Demo Abstract: Edge-assisted Collaborative Image Recognition for Augmented Reality

Jovan Stojkovic
University of Belgrade
sj160151d@student.etf.bg.ac.rs

Zida Liu
Duke University
zida.liu@duke.edu

Guohao Lan
Duke University
guohao.lan@duke.edu

Carlee Joe-Wong
Carnegie Mellon University
cjoewong@andrew.cmu.edu

Maria Gorlatova
Duke University
maria.gorlatova@duke.edu

ABSTRACT

Mobile Augmented Reality (AR), which overlays digital information with real-world scenes surrounding a user, provides an enhanced mode of interaction with the ambient world. Contextual AR applications rely on image recognition to identify objects in the view of the mobile device. In practice, due to image distortions and device resource constraints, achieving high performance image recognition for AR is challenging. Recent advances in edge computing offer opportunities for designing *collaborative* image recognition frameworks for AR. In this demonstration, we present CollabAR, an edge-assisted collaborative image recognition framework. CollabAR allows AR devices that are facing the same scene to collaborate on the recognition task. Demo participants develop an intuition for different image distortions and their impact on image recognition accuracy. We showcase how heterogeneous images taken by different users can be aggregated to improve recognition accuracy and provide a better user experience in AR.

CCS CONCEPTS

• **Computing methodologies → Distributed algorithms**; **Mixed / augmented reality**;

KEYWORDS

Edge computing, collaborative augmented reality, image recognition.

ACM Reference format:
Jovan Stojkovic, Zida Liu, Guohao Lan, Carlee Joe-Wong, and Maria Gorlatova. 2019. Demo Abstract: Edge-assisted Collaborative Image Recognition for Augmented Reality. In *Proceedings of The 17th ACM Conference on Embedded Networked Sensor Systems, New York, NY, USA, November 10–13, 2019 (SenSys '19),* 2 pages.
https://doi.org/10.1145/3356250.3361944

1 INTRODUCTION

Mobile augmented reality (AR), which overlays digital content with the real world around a user, has recently jumped from the pages of science fiction novels into the hands of consumers [1]. Contextual AR applications require an effective image recognition system that can identify objects in the camera view of the mobile device [4, 5]. However, building a robust image recognition system for AR is challenging. In real-world conditions, many images taken by mobile devices contain distortions that reduce recognition accuracy [2]. For instance, images taken by the smartphone or the head-mounted AR set camera frequently contain motion blur caused by the user's motion [4], while Gaussian white noise often appears in images taken in poor illumination conditions [2]. Although current advancements in deep neural networks (DNN) have shown outstanding performance in image recognition, DNNs are vulnerable to these distortions [2]. Moreover, due to the high computational complexity, DNNs encounter high latency when running on resource-constrained mobile AR devices. Achieving robust high-performance image recognition for AR "in the wild" is currently an open problem.

Due to recent advances in AR development platforms, AR applications have become accessible on a wide range of mobile devices. For instance, Google ARCore[1] and Apple ARKit[2] allow developers to build AR experiences for both Android-based and iOS-based mobile devices. Pervasive deployment of AR will offer numerous opportunities for *collaborative AR* among the users [5]. In addition, the advances in communication technologies are bringing edge computing into reality. Edge-connected mobile devices are no longer limited to operating in a standalone way. Instead, mobile devices that have the same recognition task can cooperate to overcome image distortions.

In this demo, we present CollabAR, an edge-assisted Collaborative image recognition framework for Augmented Reality. An illustration of CollabAR is shown in Figure 1. While in conventional image recognition systems individual devices complete recognition tasks independently [2, 4], in CollabAR AR devices that have the same recognition task can collaborate to improve recognition accuracy. The underlying principle of CollabAR is that, although different devices acquire images from different positions (angle, distance) and the images are heterogeneous in quality (image resolution, distortions), there are enough spatial and temporal correlations in the images to be able to appropriately fuse the results and improve recognition accuracy [3]. By aggregating images from multiple users, CollabAR can provide a more accurate result with a lower latency.

SenSys '19, November 10–13, 2019, New York, NY, USA

ACM ISBN 978-1-4503-6950-3/19/11…$15.00
https://doi.org/10.1145/3356250.3361944

[1] ARCore: https://developers.google.com/ar/
[2] ARKit: https://developer.apple.com/documentation/arkit

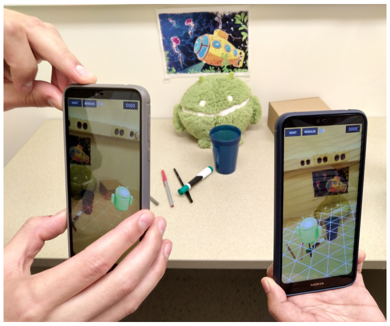

Figure 1: CollabAR in action: two mobile users collaborate to recognize the object in their shared view.

