CARE - Clinician Augmented Reality Environment: Design and Technical Feasibility of Apple Vision Pro for Image-guided Surgeries

Ze Xia Lucas Wang[1], Marianny De León[2], Boyang Zhou[1], Sai Jayanth Kalisi[1], Keith Meng Phou[1], Jennifer Han[1], Yijun Chen[1], Alicia Betty[1], Oliver Aalami[3], Alberto Paderno[4], and Rania Hussein[1*]

[1]University of Washington, USA; [2]Northeastern University, USA; [3]Stanford University, USA; [4]Humanitas University, Italy,

[*]Corresponding Author: Rania Hussein, Department of Electrical and Computer Engineering, University of Washington, 185 Stevens Way, Seattle, WA 98195, United States (rhussein@uw.edu)

ABSTRACT

Objective: Incorporating augmented reality into surgical workflows can enhance surgeon performance by providing real-time feedback and improving interaction with monitors during image-guided procedures. This study explores using the Apple Vision Pro, a head-mounted AR device, to support such procedures, simplify the operating room, and improve situational awareness.
Design and Methods: We developed and iteratively refined a prototype system with input from practicing surgeons to iteratively assess and refine its clinical usability.
Engineering and Results: Our system, CARE, uses a LAN video-casting stack to stream video from image-guided surgery equipment like the Karl Storz IMAGE1 S™, achieving latencies under 71 ms for 1080p and around 200 ms for 4K - outperforming many existing solutions.
Discussion: The Apple Vision Pro's passthrough, hand, and eye tracking features make it a strong candidate for surgical use. Our minimum viable product lays the groundwork for clinical deployment.
Conclusion: The technology stack shows promising competitive latency. Future work will involve operating room clinical trials for further usability findings.

Key words: biomedical engineering, ergonomics, human factors study, simulation, technology assessment, augmented reality

INTRODUCTION

In the evolving landscape of surgical specialties, the shift toward minimally invasive procedures has become increasingly prominent, with most surgeries now being guided by imaging techniques such as laparoscopy, endoscopy, or radiography (fluoroscopy). Monitors have become essential for these procedures, yet their placement and visibility present significant challenges in terms of user experience and ergonomics.

Augmented Reality (AR) tools offer a promising solution by allowing optimal placement of virtual monitors to enhance ergonomics and surgeon comfort. Operating room staff are continually seeking innovations that simplify the environment, improve the surgical experience, and improve patient outcomes. AR devices with strong pass-through capabilities allow computer graphics to be rendered directly in the wearer's field of vision, effectively integrating a 3-dimensional television into the wearer's view without significantly disrupting their vision[1]. Combined with novel input methods such as eye, hand, and body tracking, AR technology enables hands-free interaction with high-resolution content seamlessly integrated into the real world, minimizing latency between the user's perception and their interaction with the environment.

The introduction of the Apple Vision Pro[2] in 2023 marked a significant breakthrough in AR technology. Leveraging spatial computing, the headset projects digital content directly onto high-resolution displays, enabling applications and interfaces to appear in three-dimensional space seamlessly overlaid onto the real world.

The Apple Vision Pro can provide surgeons with real-time information directly within their view, minimizing the need to shift their sight during a procedure.

By seamlessly overlaying diagnostic imaging, anatomical models, and critical patient data onto the surgeon's immediate field of view, these emerging technologies promise to potentially improve patient outcomes across complex medical interventions.

OBJECTIVES

This paper investigates the design and technical feasibility of a native Apple Vision Pro app that surgeons can use to assist during image-guided surgery.

The CARE (2025) project builds on the original 2016 publication by Dr. Oliver Aalami, who investigated the Google Glass for intraoperative monitoring[1]. CARE (2025) leverages the Apple Vision Pro to deliver a modern reassessment of the feasibility and effectiveness of Augmented Reality in image-guided surgery. We propose a limited-scale clinical study based on Liebert et al.[1] The project's mission is to contribute to "simplifying the operating room" and improving surgical efficacy.

