Towards Real-time Computer Vision and Augmented Reality to Support Low Vision Sports: A Demonstration of ARTennis

Jaewook Lee1  Devesh P. Sarda1,2  Eujean Lee1  Amy Seunghyun Lee3  
Jun Wang1  Adrian Rodriguez1  Jon E. Froehlich1

1University of Washington, Seattle, USA  2University of Wisconsin-Madison, Madison, USA  3University of California, Davis, USA

ABSTRACT

Individuals with low vision (LV) can experience vision-related challenges when participating in sports, especially those with fast-moving objects. We introduce ARTennis, a prototype for wearable augmented reality (AR) that utilizes real-time computer vision (CV) to enhance the visual saliency of tennis balls. Preliminary findings indicate that while ARTennis is helpful, combining both visual and auditory cues may be more effective. As AR and CV technologies continue to improve, we expect head-worn AR to broaden the inclusivity of sports.

CCS CONCEPTS
• Human-centered computing → Accessibility systems and tools; Mixed / augmented reality.

KEYWORDS
augmented reality, accessibility, sports, computer vision

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).

UIT ’23 Adjunct, October 29-November 1, 2023, San Francisco, CA, USA
© 2023 Copyright held by the owner/author(s).
https://doi.org/10.1145/3586182.3615815
2 ARTENNIS PROTOTYPE

ARTennis is implemented on a HoloLens 2 and is composed of: (1) a real-time bidirectional data stream running at approximately 25 FPS, and (2) a YOLOv7 model fine-tuned on a custom dataset.

2.1 System Implementation

![Figure 2: ARTennis overview and implementation details](image)

**Real-time Bidirectional Streaming.** To enable real-time video streaming, we implemented custom data senders and receivers based on the Web Real-Time Communication (WebRTC) protocol using the MixedReality-WebRTC and aiortc libraries. When the server receives a frame, it runs a fine-tuned YOLOv7 model to locate tennis balls. Once the model has processed the frame, the result is sent back to the HoloLens, where it is visualized.

**Visualizing the Result.** To increase the visual saliency of a tennis ball, ARTennis enhances the ball’s color contrast and adds a crosshair in real-time. We selected this design after a brainstorming session with an LV research team member. Designing improved visualizations is an open research area.

**Initial Data Collection.** To train the YOLOv7 model, we first collected 13 minutes of first-person FOV video (23k frames) from the HoloLens of a research team member playing tennis. While there exists large amounts of tennis video datasets online, these recordings are third-person FOV rather than first-person.

**Custom Dataset Generation.** Given the large number of frames, we automated the dataset generation process rather than manually labeling each frame. When using a pre-trained YOLOv7 model, we were only able to detect a tennis ball in 30% of the images. Thus, we divided each image into a 10x10 grid and defined a search region based on the ball’s grid location in the previous two frames, expanding the search region if necessary. This approach found a tennis ball in 60% of the images. When we randomly sampled 10% of those images, we found that this method had generated an accurate bounding box for 96% of images in the sample.

**Fine-tuning YOLO Model.** We then fine-tuned the YOLOv7 model using our custom dataset and achieved an mAP@0.5 of 91%. Given that the average human reaction time to visual stimuli ranges from 180 to 200 ms [14], we evaluated our model by dividing a test recording into 150 ms segments and found that it successfully identified a tennis ball in 85% of the segments, demonstrating its potential.

2.2 Pilot Evaluation

To evaluate our prototype, an LV research team member played tennis using ARTennis (see demo video). His left eye has no light perception and a Coloboma dominates the right superior portion of his right eye. Consequently, while he can discern colors and shapes, he does not have depth perception, and his visual acuity is 20/450.

After a 45-min play session—he's first ever in tennis—he noted that ARTennis (1) provided him "assurance that the ball is in play", (2) "detected the ball from nearly the span of the court" in many instances, and (3) "matched the best tracking performance I've tried in an AR-based system". However, ARTennis’s visual guidance often extended beyond his field of vision (Figure 3), which rendered the system into "unhelpful flashes of green" when the ball is too high or too far. Specifically, he stated, "beyond 10 or so feet, since I didn’t have much lateral vision, I still only saw one arrow, and this arrow occupied the majority of the usable field-of-view."

He hypothesized that "the HoloLens’ small field-of-view (FOV) is a limitation because it exacerbates discontinuity in the UI as the ball tends to enter and exit the display" and suggested that "audio cues could be a way to overcome this issue". Since the HoloLens 2 has a limited diagonal FOV (52°) [16], we hope to utilize a wearable AR device with a larger FOV in future system versions.

![Figure 3: How a member of the research team who has low vision saw the visualization in Figure 1 due to limits with lateral vision and the HoloLens 2’s 52° FOV.](image)

3 FUTURE WORKS AND CONCLUSION

While our prototype currently focuses on tennis balls, we aim to expand to other tennis-related elements like the net and court lines, as well as other ball-based sports such as basketball, baseball, and badminton. Tennis poses unique challenges for a real-time wearable system due to the fast pace and diminishing size of the ball as it moves away from the player. However, because our approach achieved accurate tracking in such scenarios, we are confident in its potential to support a wide variety of sports. Moreover, as we introduce additional functionalities like ball trajectory estimation, we envision supporting beginners seeking guidance during sports activities. Furthermore, while our focus has primarily been on playing sports, we also hope to explore how our prototype can enhance the viewing experience for sports fans with varying abilities.

In closing, while the HoloLens 2’s size, bulkiness, and limited FOV restricts practical sports deployment, our ARTennis prototype enabled us to create and explore innovative ball tracking and real-time overlay visualizations in head-worn AR. With more lightweight and powerful AR headsets, we envision AR glasses as a powerful tool to enable LV individuals to participate in sports that they otherwise could not.

4 ACKNOWLEDGMENTS

This work has been supported by an NSF GRFP Fellowship and NSF Award #2125087.

---

1. https://webrtc.org
2. https://github.com/microsoft/MixedReality-WebRTC
3. https://github.com/aiortc/aiortc
REFERENCES