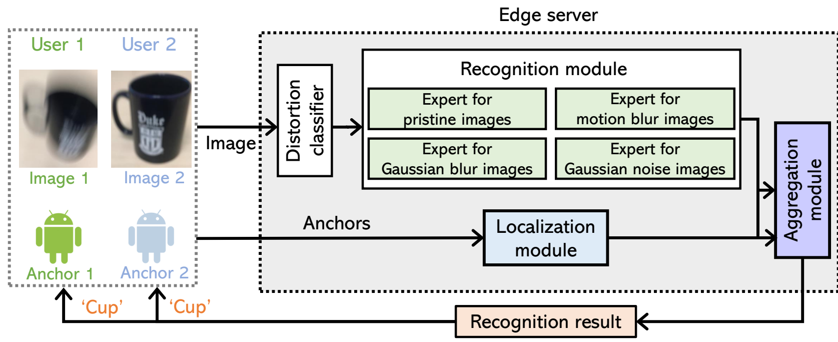

Figure 2: The overview of CollabAR. The users send images to the edge server for image recognition. The server detects the distortion in the image and applies corresponding recognition expert for the recognition. Images that are taken by users facing the same object are aggregated to improve the overall accuracy.

2 SYSTEM DESIGN

Figure 2 shows the overview of CollabAR, which aggregates inputs from multiple users to perform collaborative image recognition on the edge. The users take images of the object using Android phones, and send the images to the edge server along with associated anchors. The server contains four modules: (1) image *distortion classifier*, (2) *recognition module*, (3) *localization module*, and (4) multi-user *aggregation module*.

The *distortion classifier* detects what type of distortion is contained in the images. In addition to pristine (non-noisy) images, we consider three types of distortions that commonly appear in mobile AR scenarios: motion blur, Gaussian blur, and Gaussian noise. The *recognition module* includes four different *recognition experts*. Each of the experts is a Convolutional Neural Network (CNN) trained to recognize images with a particular type of distortion. Based on the output of the distortion classifier, the edge server selects a dedicated expert for the image recognition. The *localization module* determines whether different users are looking at the same scene. We use Cloud anchors provided by Google ARCore to identify the location and orientation of the AR devices in the physical space. In our design, the anchors are created and managed by the administrator in advance. The created anchors can be resolved and shared between multiple users if the users are in the same physical space. When sending images to the server, users also automatically upload the anchors in their view to the server. The localization module compares the sets of anchors that are uploaded by different users. For any two users, if some anchors appear in the views of both users, the server will consider these users as looking at the same scene. Finally, the *aggregation module* combines the recognition results of different users if they are looking at the same scene.

3 INTERACTIVE DEMONSTRATION

The demonstration follows the workflow shown in Figure 2. It allows the participants to develop an intuition for different image distortions and their impacts on image recognition accuracy. We also showcase how heterogeneous images (i.e., with different distortions) taken by multiple users can be aggregated to improve the accuracy. An annotated video of the demo is available online.[3]

[3]Link to the demo video: https://youtu.be/RFCxe9ZAVQw

Similar to the illustration in Figure 1, the participants interact with the demo by taking images of different objects (we use six types of objects: cups, bottles, pens, books, smartphones, and bags) using the client application running on Android mobile phones. The participants can select two operating modes when running the demo, single-user and collaborative. In the single-user mode, the recognition result of the user is not aggregated with the others. In the collaborative mode, outputs of multiple users are fused by the aggregation module to generate the final result. The results received from the server are displayed on the client smartphone in two ways. First, the original Cloud anchor on the client screen changes into a colorful holographic 3D model of the recognized object. Second, the distortion type of the image and the recognition results of the single-user and the collaborative modes are shown. The participants can observe how image distortion affects the recognition accuracy of the single-user mode, and how the use of the collaborative mode improves the accuracy. Compared to the single-user mode, the participants experience a significant accuracy improvement (from 60% to 98%) when using the collaborative mode.

ACKNOWLEDGMENTS

This work was supported in part by the NSF award CSR-1903136 and by the Defense Advanced Research Projects Agency under contract No. HR001117C0048. The opinions, findings and conclusions expressed in this material are those of the author(s) and do not necessarily reflect the views of the Defense Advanced Research Projects Agency.

REFERENCES

[1] Dimitris Chatzopoulos, Carlos Bermejo, Zhanpeng Huang, and Pan Hui. 2017. Mobile augmented reality survey: From where we are to where we go. *IEEE Access* 5 (2017), 6917–6950.

[2] Samuel F Dodge and Lina J Karam. 2018. Quality robust mixtures of deep neural networks. *IEEE Transactions on Image Processing* 27, 11 (2018), 5553–5562.

[3] Peizhen Guo, Bo Hu, Rui Li, and Wenjun Hu. 2018. FoggyCache: Cross-device approximate computation reuse. In *Proceedings of MobiCom*. ACM, 19–34.

[4] Puneet Jain, Justin Manweiler, and Romit Roy Choudhury. 2015. OverLay: Practical mobile augmented reality. In *Proceedings of MobiSys*. ACM, 331–344.

[5] Wenxiao Zhang, Bo Han, Pan Hui, Vijay Gopalakrishnan, Eric Zavesky, and Feng Qian. 2018. CARS: Collaborative augmented reality for socialization. In *Proceedings of HotMobile*. ACM, 25–30.