DESIGN AND METHODS

Limitations of Current Technology

One advantage of a head-mounted display is its customizability, as different professionals can use the technology in different ways depending on their current needs.

The CARE design team interviewed nine anesthesiologists and three surgeons, revealing three recurring challenges: uncertainty in accessing real-time surgical information, physical constraints from fixed monitor setups, and the need for frequent reorientation in dynamic medical situations.

With an AR implementation, surgeons could independently customize the placement of multiple monitor screens according to their optimal preference without relying on a perioperative nurse. Therefore, they could retrieve patient data and comfortably adjust the monitor screens during surgery, thereby mitigating issues with user experience and ergonomics. In addition, the incorporation of augmented reality will reduce the frequency of sterilization required during the operation, as the monitor screens are entirely virtual. **Figure 1A** shows the operational process without the use of augmented reality. While **Figure 1B** shows the operational process using augmented reality.

The minimum viable product functions primarily through hand gesture interaction. According to A. Paderno, the operating room is a relatively loud environment with multiple audio signals and alarms, where voice commands can be easily drowned out. However, hand gestures and eye capture could still be perceived in high-stakes scenarios, where these gestures are relatively safe within the conventions of the surgical environment. Based on interviews and feedback, two native Apple Vision Pro features below are considered acceptable to be used in future research and product development in the AR surgical space:

- ***Feature 1*** *Drag Surgical Screen:* The surgeon selects a window for movement, and a handle appears at the bottom edge, enabling a pinch-and-drag action to reposition the window within the space.
- ***Feature 2*** *Resize surgical screen:* When the surgeon gazes at the bottom corners of a window, a curved handle will appear. By pinching to select this handle and moving their hands outward diagonally, they can increase the scale of the view. Conversely, moving their hands inward will scale the view down.

However, surgeons may not always have a free hand to perform gestures. To address this limitation, an additional feature was introduced: monitor selection via eye capture. This interaction extends beyond the native capabilities of the Apple Vision Pro and was developed specifically to overcome the constraints of gesture-only control.

- ***Feature 3*** *Select monitor:* Eye capture to display a monitor where a surgeon can reference a selection of monitors at the bottom of their vision, stare to select a monitor, and the selected monitor will be displayed in front of them.

The result is a solution that offers the ability to display and access multiple augmented monitors simultaneously, allowing for varying sizes and placements. This solution would seamlessly sync with the surgeon's workflow, enhancing the overall surgical experience.

ENGINEERING AND RESULTS

Technology Stack Testing

This section includes commentary and implementation of streaming, capturing video via a capture card, measuring latency, and leveraging a local network.

Our experimental setup was designed to measure latency accurately in a simulated environment while maintaining consistent baseline conditions. Before exploring specific components, it is important to understand the overall architecture. **Figure 2** shows that video content streams from a source through a capture card to a streaming server, which then transmits to a reference display (4K TV) and the Vision Pro headset over a controlled network environment. This closely follows the framework of Dr Aalami[1], who used a Local Area Network (LAN) to stream surgical vitals to Google Glasses.

