ARephotography: Revisiting Historical Photographs using Augmented Reality

Tommy Hasselman
University of Otago
Dunedin, New Zealand
tommyhasselman@gmail.com

Wei Hong Lo
University of Otago
Dunedin, New Zealand
weihong.lo@otago.ac.nz

Tobias Langlotz
University of Otago
Dunedin, New Zealand
tobias.langlotz@otago.ac.nz

Stefanie Zollmann
University of Otago
Dunedin, New Zealand
stefanie.zollmann@otago.ac.nz

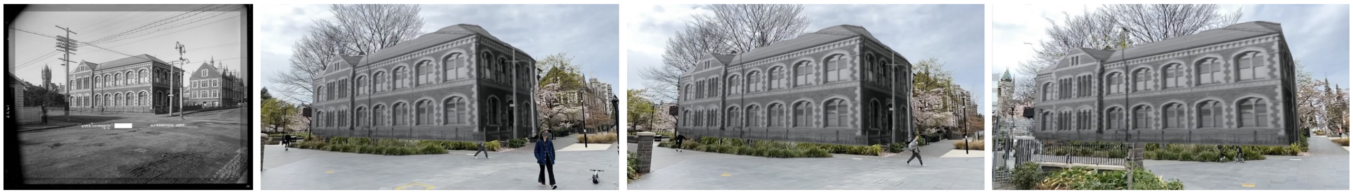


Figure 1: With ARephotography we combine the concepts of Rephotography and Augmented Reality (AR) to create experiences where one can view historical street views and buildings as they once looked. Our approach takes a historical photograph (Left) of a building and produces a textured 3D model that is visually overlaid over the current view of the building using AR (Middle Left, Middle Right and Right). Original image courtesy of Te Papa Tongarewa, reference C.012241.

ABSTRACT

Augmented Reality (AR) opens up new possibilities for interactive experiences which can be used in a variety of circumstances. Rephotography is a photo technique commonly presented on dedicated internet pages that align a past view with a current photo, allowing you to have a comparative view of the past with the present. This project aims to combine these two concepts to create AR experiences where you can view buildings and street views from historical photography seamlessly embedded in the present environment. We report on our automated pipeline that can take a historical photograph of a building and produces a textured 3D model that can be placed in AR over the current view of the building using techniques from machine learning while also reporting on first feedback from a preliminary user study.

CCS CONCEPTS

• **Human-centered computing** → **Mixed / augmented reality**; • **Computing methodologies** → **Computational photography**; **Image segmentation**.

KEYWORDS

Augmented Reality, Rephotography, Instance Segmentation, Inpainting, 3D reconstruction

ACM Reference Format:
Tommy Hasselman, Wei Hong Lo, Tobias Langlotz, and Stefanie Zollmann. 2023. ARephotography: Revisiting Historical Photographs using Augmented Reality. In *Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems (CHI EA '23), April 23–28, 2023, Hamburg, Germany.* ACM, New York, NY, USA, 7 pages. https://doi.org/10.1145/3544549.3585646

1 INTRODUCTION

In recent years, the potential of Augmented Reality (AR) in tourism [10] and cultural heritage [2] has been explored. Apart from sightseeing and navigation, there are also opportunities in using an AR interface to experience historical surroundings as they once looked when captured in a historical photograph. In this work, we explore an approach that allows for interactively browsing historic photographs using an AR interface.

The general concept of rephotography has a long tradition and is often used in applications where it is helpful to understand changes in landscapes, urban environments, or historical places over time. For example, there are applications comparing the current state of historical sites with views from the past by creating side-by-side visualisations, creating a now-and-then view. Our approach is inspired by the concept of Computational Rephotography proposed by Bae et al. [1] who automatically align historic photographs with recent photographs. This use of computer vision methods supports the previously pure manual alignment through the photographer. However, such computational rephotography methods do not allow

CHI EA '23, April 23–28, 2023, Hamburg, Germany
© 2023 Copyright held by the owner/author(s).
ACM ISBN 978-1-4503-9422-2/23/04.
https://doi.org/10.1145/3544549.3585646

users to explore and understand their surroundings, as the photographs are only aligned to a static photo representing a specific perspective. Users who often browse the result of the computational rephotography, e.g. on internet pages, are still required to spatially map now-and-then photographs into their field of view, a task that introduces a significant mental workload. Consequently, the use of rephotography on-site is often neglected and is more commonly seen on the Internet, books, and in museum installations.

