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

Jaewook Lee[1] Devesh P. Sarda[1,2] Eujean Lee[1] Amy Seunghyun Lee[3]
Jun Wang[1] Adrian Rodriguez[1] Jon E. Froehlich[1]
[1]University of Washington, Seattle, USA [2]University of Wisconsin-Madison, Madison, USA
[3]University of California, Davis, Davis, USA

Figure 1: Comparison of first-person POV footage captured directly from the HoloLens with (right) and without (left) ARTennis

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

ACM Reference Format:
Jaewook Lee, Devesh P. Sarda, Eujean Lee, Amy Seunghyun Lee, Jun Wang, Adrian Rodriguez, Jon E. Froehlich. 2023. Towards Real-time Computer Vision and Augmented Reality to Support Low Vision Sports: A Demonstration of ARTennis. In *The 36th Annual ACM Symposium on User Interface Software and Technology (UIST '23 Adjunct), October 29-November 1, 2023, San Francisco, CA, USA.* ACM, New York, NY, USA, 3 pages. https://doi.org/10.1145/3586182.3615815

UIST '23 Adjunct, October 29-November 1, 2023, San Francisco, CA, USA

ACM ISBN 979-8-4007-0096-5/23/10.
https://doi.org/10.1145/3586182.3615815

1 INTRODUCTION

People with low vision (LV) often encounter obstacles when participating in sports or exercise, which can negatively affect physical activity [1, 7, 13]. Researchers have devised various prototypes to address this issue, including Kobayashi *et al.*'s camera- and audio-based assistive tools for wall climbing [4], bowling [5], and Mölkky [6], as well as accessible fitness games featuring tennis [8], bowling [9], and yoga [11]. Among different physical activities, competitive sports such as tennis and soccer are particularly challenging as they involve fast, continuously moving elements such as balls and players [12]. Our goal is to overcome this challenge by enabling LV individuals to participate in ball-based sports using real-time computer vision (CV) and wearable augmented reality (AR) headsets with an initial focus on tennis.

Traditionally, tracking tennis balls in real-time required expensive hardware [10]. However, recent advancements in deep learning have led to models like TrackNet [3], a neural network capable of tracking tennis balls in third-person recordings of tennis games. To explore how LV individuals may be able to play tennis with AR, we designed *ARTennis* (Figure 1), a prototype system capable of tracking and enhancing the visual saliency of tennis balls from a first-person point-of-view (POV). We achieve this by streaming video from a pair of AR glasses to a backend server, analyzing the frames using a custom-trained deep learning model, and sending back the results for real-time overlaid visualization.

Our key contributions include: (1) an initial real-time prototype capable of tracking and visually augmenting a tennis ball in wearable AR, and (2) findings from a pilot evaluation with a single LV user. At the UIST demo session, we envision attendees trying ARTennis by manipulating and throwing a tennis ball with the research team.

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

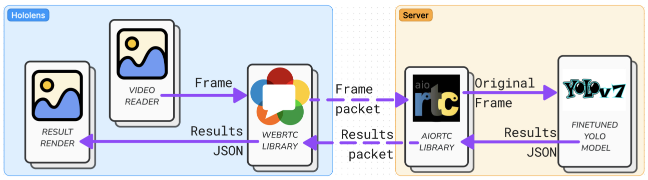

Figure 2: ARTennis overview and implementation details

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[1]) protocol using the *MixedReality-WebRTC*[2] and *aiortc*[3] libraries. When the server receives a frame, it runs a fine-tuned YOLOv7 model [15] 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 [2, 17] 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 POV 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 POV 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 video 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—his 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' 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.

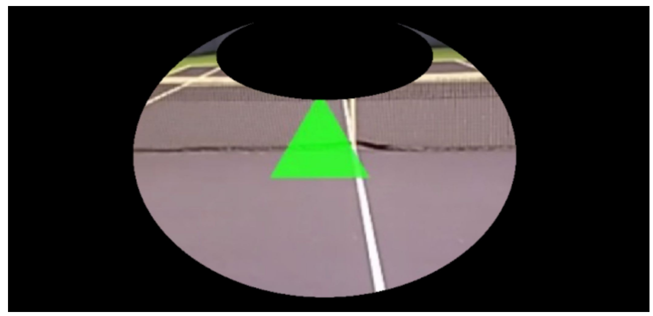

Figure 3: How a member of the research team who has low vision saw the visualization in Figure 1 due to limits with lateral vision and the HoloLens 2's 52° FOV.

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.

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

[1] Michele Capella-McDonnall. 2007. The Need for Health Promotion for Adults who are Visually Impaired. *Journal of Visual Impairment & Blindness* 101, 3 (2007), 133–145. https://doi.org/10.1177/0145482X0710100302 arXiv:https://doi.org/10.1177/0145482X0710100302

[2] R. Cavallaro. 1997. The FoxTrax hockey puck tracking system. *IEEE Computer Graphics and Applications* 17, 2 (1997), 6–12. https://doi.org/10.1109/38.574652

[3] Yu-Chuan Huang, I-No Liao, Ching-Hsuan Chen, Tsì-Uí İk, and Wen-Chih Peng. 2019. TrackNet: A Deep Learning Network for Tracking High-speed and Tiny Objects in Sports Applications. In *2019 16th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS)*. 1–8. https://doi.org/10.1109/AVSS.2019.8909871

[4] Makoto Kobayashi. 2010. A Basic Inspection of Wall-Climbing Support System for the Visually Challenged. In *Computers Helping People with Special Needs*, Klaus Miesenberger, Joachim Klaus, Wolfgang Zagler, and Arthur Karshmer (Eds.). Springer Berlin Heidelberg, Berlin, Heidelberg, 332–337.