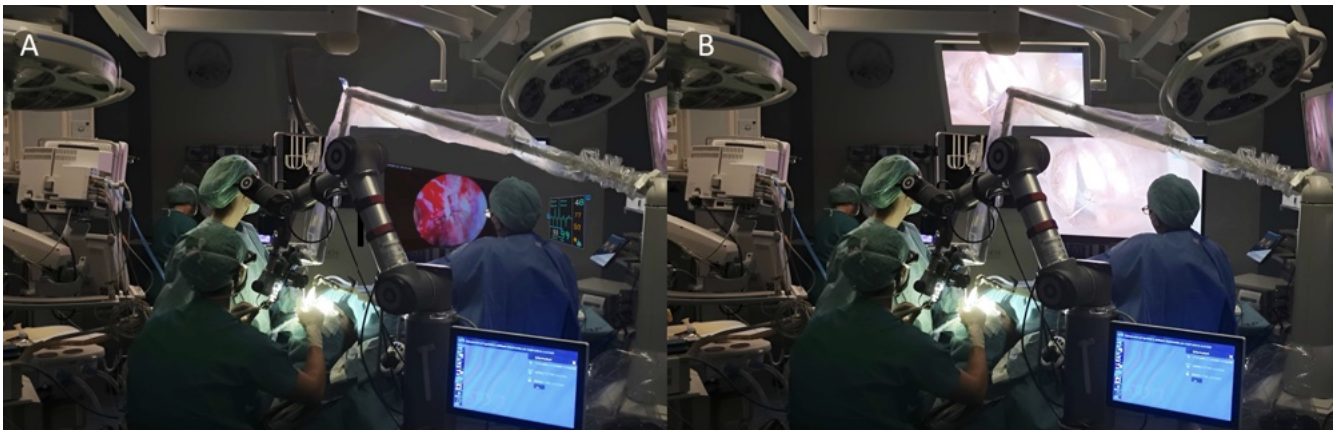

Figure 1. The concept image above compares the operating room with (B) and without (A) the use of the CARE virtual monitor interface, edited with Adobe Photoshop. Note that image (A) only shows the virtual displays from the perspective of the right-most surgeon. The other two surgeons would each have their own customized virtual display layouts, which are exclusively visible to them.