In this work, we propose ARephotography, a concept that will bridge this gap by exploring how to directly visualize a now-and-then view in the users' field of view on a head-mounted display or mobile phone. By doing so, we create a new interface to historic photos and with new opportunities for how rephotography can be used, for example, in tourism, cultural heritage, and even on-site environmental monitoring. The challenges within this project consist of how to align historic photographic information with the current field of view of the users and provide them with an interface to explore the photograph. We also provide some first feedback on how our approach was perceived by users of the system.

2 RELATED WORK

In the past, there have been several approaches using AR to explore historical or tourist sites. For instance, with "Riverwalk", Cavallo et al. [3] presented an approach for superimposing 2D photographs onto the real world in an AR application. They used feature detection and tracking with a custom AR application to assist in a museum-like experience where historical images were superimposed on specific views in the users' surroundings. A similar application has been proposed by [16]. CityViewAR has been proposed by Lee et al. [11] which uses a variety of content to show users past views of Christchurch (NZ) city's downtown area prior to the 2011 earthquake. This uses the users' GPS position and sensors to determine camera location and offer related content at specific locations. Matviienko et al. [13] also overlaid historic photographs on smartphones and AR glasses.

The common issue with these approaches is the nature of the 2D visual content. Aligning a 2D photograph of a building with an existing structure in AR requires the user to have a specific point of view. Deviation from the original viewpoint would lead to misalignment in the real world. This creates limitations on the experience as movements are restricted. In this project, we aim to remove these limitations by superimposing 3D models rather than 2D images in AR. This allows for a wide range of viewing angles as the user is able to move around the 3D model.

There are other works that focus on extracting a 3D model from single photographs or videos. The approach proposed by Debevec et al. [5] starts by presenting the user with an interactive 3D modelling program. Here, the user aligns primitive volumes to edges in a single photograph. Their algorithm then generates a 3D model to align with the user's input, where the user can then refine the 3D model until it is satisfactory. Once the base 3D model is complete, Debevec et al. [5] then texture it by projecting the source image, or images, onto its surface. This approach generates good quality 3D models but relies heavily on user input.

Nishida et al. [15] propose a method to generate a procedural grammar from a single image that can then be used to reproduce the building as a 3D model. Their approach generates the 3D model by first determining the shape of the building and then the density and style of features such as windows. Although structurally the 3D model produced by this method is very good and would be suitable for use in an augmented or virtual environment, the approach neither support complex-shaped buildings nor preserve the original texture.

Other methods focus on reconstructing the surfaces visible in the source image. For instance, Horry et al. [9] propose *Tour Into the Picture*, an approach designed to generate a simple 3D model from a one-point-perspective photograph to allow novel views further "into" the photograph. They take user input to define the "rear wall" of a photograph along with its vanishing point to define five orthogonal planes.

Saxena et al. [18] estimated depth to recover 3D information rather than using geometric methods. They calculate super pixels and use a probabilistic approach to calculate the parameters of the plane they lie on and their depth in 3D space. This is then simplified with the assumption that the whole 3D model should contain a low number of planes.

Hoiem et al. [7] proposed a method of transforming a single image into a 3D model by "folding" it, based on the inferred structure in the image. They use a process of labelling regions in the image, defining lines that represent cuts and folds to be made, and then using these lines to transform the image into a 3D model (like in a pop-up book). Hoiem et al. [7]'s approach creates a simple 3D model which allows for moderate changes in perspective while preserving the original texture of the input image. However, this approach does not handle foreground content well nor does it allow for the integration into an AR view.

More recently Neural Radiance Fields (NeRF) have enabled major advancements in view synthesis [14]. While these approaches often require a larger set of input images, Yu et al. [22] proposed a method that allows the computation of a NeRF representation from few or even a single image. However, NeRF are still challenging to render in real-time on mobile devices. Thus they are currently not suitable for AR.

Up until now, the idea of combining AR and rephotography has not been explored. The use of single historic photographs as input for creating a "now-and-then" AR experience is lacking. So far, it is unclear whether the quality of such a 3D extraction is suitable for AR. Our work tries to bridge this gap by proposing an approach for creating such models and exploring the opportunities that arise from such an approach. While our approach builds on existing methods such as instance segmentation [21], image inpainting [19] and single view metrology [4], we combine them an in novel way and customise them to enable the combination of AR and rephotography possible.

