Poster: BystandAR: Protecting Bystander Visual Data in Augmented Reality Systems

Matthew Corbett
Virginia Tech
Blacksburg, Virginia, USA
matthewc84@vt.edu

Brendan David-John
Virginia Tech
Blacksburg, Virginia, USA
bmdj@vt.edu

Jiacheng Shang
Montclair State University
Montclair, New Jersey, USA
shangj@montclair.edu

Y. Charlie Hu
Purdue University
West Lafayette, Indiana, USA
ychu@purdue.edu

Bo Ji
Virginia Tech
Blacksburg, Virginia, USA
boji@vt.edu

Abstract

Augmented Reality (AR) devices are set apart from other mobile devices by the immersive experience they offer. While the powerful suite of sensors on modern AR devices is necessary for enabling such an immersive experience, they can create unease in bystanders (i.e., those surrounding the device during its use) due to potential bystander data leaks, which is called the bystander privacy problem. In this poster, we propose BystandAR, the first practical system that can effectively protect bystander visual (camera and depth) data in real-time with only on-device processing. BystandAR builds on a key insight that the device user's eye gaze and voice are highly effective indicators for subject/bystander detection in interpersonal interaction, and leverages novel AR capabilities such as eye gaze tracking, wearer-focused microphone, and spatial awareness to achieve a usable frame rate without offloading sensitive information. Through a 16-participant user study, we show that BystandAR correctly identifies and protects 98.14% of bystanders while allowing access to 96.27% of subjects. We accomplish this with average frame rates of 52.6 frames per second without the need to offload unprotected bystander data to another device.

CCS Concepts

• **Security and privacy** → **Domain-specific security and privacy architectures**; **Privacy protections**; • **Human-centered computing** → **Mobile devices**.

Keywords

bystander privacy, visual data, augmented reality, eye tracking

ACM Reference Format:
Matthew Corbett, Brendan David-John, Jiacheng Shang, Y. Charlie Hu, and Bo Ji. 2023. Poster: BystandAR: Protecting Bystander Visual Data in Augmented Reality Systems . In *The 21st Annual International Conference on Mobile Systems, Applications and Services (MobiSys '23), June 18–22, 2023, Helsinki, Finland.* ACM, New York, NY, USA, 2 pages. https://doi.org/10.1145/3581791.3597377

MobiSys '23, June 18–22, 2023, Helsinki, Finland

ACM ISBN 979-8-4007-0110-8/23/06.
https://doi.org/10.1145/3581791.3597377

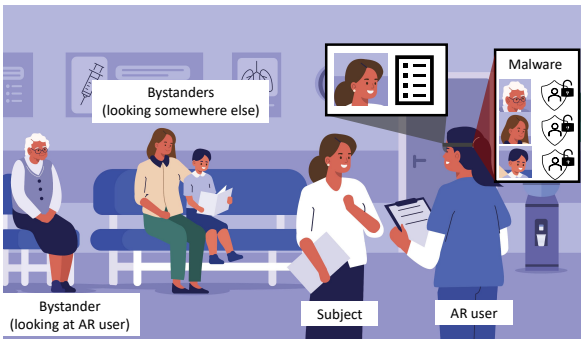

Figure 1: An illustration of the medical use case of AR, where a nurse wearing an AR device is interacting with a patient while there are bystanders present (watching or not watching the nurse). In this situation, while the patient's medical record information needs to be presented to the nurse via the AR device, bystander information must be protected.

1 Introduction

Augmented Reality (AR) applications rely on the unique capabilities of AR devices, namely the ability to understand the physical world, and seamlessly blend the physical world and the holographic, digital world. This ability to create a virtual mapping of a physical space through Simultaneous Localization and Mapping (SLAM), establish synthetic holographic contact, and sense user eye gaze and hand gestures, is made possible by the integrated and powerful suite of sensors on modern AR devices. Such sensors, while essential to the immersive experience that makes AR devices unique and powerful, do not discriminate in the data they collect. AR devices capture data required for well-intentioned tasks (e.g., SLAM, pose estimation, and gesture recognition), but also capture visual (e.g., camera and depth) data about bystanders (i.e., persons surrounding the device during its use), which can potentially be used to identify sensitive information (age, gender, emotion, gait, etc.) of bystanders for malicious purposes [1–3, 6–8]. This threat of bystander data leak is called the *bystander privacy problem or BPP* [4, 5, 9].

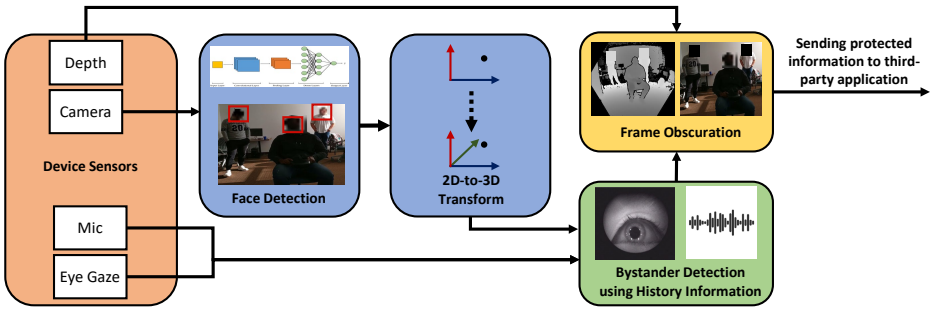

Figure 2: BystandAR Architecture. Raw data is captured from the device's sensors and is used both in face detection and learning eye gaze and voice history information for bystander detection. Afterward, bystander detection is used to obscure human faces not designated subjects in both camera and depth frame data.