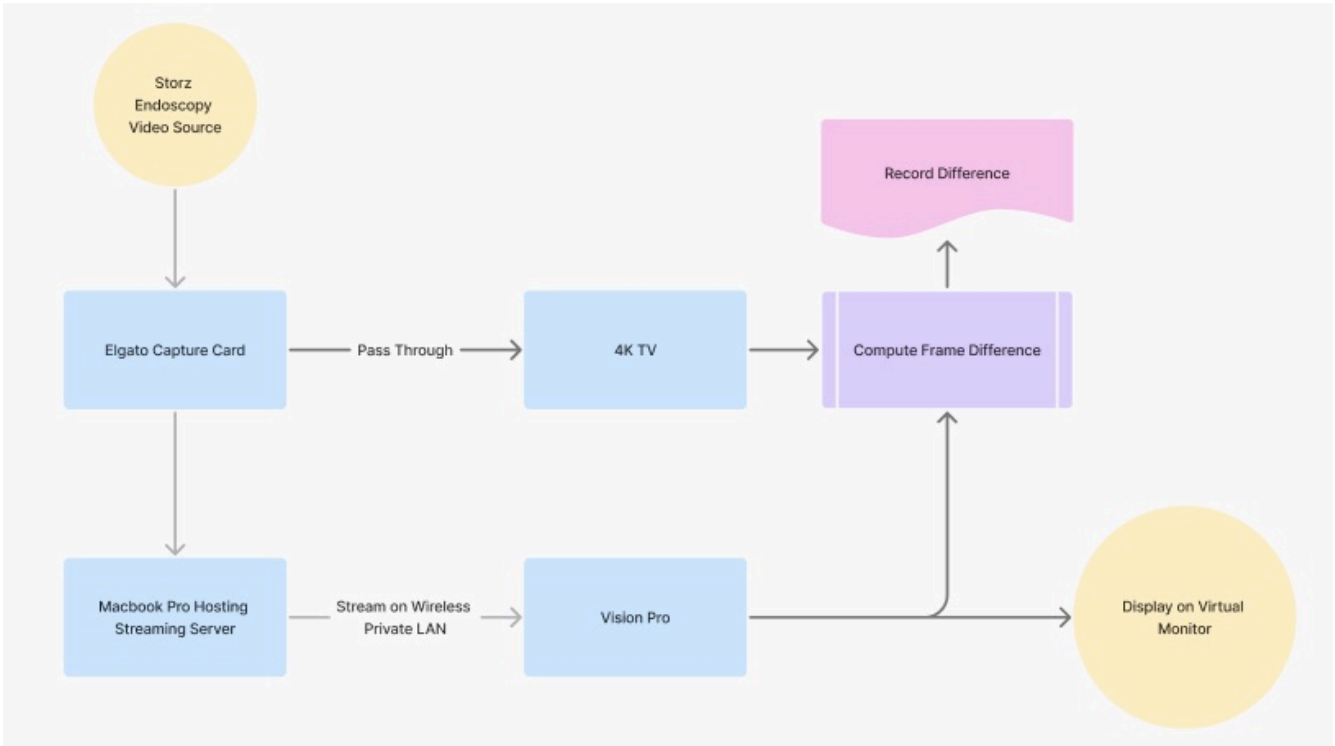

Figure 2. The Image above portrays the flow of information, and how recording took place. For a wired transmission of data, a USB-C cable was used instead of a Private LAN.

In laboratory experiments, prerecorded videos streamed from an internet-connected laptop were used to simulate input to the CARE system. In a clinical setting, this input could originate from endoscopes, vital sign monitors, or other display-enabled medical devices. Our experiments specifically measured the system latency introduced solely by the CARE system, from the video source to the virtual display. Using pre-existing video content provided a consistent baseline, producing accurate and precise latency measurements.

The video would then be input via HDMI to an Apple Macbook Pro (M2 Max Chip, macOS 15 Sequoia) which is part of the same local network as the Vision Pro. A client-server relationship will be established using a streaming service to transmit video information to the Vision Pro, displaying it to the surgeon.

Of the many streaming services explored (REI re:streamer[3], Ensemble[4], and castaway[5]), the Ensemble Framework showed the most promise. Re:streamer came with latency problems, sporting a measured peak delay of 28 seconds, which is unacceptable for clinical work. Although Castaway performed better with a significantly lower latency of about 1 second, the software is proprietary. The Ensemble framework was finally selected as the primary streaming framework due to its open-source customization and testing capabilities. In addition, the latency measured was very low as well. As seen in **Table 1**, the average latency was less than 100 milliseconds for all 1080p videos.

The Open-Source nature of our source material allows us to "stand on the shoulders of giants." Ensemble implemented their frame compression using an lz4[6] compression algorithm. We compared lz4 to lzfse[7] and zlib[8]. We performed preliminary user experience analysis loosely following papers comparing such algorithms[9] to find jitteriness and lag in the zlib and lz4 algorithms compared to lzfse.

A capture card (Elgado 4K X) was used to streamline the testing and forwarding of the video footage from the source to the Vision Pro (Vision OS version 2.0 minimum). **Figure 2** shows how the pass-through functionality enables simultaneous streaming to a 4K TV, which acts as a latency baseline, and a vision pro virtual monitor acting as the surgeon's point of view. OBS[10] (or Quicktime[11]) is required to stream and control video output to the Apple Vision Pro. The video has "black bars" which limit the surgeon's field of view. Removing this involved using Deskpad[12] to act as a virtual monitor. This capture card allowed us to see the instantaneous output from the video source, acting as a visual baseline.

This affords a more precise measurement of the latency, which was done via a frame-wise analysis of videos taken on the Vision Pro. The practice of counting frames to determine the latency of a connection, as seen in Saunders'[13] and Feldstein's[14], allows us to use the number of frames to estimate latency. Thus, a reference frame was considered after a major frame change or transition. The number of frames between the rendering of this frame in the passthrough and the rendering on the Vision Pro was counted and recorded.

We used a dedicated router (CenturyLink C4000) that acts as both a host and a network controller for network configuration. This ensured minimal external network interference during measurements. The same setup was first performed with Vision Pro and an Apple Macbook Pro 2023 (M2 Max) on this local network, testing 4K video and 1080p video. A final baseline was taken with a wired connection using the Vision Pro Developer Developer Cable[15] between the Macbook and Vision Pro, to determine the lowest latency commercial tools enable this setup to provide. For the Wired connection, when connecting to a Macbook Pro, the Vision Pro creates a LAN between itself and the Macbook. This ensures that the streaming setup is left unchanged, creating a valid baseline.