3 AREPHOTOGRAPHY APPROACH

We combine the concepts of AR and Rephotography in a novel approach called "ARephotography". ARephotography takes an input photograph through a series of stages to produce the final 3D model. The ARephotography approach includes 1) a cleaning stage that detects and removes objects in the image, 2) an extraction stage masking out irrelevant content from the image, 3) transforming

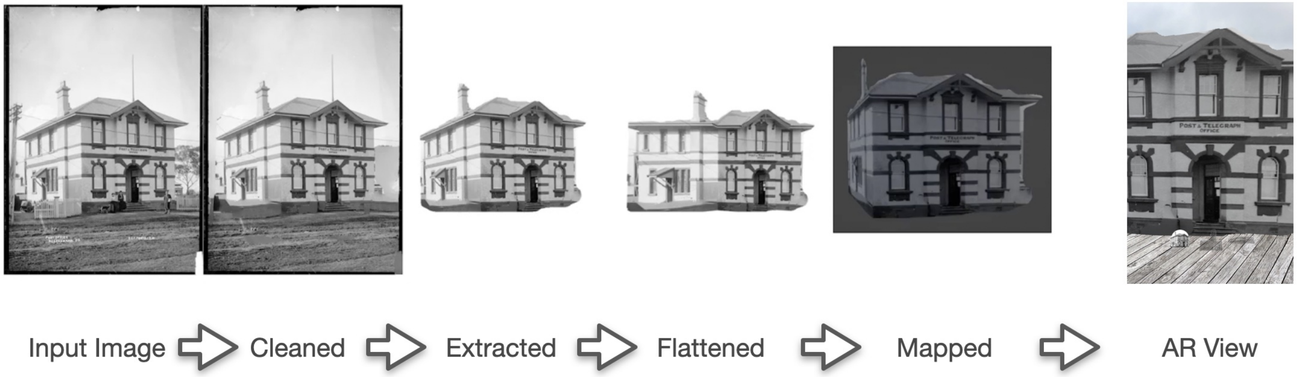


Figure 2: ARephotography aims to combine the concept of rephotography and Augmented Reality (AR) to create experiences where one can view historical buildings as they once looked. Our approach takes a historical photograph (Left) of a building and produces a textured 3D model that can be placed in AR over the current view of the building (Middle and Right). Original photograph: Green & Colebrook (Firm). Ngaruawahia Post Office and mail coach, 1910 - Photograph taken by G & C Ltd. Price, William Archer, 1866-1948 :Collection of post card negatives. Ref: 1/2-001602-G. Alexander Turnbull Library, Wellington, New Zealand. /records/23009835

relevant image parts into a flattened texture, 4) creating a 3D geometry with the flattened texture applied and finally 5) render this geometry in AR (Figure 2).

3.1 Cleaning

As the historic photograph may contain foreground objects that can occlude the building of interest, first, we produce a "cleaner" unobstructed view of the building. This stage uses a machine learning model that we trained with a custom dataset to identify foreground elements. We then use an inpainting algorithm to fill in areas of the image based on the surrounding pixels.

The machine learning model used is a convolutional neural network (CNN) [21] that we trained with our custom dataset of historic buildings photographs. The dataset was hand-labelled with objects we consider to be occluding objects (occluders). The custom dataset was used to fine-tune the model to the visual look of historic photography. We used transfer learning to fine-tune the model, as the dataset was not large enough to train a CNN from scratch. We used the open-source tool COCO Annotator[1] to label the dataset.

Once the occluders are detected, we use the segmentation data to produce a binary mask that indicates the areas to be cleaned from the image. Finally, we feed these masks to an inpainting algorithm [19][2] along with the source image to produce the cleaned version. Inpainting is a technique used to fill in areas of an image based on the surrounding pixels with a variety of approaches being available.

3.2 Segmentation

We then use the final cleaned version of the photograph to inform our 3D modeling approach to generate the final 3D model. The generated 3D model is then used in our AR application.

In Hoiem et al. [7]'s approach , the image was segmented into sky, vertical, and ground areas; with the vertical and ground areas both appearing in the final 3D model. However, in our case, we only require the vertical area, the building, since the ground area is not need for the AR application. We apply another CNN for the segmentation process [21], that we trained on a second custom dataset of historic photographs using transfer learning. We decided to train the network with a custom dataset here to account for the unique visual appearance of historical photographs and buildings. The custom dataset is made up of 51 photographs each with a single labelled instance of a building. Due to the small dataset size, we again utilize transfer learning for this model. The photographs we used were sourced by running the inpainting process of the training set for the cleaning stage, to train the model on photographs that had been altered with inpainting since this would be the input in practice.

