Accessible Computer Vision Tools For Social Augmented Reality

Anonymous
Anonymous
Anonymous, Anonymous
Anonymous

ABSTRACT

Computer Vision (CV) technology is widely used in Augmented Reality (AR) social media applications. Users take videos of themselves to message their friends with computer graphics (CG) augmenting their body and environment, all with minimal awareness of how the CV technology is running on their smartphone. As the technology matures, use cases are being explored that requires users to handle more complex interactions and decision-making with the underlying CV technology. This paper explores the question of how to design an accessible user experience (UX) and user interface (UI) for a CV scanning tool that allows novice users to navigate the complex computational processes that normally require an advanced CV degree to understand. Through two user studies, we refined our UI/UX to the point that testers were able to understand the workflow to make well-guided decisions on building effective CV data-sets for location-based AR experiences on their smartphone.

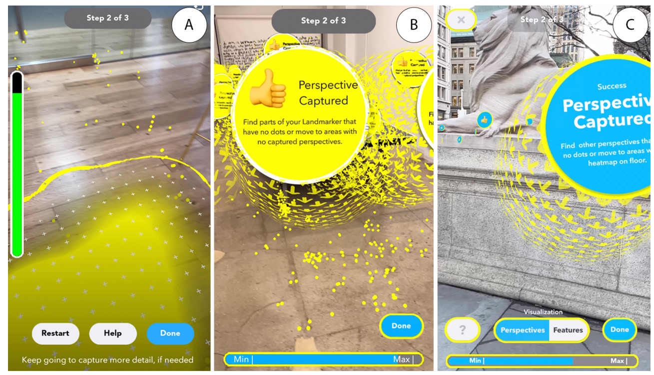

Figure 1: Incremental UI designs consisting of screen- and world-based UI elements for a smartphone CV scanning tool. A) first draft UI, B) second draft UI, C) final draft UI.

CCS CONCEPTS

• **Human-center computing** → **Augmented reality; accessibility technologies; visualization techniques; usability testing; interaction design theory, concepts and paradigms**;

KEYWORDS

Smartphone, augmented reality, computer vision, interaction design, user interface, 3D visualization, accessibility

ACM Reference Format:

Anonymous. 2022. Accessible Computer Vision Tools For Social Augmented Reality. In *Proceedings of In Special Interest Group on Computer Graphics and Interactive Techniques Conference Posters (SIGGRAPH'22 Posters).* ACM, New York, NY, USA, 2 pages. https://doi.org/10.1145/3450618.3469166

1 INTRODUCTION

The use of Augmented Reality (AR) in social media applications, such as anonymous social media platform, has become mainstream as users routinely take videos of themselves to message their friends that are augmented with computer graphics (CG) [4]. As use case exploration gets deeper into the computer vision (CV) tech stack, more robust user experience (UX) and supporting user interface (UI) designs are needed for users to handle the more complex interactions and cognitive ideas needed for them to process. This paper explores one such use case, how to allow anonymous social media platform users to build their own location-based AR experiences to share with their friends when they are in the same location.

The relevant CV concepts we will be covering in this paper are features and keyframes. A feature is a measurable piece of data in your image which is unique to this specific object [5]. It may be a distinct color in an image or a specific shape such as a line, edge, or an image segment. A good feature is used to distinguish objects from one another. Keyframe extraction is a primary step of a CV algorithm. The keyframe is the recorded location and orientation of the measuring device when registering a series of features [5].

When designing this system, we referenced relevant AR usability research. There have been several studies revealing attention impacting factors, such as attention tunneling [7], distracting visuals or backgrounds, and visual focal switching between screen and world-space elements [1]. Some studies have offered UX solutions to these problems that we incorporated in our system, such as minimizing text and graphics [2], keeping a user focused either in the world or on screen elements, and to use subtle cueing to guide the user's attention from one focal distance to another [6].

2 SYSTEM AND STIMULI DESIGN

The requirements for the CV scanning tool were based on the iPhone Pro's LiDAR sensor combined with anonymous social media's emerging technology for localizing AR content on user-generated CV data-sets. This meant the user flows needed to accommodate a LiDAR scanning phase for building 3D geometry, and a CV scanning phase for building a CV data-set. The 3D geometry is used for placing AR content and the CV data-set is used for localizing that AR content at run time. The creation process continues on-device for the user, but is beyond the scope of this system.

SIGGRAPH'22 Posters, August 8–11, 2022, Vancouver, CA

ACM ISBN 978-1-4503-8371-4/21/08.
https://doi.org/10.1145/3450618.3469166

The most difficult workflow of this system to make accessible for users was the CV data-set generation scan, which is why it is the focus of this paper. As shown in Figure 1, multiple iterations of the UI/UX were refined for this workflow as user testing revealed it needed the most work to achieve satisfactory accessibility.

The first draft UI shown in Figure 1A presented both CV elements, features and keyframes, as stimuli in the form of yellow dots and a heatmap on the floor respectively. The second draft UI shown in Figure 1B presented features as yellow dots and keyframes changed to "perspective captured" visuals positioned where the keyframes were recorded. The third draft UI show in Figure 1C minimized the stimuli and presented the features and keyframes in separate views along with minimizing "perspective captured" visuals when far from the user's device.