2 BystandAR Design

BystandAR is designed to prevent malicious AR applications running on AR devices from collecting sensitive information from visual data of bystanders of interpersonal interactions by exploiting the BPP. Fig. 2 shows the architecture of our proposed BystandAR system. The camera and depth frames are continuously captured by the AR device camera. At a given sampling interval, the face detection module infers the 2D location of any faces present in the frame, and BystandAR locates these faces in 3D after 2D-to-3D transformation. Using this location, we create a 3D bounding box, invisible to the user, that serves as the 3D anchor for each detection. By default, these faces are labeled bystanders. As sampled face detection continues and the position of the face changes, BystandAR updates the location of the face and moves the 3D bounding box accordingly.

In parallel with the above face detection and tracking process, BystandAR collects information about the user's eye gaze and voice using the AR device's onboard eye gaze tracking and wearer-focused microphone. For every camera frame, BystandAR tracks on which face the user's attention is currently focused on and maintains a history of this information for all currently detected faces. Once the history of the user's attention (eye gaze or simultaneous eye gaze and voice input) meets a pre-specified threshold, the detection is labeled a subject. With this context, the face obscuration module obscures the faces of each detection as required. After bystander visual data has been removed from each frame, the frame is safe for release to any third-party application.

3 Evaluation

Through an evaluation involving 16 participants, BystandAR was successful in protecting 98.14% of bystander faces through obscuration and in identifying the subject of an AR interaction in 96.27% of output frames. This ensures that the visual data of identified subjects remain available for legitimate uses. Our evaluation also shows an improvement in bystander protection by 12% over the most accurate existing solution and shows a marked increase in bystander perceptions of privacy. These improvements are gained while keeping bystander data *on-device*, removing the need to offload unprotected bystander data to another device, and maintaining frame rates as high as 52.6 frames per second (FPS).

4 Conclusion

In this work, we harnessed the dynamics of human interaction to improve bystander visual data protection in AR devices by creating a novel system called BystandAR. This is achieved *on-device* while maintaining usable frame rates on AR devices. We believe that this work expands the understanding of the capability of modern AR devices to protect bystander privacy and to further the trust of bystanders that their privacy is protected, using unique capabilities that only these exciting, advanced AR devices possess. This is the poster abstract for the full paper titled "BystandAR: Protecting Bystander Visual Data in Augmented Reality Systems". Please see the full paper (doi: 10.1145/3581791.3596830) for more details.

Acknowledgment This work is supported in part by the Commonwealth Cyber Initiative (CCI) and the NSF grants under CNS 2112778 and 2153397.

References

[1] Ricard Borràs, Àgata Lapedriza, and Laura Igual. 2012. Depth Information in Human Gait Analysis: An Experimental Study on Gender Recognition. In *Image Analysis and Recognition*, Aurélio Campilho and Mohamed Kamel (Eds.). Springer Berlin Heidelberg, Berlin, Heidelberg, 98–105.

[2] Zijun Cheng, Tianwei Shi, Wenhua Cui, Yunqi Dong, and Xuehan Fang. 2017. 3D face recognition based on kinect depth data. In *2017 4th International Conference on Systems and Informatics (ICSAI)*. IEEE, 445 Hoes Lane, Piscataway, NJ 08854., 555–559. https://doi.org/10.1109/ICSAI.2017.8248353

[3] Edward Chou, Matthew Tan, Cherry Zou, Michelle Guo, Albert Haque, Arnold Milstein, and Li Fei-Fei. 2018. Privacy-Preserving Action Recognition for Smart Hospitals using Low-Resolution Depth Images. *CoRR* abs/1811.09950 (2018). arXiv:1811.09950

[4] Brendan David-John, Diane Hosfelt, Kevin Butler, and Eakta Jain. 2021. Let's SOUP up XR: Collected thoughts from an IEEE VR workshop on privacy in mixed reality. In *VR4Sec: Security for VR and VR for Security, SOUPS 2021 Workshop*.

[5] Jaybie A. De Guzman, Kanchana Thilakarathna, and Aruna Seneviratne. 2019. Security and Privacy Approaches in Mixed Reality: A Literature Survey. *ACM Comput. Surv.* 52, 6, Article 110 (oct 2019), 37 pages. https://doi.org/10.1145/3359626

[6] Sarwesh Giri, Gurchetan Singh, Babul Kumar, Mehakpreet Singh, Deepanker Vashisht, Sonu Sharma, and Prince Jain. 2022. Emotion Detection with Facial Feature Recognition Using CNN and OpenCV. In *2022 2nd International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE)*. IEEE, 445 Hoes Lane, Piscataway, NJ 08854., 230–232. https://doi.org/10.1109/ICACITE53722.2022.9823786

[7] Ahmad Jalal, Shaharyar Kamal, and Daijin Kim. 2016. Human Depth Sensors-Based Activity Recognition Using Spatiotemporal Features and Hidden Markov Model for Smart Environments. *Journal of Computer Networks and Communications* 2016 (04 Oct 2016), 8087545. https://doi.org/10.1155/2016/8087545

[8] Pavan Kunchala. 2021. Real-time age gender detection using opencv. https://medium.com/analytics-vidhya/real-time-age-gender-detection-using-opencv-fa705fe0e1fa

[9] Alfredo J Perez, Sherali Zeadally, and Scott Griffith. 2017. Bystanders' privacy. *IT Professional* 19, 3 (2017), 61–65.