Our customised segmentation step detects historic buildings well and in their entirety (Figure 2, Extracted). We use the segmentation step to create a mask that only contains pixel that cover the historic building.

3.3 3D Model Computation

After isolating the building from the rest of the image, we start to transform it into our final 3D model. This stage is based on the assumption that we are able to identify a set of vertices describing the building corners. It is possible to detect lines and calculate corners based on their intersection or we could train another machine learning model for estimating these points. In our approach, we implemented a user-guided step that asks the user to place six points marking a set of vertices on the building. At this stage, we have a simplified 3D model of the building represented by the six corner points. We then map the texture from the historic photograph onto

[1] https://github.com/jsbroks/coco-annotator
[2] https://github.com/saic-mdal/lama

this 3D model by using a homography that maps from the image plane to the simplified 3D building planes [6].

At this stage, the 3D model does have a predefined dimension. In order to render the 3D model in an AR application aligned to the real-world, we need to estimate the relative dimension of the building close to its real-world dimensions. For this purpose, we use a method inspired by classical drawing techniques for creating relative measurements in two-point-perspective illustrations. Using the six points we have retrieved, our approach reconstructs the geometry necessary for this method as described by MacEvoy [12]. We calculate measure points on the horizon line for each set of vanishing lines. We then cast a line from these points through the vertices of the building, the intersection creates a measuring line. The distance from the measuring lines anchor point on our building to the intersections gives the relative distances.

For photographs with reasonable levels of occlusion and a good view of the subject building, the pipeline performs well over both segmentation tasks and the generation of the 3D model. However, images with large amounts of occlusion and/or very poor viewing angles can be challenging for our approach.

The relative dimensions then provide us with a 3D model that is close to the real-world object up to a scale. We implemented an AR application that renders the historic building on top of the real-world view. In our AR application, we use a LIDAR scan to capture the real-world environment as it looks today and apply a point cloud registration step between the LIDAR scan point cloud and the extracted historic 3D model. For this purpose, the user has to select at least three reference points in both 3D models [8]. A more automated solution for this could be to apply the Iterative Closest Points (ICP) method [17] to align the 3D model with the real-world automatically. The input of the LIDAR scan can then be used for AR localisation[3] or can be integrated with marker tracking [20]. We then either render this as video-see-through AR on a mobile phone or as optical-see-through AR on AR glasses[4].

4 USER STUDY

We gathered feedback in a preliminary user study to get a better understanding of whether ARephotography would help users to explore historic photographs. To do this, we conducted a user study using three videos of historic imagery placed over a real-world location in 2D or AR (Figure 3). We use three different conditions: 1) a static 2D view that overlays the segmented building in the center of the screen (Figure 3, top), 2) a tracked 2D view that tracks the segmented 2D building to the building and 3) an AR view with the extracted 3D building aligned with the real world. By having the participants provide ratings for each of the views, we want to explore the following hypotheses: H1: The AR view with our 3D reconstruction is better in visual quality, realism, and coherency than the 2D alternatives. H2: The AR view with our 3D reconstruction provides a better understanding of the spatial relationship and alignment between the historic photograph and the real building than the 2D alternatives.

We conducted this study remotely and unsupervised by having participants respond to a Google Form on their own device using a link that we shared. This study was approved by the ethics committee of the University of Otago (D22/256).

4.1 Procedure

In the first part of the study, participants were required to answer a set of non-identifiable demographic questions. This covered age, gender, ethnicity, vision, and experience with AR. Following these, participants answered a set of five questions relating to our hypotheses for each of the three AR views. The AR views were displayed as an embedded YouTube video for the participant to watch prior to answering by giving ratings on a 1 to 7 Likert-like scale for each question. The precaptured videos were created by implementing each of the overlays (2D static, 2D Tracked and 3D Tracked) as AR lenses in Lensstudio[5] and displaying and precapturing them on a mobile phone.

After reviewing each AR view and giving their ratings, participants were asked for their opinion on three questions relating to content in AR and to note any extra thoughts they had for the project.

4.2 Participants

A total of 18 participants ranging from 22 to 65 years old (72.2% being under the age of 30) with 66% identifying themselves as female and 33% identifying as male participated in the study. No participants identified as gender diverse.

77.8% of participants reported that they have normal or corrected to normal vision. 55.6% of participants reported that they had not used AR before.