Results Summary

Table 1 compares various video formats and wiring mode latencies. It should be noted that 4K video streaming was significantly more delayed compared to 1080p video streaming. In the wireless modes, we noted latencies of roughly 198.8ms for 4K videos versus only a 70.77ms latency for 1080p videos.

Our 4K videos' latency figures show similar, but smaller latency values than those found in existing literature from Ruijters (300 ms)[16], Askeland (285 ms)[17]. Do note that a few of these slower sources transmit videos that have a lower refresh rate (ex. 20fps) and reduced video quality (ex. 1080p). Our 1080p latency figures outperform both sources further.

However, we do not outperform all preexisting methods. Kibsgaard[18] boasts a 54-70.07ms latency, which even outperforms our 1080p latency figures. Even more impressive is Tokuda[19] who boasts a <10ms latency for a 32 fps video at 1080p.

Regarding quality, the 4K streaming video format was comparatively more jittery. Transition delay was more prevalent. 1080p was much smoother.

Included **Figure 3C** shows limited multi-streaming capability. The Wired pass-through is 1-2 frames ahead of the two virtual monitors. This is expected behavior. Also included are multiple images **Figure 3A** and **Figure 3B**, which show the ideal case and the out-of-sync imperfections in streaming video feeds. **Figure 3D** shows one example video used to benchmark system latency.

Note that these images and, by extension, all screen recordings from the Apple Vision Pro are inherently blurry due to the nature of taking screenshots on the