3 USER STUDY METHOD

To understand how to improve the usability of the CV scanning system, we conducted a user study with six participants from within the company. We choose the first six employee volunteers from our NYC office. They ranged in AR/CV understanding from none to a lot, due to their various job roles including business, operations, and engineering. Five out of six participants were within the target demographics of our social media audience, which were between 20-30 years of age, and having no to little AR/CV understanding.

3.1 Procedure

The researcher spent five minutes before each experiment stating participant's goals, and preparing the participant for the experiment. The experiment lasted 15 minutes, while the researcher observed the participant, writing notes, and helping them if they got stuck. The post interview lasted 20 minutes.

The participants' goal was to select a permanent feature of a public space to scan so that later on they can add messages for their friends to see when they are at that location. The participant was only responsible for understanding how to scan the location and attempt to understand the quality of their geometry and CV scanning data. They were not required to create a message for their friends as that is outside the scope of the system and study. Participants were then taught how to conduct a cognitive walkthrough, which involves them verbalizing their thoughts, feelings, and intentions as they use a piece of software to achieve a goal.

3.2 Measurements

After each experiment, the researcher conducted a semi-structured interview. There were a total of 10 questions asked intended to better understand how different aspects of the scanning process UI/UX impacted usability. A sample of our questions include:

(1) What did you think about the scanning process?
(2) What did you think about the directions given to you?
(3) What did you think about the user interface?
(4) What did you think about the 3D graphics shown to you?
(5) What parts of the experience could be clearer to you?
(6) How would you improve the experience?

A three-stage coding process was used as described by Cambell et al., for the measurement of intercoder reliability for semi-structured interviews [3]. The interview lead read through all interview transcripts and generated a list of codes. A second researcher reviewed 10% of the interview transcripts and reached a minimum of 87% intercoder reliability with the interview lead, which was 89% intercoder reliability. Once completed, both researchers coded the remaining open-ended responses and counted code occurrences.

Some code samples generated from the interview questions:

(1) Attention Discrepancies Between World and Screen UI elements: Participants reported not paying attention to screen-based UI when world-based 3D graphics were present.
(2) Poor Interpretation of 3D Visuals: Participants reported misunderstanding of the 3D visuals for features and keyframes.
(3) Poor Understanding of CV Terminology: Participants reported not understanding technical terms as there was no context to their normal experiences.

4 DESIGN ITERATION

Based on participant feedback from the user study, we iterated on our UI/UX designs. We updated the UI from first draft in Figure 1A to the second draft in Figure 1B. The heatmap representing the location of the keyframes was not conveying the correct interpretation, and was blending too much with the yellow dots of the features. In addition, the verbiage needed to be refined from unfamiliar CV terminology to concepts that were more familiar to novice CV/AR users. The idea of capturing perspectives was discussed and seemed to work better with the ideas already in the testers' minds. After updating the UI/UX, a second user study was conducted to confirm that the CV workflows were more accessible and a last round of polishing lead the final version of the UI shown in Figure 1C.

5 CONCLUSION

This paper explored how to design an accessible UI/UX for a CV scanning tool that allows novice users to navigate complex computational processes that normally require an advanced CV degree to understand. Through two user studies, we refined the UI/UX to the point that testers were able to understand the workflows enough to make well-guided decisions on building effective CV data-sets for location-based AR experiences on their smartphone.

REFERENCES

[1] Mohammed Arefin, Nate Phillips, Alexander Plopski, and J. Edward Swan. 2021. Effects of a Distracting Background and Focal Switching Distance in an Augmented Reality System. In *IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct)*. IEEE ISMAR, 96–99.

[2] Amit Barde, Matt Ward, Robert Lindeman, and Mark Billinghurst. 2019. Less is More: Using Spatialized Auditory and Visual Cues for Target Acquisition in a Real-World Search Task. In *IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct)*. IEEE ISMAR, 340–341.

[3] John Campbell, Charles Quincy, Jordan Osserman, and Ove Pedersen. 2013. Coding in-depth semistructured interviews: Problems of unitization and intercoder reliability and agreement. *Sociological Methods and Research* 42, 3 (2013), 294–320.

[4] Matthew Donaruma. accessed on 4/26/2022, published 2021. Here's Why Augmented Reality Matters. *https://opticskypro.com/updates/social-media-marketers-heres-why-augmented-reality-matters/* (accessed on 4/26/2022, published 2021).

[5] Mohamed Elgendy. 2020. *Deep Learning for Vision Systems.* Manning, New York, NY, USA.

[6] Weiquan Lu, Duh Been-Lim, and Steven Feiner. 2012. Subtle Cueing for Visual Search in Augmented Reality. In *IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. IEEE ISMAR, 161–166.

[7] Brandon Syiem, Kelly Ryan, Jorge Gonclaves, Eduardo Velloso, and Tilman Dingler. 2021. Impact of Task on Attentional Tunneling in Handheld Augmented Reality. In *CHI Conference on Human Factors in Computing Systems (CHI '21)*. SIGCHI, 1–14.