4.3 Results

We analyzed the results using non-parametric tests for their statistical analysis as we worked with Likert-like scales, first the Friedman test and then the Wilcoxon test (with a 2-tailed hypothesis) for direct comparison between AR views. To answer our first hypothesis, our study included three questions to collect ratings from participants on the visual quality, realism, and coherency of each AR view presented. The differences in participant ratings for all three of these questions were statistically significant when subjected to a Friedman test ($p < 0.05$). Specifically, the ratings for quality had a p-value of 0.00785, realism 0.00503, and coherency 0.00387. We then performed 2-tailed hypothesis Wilcoxon tests between each pair of AR views for the three questions (Table 1).

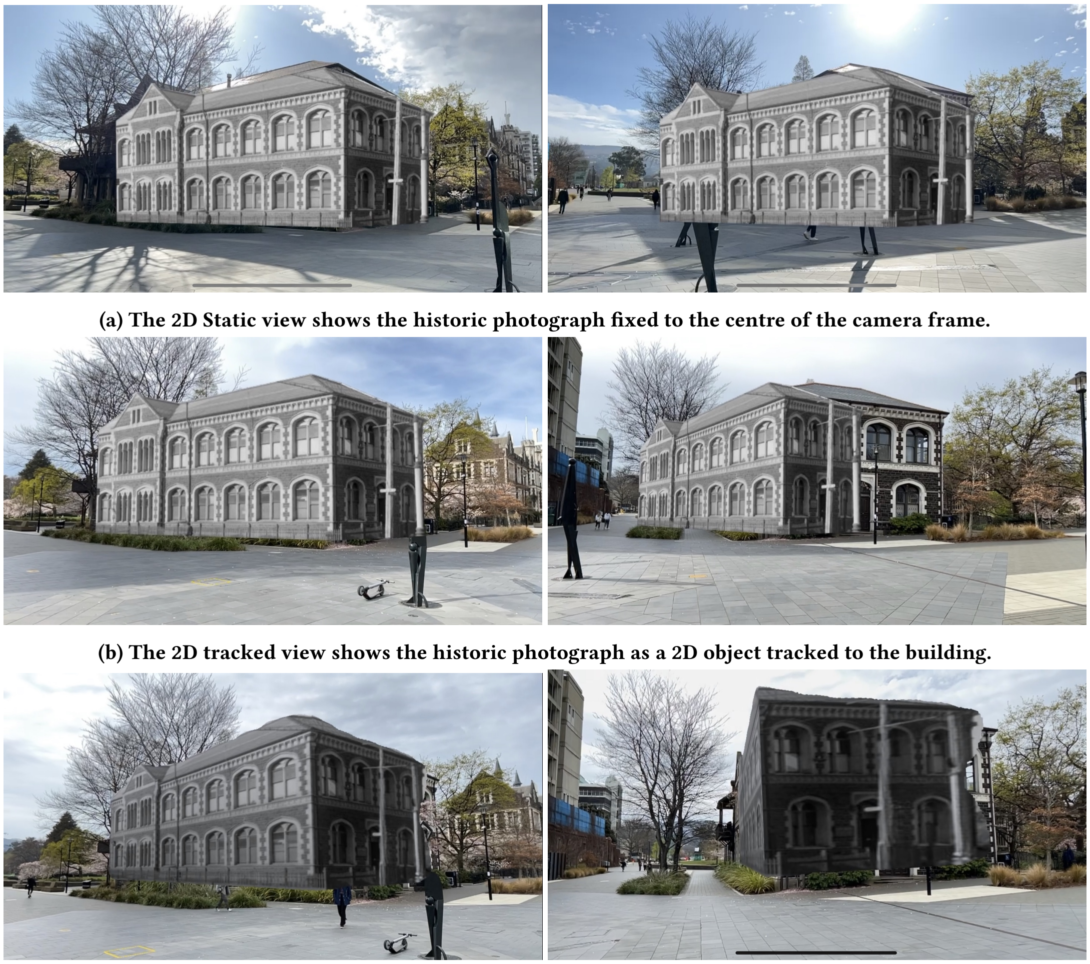
We found that the differences in ratings between the 2D Static AR view and the 3D Tracked AR view were statistically significant across all three visual attribute questions, with $p < 0.05$. We did not find statistically significant differences between 2D Static and 2D Tracked, and 2D Tracked with 3D Tracked.