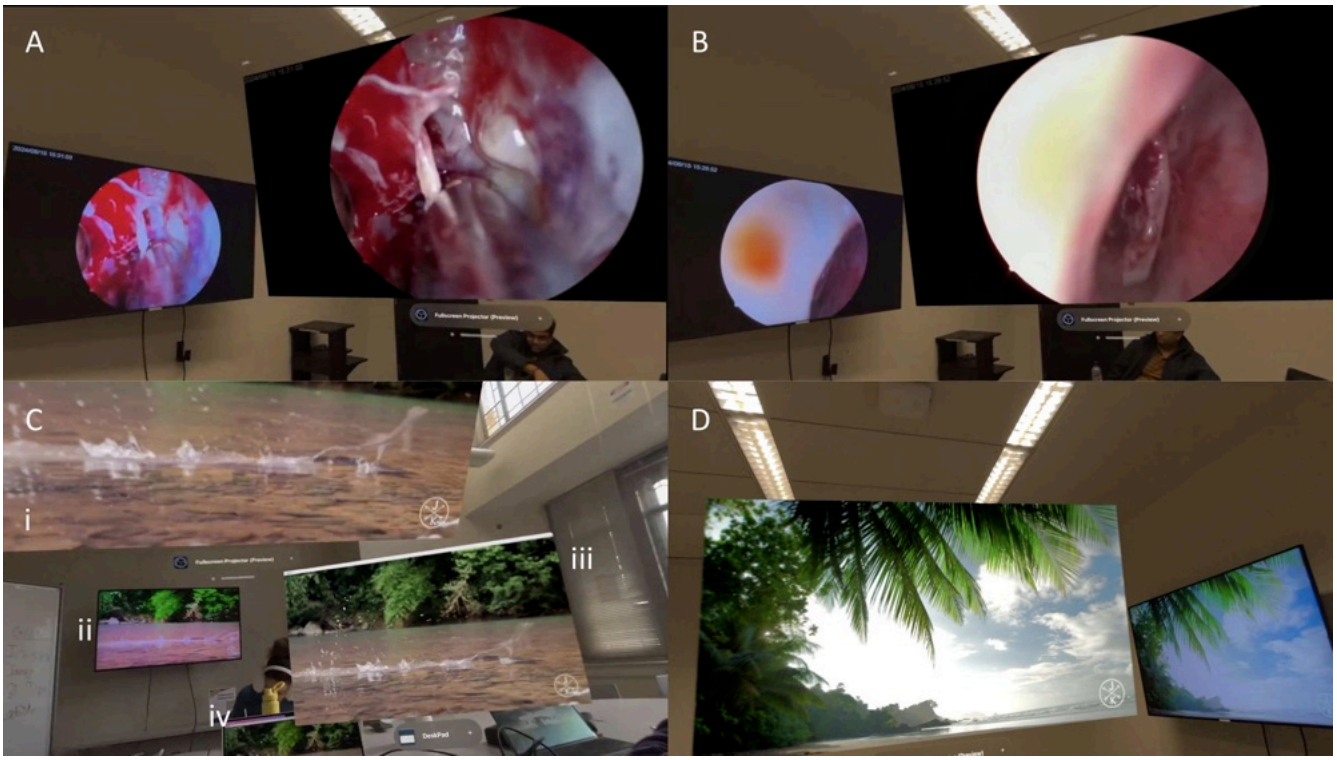

Figure 3. Subimage (A) shows an example of a standard in-sync video streaming during endoscopy. Subimage (B) shows that there are times when the video is a few frames out-of-sync during endoscopy would look like. Subimage (C. i) details multiple monitors streaming the same video. (i) and (iii): 2 streaming Vision Pro Virtual Monitors. (iv): Laptop monitor that sends video to both the virtual monitors and the Pass-through 4K TV. (ii): 4K TV using pass-through functionality, which acts as the wired baseline. Subimage (D) shows a baseline landscape video used to stream and collect latency measurements.

Video Format	Wired Mode	Average Latency (ms)	Median Latency (ms)
1080p	Wireless	70.77	66.66
4K	Wireless	198.8	166.5
1080p	Wired	54.16	66.66
4K	Wired	214.79	199.98

Table 1. Table detailing video format and wired modality and their corresponding latency transmitting 60fps video which was further bottlenecked to 30fps due to the capture card's internal limitations.

Apple Vision Pro. We noticed that the quality of the image and video on the Apple Vision Pro was high-quality 5K, but the pictures we show are condensed and blurry at the edges. This could be due to a projection mapping of a curved field of view onto a flat image.

DISCUSSION AND CONCLUSION

The exploration of system design and the results of our latency testing sufficiently demonstrate that head-mounted displays with augmented reality (AR), and Apple Vision Pro in particular, have significant potential for enhancing intraoperative workflows. Through our testing, we concluded that video streaming of critical surgical data to the Apple Vision Pro can be achieved with latency levels well below 100 milliseconds for 1080p footage, and acceptable latency even for 4K streams when conditions are optimized.

Low latency performance is critical for clinical applications where timely decision-making directly impacts patient safety. Our findings suggest that the Apple Vision Pro, when paired with an efficient streaming framework such as Ensemble, can reliably display surgical information with minimal perceptual delay. These latency figures are a marked improvement over previous AR device implementations, such as the Google Glass-based systems studied by Aalami et al. in 2016, which struggled with both hardware and network limitations.

Furthermore, our interviews with clinicians highlighted the demand for flexible and customizable visual interfaces during surgery. The hand gesture and eye-tracking interaction designs proposed for our minimum viable product (MVP) directly address this need, offering a solution that enhances agency and reduces reliance on other physicians to manage information displays. Our system also mitigates potential sterilization breaches by digitizing monitor interactions, which could translate to better sanitation in the operating room.

Another promising takeaway is that there is minimal difference between the wired transmission modality - there is instead a noticeable and non-negligible difference in latency between different video qualities. The variability in the infrastructure of the hospital networks must be addressed to solve this issue. More precise measurements of latency and the causes of this latency may also provide insight into the direction of future work.

