 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). *to be published by* ACM, Ottowa, ON, Canada (2023). https://doi.org/10.1145/3607542.3617358
12. Sauter, L., Rossetto, L., Schuldt, H.: Exploring Cultural Heritage in Augmented Reality with GoFind! In: 2018 IEEE International Conference on Artificial Intelligence and Virtual Reality (AIVR). pp. 187–188. IEEE, Taichung, Taiwan (Dec 2018). https://doi.org/10.1109/AIVR.2018.00041
13. Spiess, F., Gasser, R., Heller, S., Rossetto, L., Sauter, L., Schuldt, H.: Competitive Interactive Video Retrieval in Virtual Reality with vitrivr-VR. In: Lokoc, J., Skopal, T., Schoeffmann, K., Mezaris, V., Li, X., Vrochidis, S., Patras, I. (eds.) MultiMedia Modeling - 27th International Conference, MMM 2021, Prague, Czech Republic, June 22-24, 2021, Proceedings, Part II. Lecture Notes in Computer Science, vol. 12573, pp. 441–447. Springer (2021). https://doi.org/10.1007/978-3-030-67835-7_42