For our second hypothesis, we asked participants to rate how well they understood the spatial relationship between the historic photograph and the real building in the AR views, as well as how well aligned they found the visual content (Figure 4). We found statistically significant differences when subjected to a Friedman test ($p < 0.05$, spatial understanding $p = 0.00371$, alignment $p =$

[3] https://docs.snap.com/lens-studio/references/templates/landmarker/custom-landmarker-scan

[4] https://www.spectacles.com

[5] https://ar.snap.com/lens-studio



(a) The 2D Static view shows the historic photograph fixed to the centre of the camera frame.

(b) The 2D tracked view shows the historic photograph as a 2D object tracked to the building.

(c) The 3D tracked view shows our 3D reconstruction tracked to the building.

Figure 3: Screenshots of the three visualizations used in our study. The left column shows the content when viewing the building at an angle, and the right column shows the content when viewing the building from the far left.

Attribute	2DS and 2DT	2DS and 3DT	2DT and 3DT
Quality	0.19706	0.00714(**)	0.05
Realism	0.09296	0.00328(**)	0.09296
Coherency	0.1031	0.0027(**)	0.08726
Spatial Relationship	0.5892	0.00148(**)	0.00672(**)
Alignment	0.13622	0.0003 (**)	0.00194(**)

Table 1: P-values for pairwise Wilcoxon calculation on different attribute ratings. 2DS = 2D Static view, 2DT = 2D Tracked view, 3DT = 3D Tracked AR view.

0.00009). Given that both questions had significance across the ratings we also performed the Wilcoxon tests between each of the AR view pairs (Table 1).

For both spatial questions, the differences in ratings between the 2D Static and 3D Tracked, and 2D Tracked and 3D Tracked were significant ($p < 0.05$, Table 1). The 3D-tracked AR view was generally rated higher than 2D Static and 2D Tracked by participants. We did not find statically significant differences between the 2D Static view and the 2D Tracked view.

At the end of the survey, we asked to provide additional feedback. The majority of participants who left extra comments on the content simply expressed their recognition of the potential for 3D content in historic AR applications. With some expressing how 3D provided better spatial understanding and more visual information compared to the other options. A few participants similarly comment that a full AR application could benefit from more context surrounding the visualisation, be this text-based or visual. Another participant commented how they preferred the 3D content but indicated that the preservation of the actual original is more important than the digital replica.

4.4 Discussion

With our user study, we found that there is a statistical significant difference in visual quality, realism, and coherency between simply

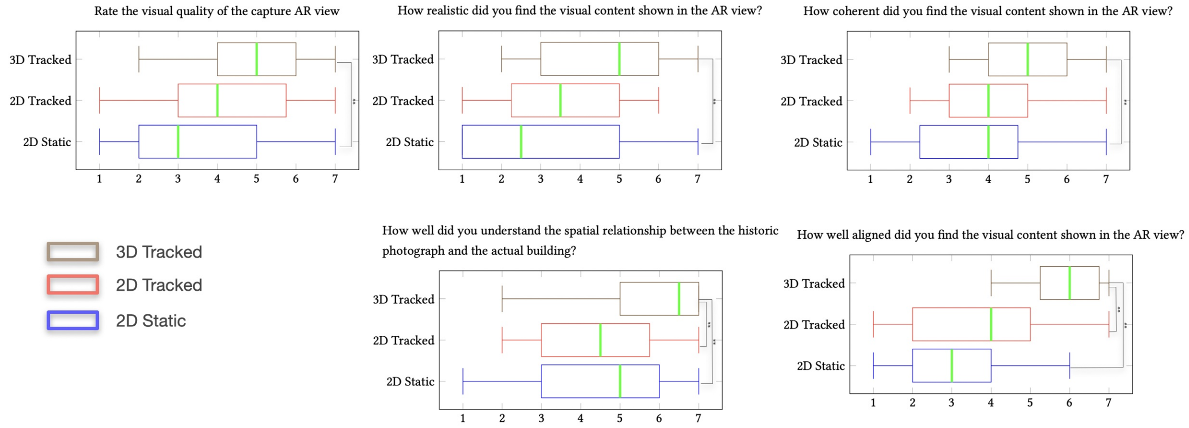


Figure 4: Results from rating different aspects of three different renderings conditions (2D Static, 2D Tracked and 3D Tracked (AR)) using Likert-like scales.

overlaying a 2D cutout of a building (2D) and the AR view (3D tracked) with the AR view being rated the highest. This partly confirms H1. However, no differences were measured between 2D tracked and 3D tracked. This might indicate that the participants found the visual attributes of quality, realism, and coherency to be comparable for both views that had tracked content. Ratings were also non-significant between both 2D views.