The Vision Pro's high-resolution display and intuitive input systems offer considerable advantages, but concerns regarding the device's weight, comfort during prolonged wear, and battery life persist.

Finally, while our latency measurements were precise under laboratory conditions, we need more studies to validate performance in actual clinical environments, where system performance may be affected by changes in lighting, differing room layouts, and variations in existing workflows.

It is imperative to collect feedback from surgeons at every stage of the process to evaluate the effectiveness of the prototype and refine it accordingly. As a result, future work will focus on clinical testing in the surgery room. We plan to work further with Drs. Aalami and Paderno to carry out preliminary clinical studies based on the study methodologies of Liebert et al. to evaluate the feasibility of the Apple Vision Pro in supporting image-guided clinical procedures.

ACKNOWLEDGMENTS

The CARE team would like to thank our clinical partners, particularly Dr. Alberto Paderno and Dr. Oliver Aalami, who gave us insight into a surgical perspective.

We would also like to thank the University of Washington (Remote Hub Labs) and Stanford University (Spezi Labs). Both provided research resources for a literature review of prior research.

We also thank the contributors for the open-source software that we used, such as Ensemble, OBS, and DeskPad. The foundational work of these communities enabled us to create our streaming workflow for the Vision Pro.

We thank the University of Washington: CoMotion and the NSF I-Corps Hub Northwest Regional program for supporting us through the stakeholder discovery process.

AUTHOR CONTRIBUTIONS

ZXLW contributed to the conception and design, engineering and development of the product, data acquisition, analysis, and drafting of the manuscript. They critically revised the manuscript, were involved in final

approval, and agreed to be accountable for all aspects of the work, ensuring integrity and accuracy.

MD contributed to writing the final paper, research, and review of the manuscript. They contributed to project planning and design supervision, were involved in final approval, and agree to be accountable for all aspects of work, ensuring integrity and accuracy.

SJK contributed to the engineering and development of the product, data acquisition, analysis, and interpretation. They drafted key portions of the manuscript. They were involved in giving final approval, and agree to be accountable for all aspects of work, ensuring integrity and accuracy.

KMP contributed to writing the final paper, engineering, and development of the product. They drafted key portions of the manuscript. They were involved in giving final approval, and agree to be accountable for all aspects of work, ensuring integrity and accuracy.

BZ contributed to paper planning, research into publications, review of the manuscript, and engineering. They were involved in giving final approval, and agree to be accountable for all aspects of work, ensuring integrity and accuracy.

JH contributed to product design, surgeon contact, and interview. They drafted key portions of the manuscript. They were involved in giving final approval, and agree to be accountable for all aspects of work, ensuring integrity and accuracy.

CC contributed to product design, surgeon contact, and interview. They drafted key portions of the manuscript. They were involved in giving final approval, and agree to be accountable for all aspects of work, ensuring integrity and accuracy.

AB contributed to product design, surgeon contact, and interview. They drafted key portions of the manuscript. They were involved in giving final approval, and agree to be accountable for all aspects of work, ensuring integrity and accuracy.

Subteam distribution

Design: Chloe Chen, Alicia Betty, Jennifer Han

Engineering: Sai Jayanth Kalisi, Keith Meng Phou, Boe Zhou

Research Administration: Marianny De Leon, Ze Xia Lucas Wang

Supervision: Oliver Aalami, Alberto Paderno, Rania Hussein

Stakeholder Discovery: Chloe Chen, Alicia Betty, Jennifer Han, Marianny De Leon, Ze Xia Lucas Wang

DECLARATION OF CONFLICTING INTERESTS

The authors have no competing interests to declare.

FUNDING

The author(s) received no financial support for the research, authorship, and/or publication of this article.