[5] Makoto Kobayashi. 2017. Automatic Pin Counting System for the Blind Bowling. *Journal of Advanced Computational Intelligence and Intelligent Informatics* 21, 1 (2017), 119–124. https://doi.org/10.20965/jaciii.2017.p0119

[6] Makoto Kobayashi and Takuya Suzuki. 2022. Accessibility Improvement of Leisure Sports "Mölkky" for Visually Impaired Players Using AI Vision. In *Computers Helping People with Special Needs*, Klaus Miesenberger, Georgios Kouroupetroglou, Katerina Mavrou, Roberto Manduchi, Mario Covarrubias Rodriguez, and Petr Penáz (Eds.). Springer International Publishing, Cham, 73–78.

[7] Rosie K Lindsay et al. 2021. Correlates of Physical Activity among Adults with Sight Loss in High-Income-Countries: A Systematic Review. *International Journal of Environmental Research and Public Health* 18, 22 (9 11 2021), 11763. https://doi.org/10.3390/ijerph182211763

[8] Tony Morelli, John Foley, Luis Columna, Lauren Lieberman, and Eelke Folmer. 2010. VI-Tennis: A Vibrotactile/Audio Exergame for Players Who Are Visually Impaired. In *Proceedings of the Fifth International Conference on the Foundations of Digital Games* (Monterey, California) *(FDG '10)*. Association for Computing Machinery, New York, NY, USA, 147–154. https://doi.org/10.1145/1822348.1822368

[9] Tony Morelli, John Foley, and Eelke Folmer. 2010. Vi-Bowling: A Tactile Spatial Exergame for Individuals with Visual Impairments. In *Proceedings of the 12th International ACM SIGACCESS Conference on Computers and Accessibility* (Orlando, Florida, USA) *(ASSETS '10)*. Association for Computing Machinery, New York, NY, USA, 179–186. https://doi.org/10.1145/1878803.1878836

[10] N. Owens, C. Harris, and C. Stennett. 2003. Hawk-eye tennis system. In *2003 International Conference on Visual Information Engineering VIE 2003*. 182–185. https://doi.org/10.1049/cp:20030517

[11] Kyle Rector, Cynthia L. Bennett, and Julie A. Kientz. 2013. Eyes-Free Yoga: An Exergame Using Depth Cameras for Blind Low Vision Exercise. In *Proceedings of the 15th International ACM SIGACCESS Conference on Computers and Accessibility* (Bellevue, Washington) *(ASSETS '13)*. Association for Computing Machinery, New York, NY, USA, Article 12, 8 pages. https://doi.org/10.1145/2513383.2513392

[12] Kyle Rector, Lauren Milne, Richard E. Ladner, Batya Friedman, and Julie A. Kientz. 2015. Exploring the Opportunities and Challenges with Exercise Technologies for People Who Are Blind or Low-Vision. In *Proceedings of the 17th International ACM SIGACCESS Conference on Computers Accessibility* (Lisbon, Portugal) *(ASSETS '15)*. Association for Computing Machinery, New York, NY, USA, 203–214. https://doi.org/10.1145/2700648.2809846

[13] Lee Smith, Sarah E Jackson, Shahina Pardhan, Guillermo Felipe López-Sánchez, Liang Hu, Chao Cao, Davy Vancampfort, Ai Koyanagi, Brendon Stubbs, Joseph Firth, and Lin Yang. 2019. Visual impairment and objectively measured physical activity and sedentary behaviour in US adolescents and adults: a cross-sectional study. *BMJ Open* 9, 4 (2019). https://doi.org/10.1136/bmjopen-2018-027267 arXiv:https://bmjopen.bmj.com/content/9/4/e027267.full.pdf

[14] P D Thompson et al. 1992. Voluntary stimulus-sensitive jerks and jumps mimicking myoclonus or pathological startle syndromes. *Movement disorders: official journal of the Movement Disorder Society* 7, 3 (1992), 257–262. https://doi.org/10.1002/mds.870070312

[15] Chien-Yao Wang, Alexey Bochkovskiy, and Hong-Yuan Mark Liao. 2022. YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors. arXiv:2207.02696 [cs.CV]

[16] Zequn Wu, Tianhao Zhao, and Chuong Nguyen. 2020. 3D Reconstruction and Object Detection for HoloLens. In *2020 Digital Image Computing: Techniques and Applications (DICTA)*. 1–2. https://doi.org/10.1109/DICTA51227.2020.9363378

[17] Yuhang Zhao, Sarit Szpiro, Lei Shi, and Shiri Azenkot. 2019. Designing and Evaluating a Customizable Head-Mounted Vision Enhancement System for People with Low Vision. *ACM Trans. Access. Comput.* 12, 4, Article 15 (dec 2019), 46 pages. https://doi.org/10.1145/3361866