We were also able to confirm H2. The 3D-tracked AR view received higher ratings for spatial understanding and alignment with statistical significance compared to both 2D views. We did not find a difference between the 2D views. This indicates that the 3D-tracked content might be better suited to the exploration of historic spaces over 2D alternatives because of how it is able to support accurate motion and novel views as the user moves through the environment. The 3D-tracked AR view scored particularly higher than the alternatives on alignment (Figure 4).

Overall, our preliminary user study indicates the potential benefits of using the output from our ARephotography approach for AR exploration of historic areas over 2D alternatives. However, it is important to note that these findings are based on pre-captured results. More studies are needed to investigate whether the same findings can be confirmed in an on-site experiment. It is also important to mention that we did not further specify visual quality, realism and coherency in our questionnaires, and as such the interpretation of those ratings could be different for each participant. As this is a preliminary study to collect first feedback we kept the questions more general. However, future research could investigate these aspects in more detail.

5 CONCLUSION AND FUTURE WORK

In this paper, we propose the concept of ARephotography that puts historic photos back into their current context using an AR interface. We report on our automated pipeline that takes a single historic photograph of a building and cleans, isolates and meshes the visible facade into a 3D model. We also report on a preliminary user study that showed that using 3D content for the exploration of historic areas receives significantly higher ratings from users over similar AR experiences using 2D content with respect to coherence and realism but in particular for users' spatial understanding of how the historic photography relates to the building in real life and how well the content is aligned in the AR space.

For future work, we plan to completely automate the 3D reconstruction. We plan to do this using machine learning for the corner point extraction, however, this would likely require a large dataset to get accurate placement. We are also planning to extract finer geometry of the building's facades, like balconies and roof and develop a better integration of dynamic foreground objects occluding the historic building. Furthermore, we are also interested in exploring user interfaces for ARephotography in more depth investigating how users would interact with historic photographs in AR environments. This includes providing AR UI elements to switch between different points in time as well as the integration of customized visualization techniques.

In this, we have shown how modern machine learning and computer vision techniques can be used to produce historical content for AR applications, enabling a more immersive way for users to interactively explore historic areas in AR with content that is more realistic and provides a better spatial experience.

ACKNOWLEDGMENTS

Thanks to Snap Inc. for supplying hardware used for testing in this project.

REFERENCES

[1] Soonmin Bae, Aseem Agarwala, and Fredo Durand. 2010. Computational rephotography. *ACM Transactions on Graphics* 29, 3 (July 2010), 24:1–24:15. https://doi.org/10.1145/1805964.1805968

[2] George Caridakis and John Aliprantis. 2019. A Survey of Augmented Reality Applications in Cultural Heritage. *Int. J. Comput. Methods Herit. Sci.* 3, 2 (jul 2019), 118 – 147. https://doi.org/10.4018/IJCMHS.2019070107

[3] Marco Cavallo, Geoffrey Alan Rhodes, and Angus Graeme Forbes. 2016. Riverwalk: Incorporating Historical Photographs in Public Outdoor Augmented Reality Experiences. In *2016 IEEE International Symposium on Mixed and Augmented Reality (ISMAR-Adjunct)*. 160–165. https://doi.org/10.1109/ISMAR-Adjunct.2016.0068

[4] A. Criminisi, I. Reid, and A. Zisserman. 2000. Single View Metrology. *International Journal of Computer Vision* 40, 2 (Nov. 2000), 123–148. https://doi.org/10.1023/A:1026598000963

[5] Paul E. Debevec, Camillo J. Taylor, and Jitendra Malik. 1996. Modeling and rendering architecture from photographs: a hybrid geometry- and image-based approach. In *Proceedings of the 23rd annual conference on Computer graphics and interactive techniques (SIGGRAPH '96)*. Association for Computing Machinery, New York, NY, USA, 11–20. https://doi.org/10.1145/237170.237191

[6] Richard Hartley and Andrew Zisserman. 2004. *Multiple View Geometry in Computer Vision* (2 ed.). Cambridge University Press, Cambridge. https://doi.org/10.1017/CBO9780511811685