REFERENCES

1. Liebert CA, Zayed MA, Aalami O, Tran J, Lau JN. Novel Use of Google Glass for Procedural Wireless Vital Sign Monitoring. Surgical Innovation. 2016;23(4):366-73. PMID: 26848138. Available from: https://doi.org/10.1177/1553350616630142.

2. Inc A. Apple Vision Pro; 2023. Accessed: 2025-01-10. Available from: https://www.apple.com/vision-pro/.

3. Data R, opperman i, stabenow j, venerbeck s, dotter cj, dalmijn w, et al.. Datarhei/Restreamer: The restreamer is a complete streaming server solution for self-hosting. it has a visually appealing user interface and no ongoing license costs. Upload your live stream to YouTube, twitch, Facebook, Vimeo, or other streaming solutions like Wowza. receive video data from OBS and publish it with the RTMP and SRT server.; 2024. Available from: https://github.com/datarhei/restreamer.

4. Saagarjha. Saagarjha/Ensemble: Cast mac windows to visionos;. Available from: https://github.com/saagarjha/Ensemble.

5. Voorhees F. Castaway: Spatial hdmi monitor; 2024. Available from: https://apps.apple.com/us/app/castaway-spatial-hdmi-monitor/id6476697957.

6. Bartík M, Ubik S, Kubalik P. LZ4 compression algorithm on FPGA. In: 2015 IEEE International Conference on Electronics, Circuits, and Systems (ICECS); 2015. p. 179-82.

7. Lzfse. LZFSE/LZFSE: LZFSE Compression Library and command line tool;. Available from: https://github.com/lzfse/lzfse?tab=readme-ov-file.

8. Gailly Jl, Adler M. Zlib Home Site;. Available from: https://zlib.net/.

9. Vestergaard R, Zhang Q, Lucani DE. Enabling Random Access in Universal Compressors. In: IEEE INFOCOM 2021 - IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS); 2021. p. 1-6.

10. 7 W. Project OBS;. Available from: https://obsproject.com/.

11. Apple. QuickTime player user guide for mac;. Available from: https://support.apple.com/guide/quicktime-player/welcome/mac.

12. D S. Stengo/DeskPad: A virtual monitor for screen sharing. Github;. Available from: https://github.com/Stengo/DeskPad.

13. Saunders DR, Woods RL. Direct measurement of the system latency of gaze-contingent displays. Behavior Research Methods. 2014;46(2):439-47.

14. Feldstein IT, Ellis SR. A Simple Video-Based Technique for Measuring Latency in Virtual Reality or Teleoperation. IEEE Transactions on Visualization and Computer Graphics. 2021;27(9):3611-25.

15. Apple. Downloads and resources - visionos;. Available from: https://developer.apple.com/visionos/resources/.

16. Ruijters D, Zinger S, Do L, de With PHN. Latency optimization for autostereoscopic volumetric visualization in image-guided interventions. Neurocomputing. 2014;144:119-27. Available from: https://www.sciencedirect.com/science/article/pii/S0925231214007395.

17. Askeland C, Solberg OV, Bakeng JBL, Reinertsen I, Tangen GA, Hofstad EF, et al.. CustusX: An open-source research platform for image-guided therapy - International Journal of Computer Assisted Radiology and surgery. Springer Berlin Heidelberg; 2015. Available from: https://link.springer.com/article/10.1007/s11548-015-1292-0.

18. Kibsgaard M, Kraus M. Measuring the Latency of an Augmented Reality System for Robot-assisted Minimally Invasive Surgery. In: Proceedings of the 12th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications - Volume 1: GRAPP, (VISIGRAPP 2017). INSTICC. SciTePress; 2017. p. 321-6.

19. Tokuda J, Fischer GS, Papademetris X, Yaniv Z, Ibanez L, Cheng P, et al. OpenIGTLink: an open network protocol for image-guided therapy environment. The International Journal of Medical Robotics and Computer Assisted Surgery. 2009;5(4):423-34. Available from: https://onlinelibrary.wiley.com/doi/abs/10.1002/rcs.274.