[7] Derek Hoiem, Alexei A. Efros, and Martial Hebert. 2005. Automatic Photo Pop-Up. In *ACM SIGGRAPH 2005 Papers* (Los Angeles, California) *(SIGGRAPH '05)*. Association for Computing Machinery, New York, NY, USA, 577 – 584. https://doi.org/10.1145/1186822.1073232

[8] BKP Horn. 1987. Closed-form solution of absolute orientation using unit quaternions. *JOSA A* 4, April (1987), 629–642. http://www.opticsinfobase.org/abstract.cfm?&id=2711http://www.opticsinfobase.org/abstract.cfm?&id=2711

[9] Youichi Horry, Ken Anjyo, and Kiyoshi Arai. 1997. Tour Into the Picture: Using Spidery Mesh Interface to Make Animation from a Single Image". 225–232. https://doi.org/10.1145/258734.258854

[10] Ibrahim Ilhan and Evrim Celtek. 2016. Mobile Marketing: Usage of Augmented Reality in Tourism. *Gaziantep University Journal of Social Sciences* 15, 2 (Dec. 2016), 581–599. https://doi.org/10.21547/jss.256721

[11] Gun A. Lee, Andreas Dünser, Seungwon Kim, and Mark Billinghurst. 2012. CityViewAR: A mobile outdoor AR application for city visualization. In *2012 IEEE International Symposium on Mixed and Augmented Reality - Arts, Media, and Humanities (ISMAR-AMH)*. 57–64. https://doi.org/10.1109/ISMAR-AMH.2012.6483989 ISSN: 2381-8360.

[12] Bruce MacEvoy. 2015. Two Point Perspective. In *Handprint Watercolors*. https://www.handprint.com/HP/WCL/perspect3.html

[13] Andrii Matviienko, Sebastian Günther, Sebastian Ritzenhofen, and Max Mühlhäuser. 2022. AR Sightseeing: Comparing Information Placements at Outdoor Historical Heritage Sites Using Augmented Reality. *Proc. ACM Hum.-Comput. Interact.* 6, MHCI, Article 194 (sep 2022), 17 pages. https://doi.org/10.1145/3546729

[14] Thomas Müller, Alex Evans, Christoph Schied, and Alexander Keller. 2022. Instant Neural Graphics Primitives with a Multiresolution Hash Encoding. *ACM Trans. Graph.* 41, 4, Article 102 (July 2022), 15 pages. https://doi.org/10.1145/3528223.3530127

[15] Gen Nishida, Adrien Bousseau, and Daniel Aliaga. 2018. Procedural Modeling of a Building from a Single Image. *Computer Graphics Forum* 37 (05 2018), 415–429. https://doi.org/10.1111/cgf.13372

[16] Kari Rainio, Petri Honkamaa, and Kaisa Spilling. 2015. *Presenting Historical Photos using Augmented Reality.* Technical Report. VTT Technical Research Centre of Finland.

[17] S. Rusinkiewicz and M. Levoy. 2001. Efficient variants of the ICP algorithm. In *Proceedings Third International Conference on 3-D Digital Imaging and Modeling*. 145–152. https://doi.org/10.1109/IM.2001.924423

[18] Ashutosh Saxena, Min Sun, and Andrew Y. Ng. 2009. Make3D: Learning 3D Scene Structure from a Single Still Image. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 31, 5 (May 2009), 824–840. https://doi.org/10.1109/TPAMI.2008.132

[19] Roman Suvorov, Elizaveta Logacheva, Anton Mashikhin, Anastasia Remizova, Arsenii Ashukha, Aleksei Silvestrov, Naejin Kong, Harshith Goka, Kiwoong Park, and Victor Lempitsky. 2021. Resolution-robust Large Mask Inpainting with Fourier Convolutions. *arXiv preprint arXiv:2109.07161* (2021).

[20] Daniel Wagner, Gerhard Reitmayr, Alessandro Mulloni, Tom Drummond, and Dieter Schmalstieg. 2008. Pose tracking from natural features on mobile phones. In *2008 7th IEEE/ACM International Symposium on Mixed and Augmented Reality*. 125–134. https://doi.org/10.1109/ISMAR.2008.4637338

[21] Yuxin Wu, Alexander Kirillov, Francisco Massa, Wan-Yen Lo, and Ross Girshick. 2019. Detectron2. https://github.com/facebookresearch/detectron2.

[22] Alex Yu, Vickie Ye, Matthew Tancik, and Angjoo Kanazawa. 2021. pixelNeRF: Neural Radiance Fields From One or Few Images. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. 4578–4587.