UNIVERSITY OF SOUTHAMPTON
Faculty of Engineering and Physical Sciences
School of Electronics and Computer Science

A progress report submitted for the award of
MEng Electrical and Electronics Engineering

Supervisor: Dr. Tom Blount
Examiner: TBD

Open-Source Stereo Video Camera System and Software Implementation for Virtual Reality Lifelogging and Content Creation

by Muhammad Hazimi Bin Yusri

December 11, 2023

UNIVERSITY OF SOUTHAMPTON

ABSTRACT

FACULTY OF ENGINEERING AND PHYSICAL SCIENCES
SCHOOL OF ELECTRONICS AND COMPUTER SCIENCE

A project report submitted for the award of MEng Electrical and Electronics Engineering

by Muhammad Hazimi Bin Yusri

This project addresses challenges related to the accessibility and cost-effectiveness of stereoscopic video recording for Virtual Reality (VR) Head Mounted Display (HMD) technologies. The main goal is to democratize VR content creation by developing an open-source, modular stereo video camera system, implementing an efficient video processing pipeline, and creating an intuitive VR software for exploring stereoscopic media. At its core, the project transforms lifelogging into a personalized, artisanal content creation process. The system incorporates advanced object and scene detection algorithms, allowing for custom metadata tagging within captured videos and images. This integration enhances the lifelogging experience, enabling personalized categorization and efficient retrieval of specific moments, objects, or scenes. Ultimately, the project aims to provide a practical and accessible solution for stereoscopic lifelogging, empowering users to curate and revisit their experiences immersively in VR.

Contents

List of Figures

List of Tables

Acknowledgements

I would like to thanks Dr Tom Blount for his tremendous support, ideas and approval for this project, and also to Godot XR community especially Bastiaan Olij and Malcolm Nixon for help with shaders, stereo video example and relentless guidance.

Chapter 1

Introduction

1.1 Problem Statement

The surge in Virtual Reality (VR) Head Mounted Display (HMD) technologies has opened new creative avenues, offering users immersive experiences [1], [2]. However, the challenges of exclusivity and high costs in existing solutions for recording stereoscopic/3D/spatial video limit their potential [3]. This holds significance for VR's broader adoption, as low user retention is often linked to a lack of exclusive content maximizing the medium's advantages [4]. Viewing 2D content on VR headsets, despite superior displays elsewhere, underlines the need to empower users to effortlessly create their stereoscopic content, enhancing the appeal of VR [4]. Traditional VR content creation's focus on factors like Field of View (FOV) and 3 Degrees of Freedom (3DoF) often results in visually uninteresting videos lacking depth [4], [5]. The exclusive depth perception feature on VR HMDs and certain autostereoscopic displays has the potential to revive lifelogging, offering individuals an immersive experience to relive memories [6].

1.2 Goals

This project endeavours to surmount these challenges by:

1. Developing an Open-Source, Cost-Efficient Stereo Video Camera Hardware System.

2. Implementing a Video Processing Pipeline.

3. Creating Intuitive VR Software for Browsing and Viewing Stereoscopic Media.

This project not only addresses the critical gaps in hardware accessibility and the video processing pipeline but also represents a significant enhancement to existing stereoscopic video VR software. By fostering an open-source, modular, and cost-effective approach, this initiative strives to democratize VR content creation, making it more widely accessible and fostering innovation in the field.

Chapter 2

Background

2.1 Lifelogging

Lifelogging, a contemporary term encapsulating habitual documentation of one life akin to social media practices, distinguishes itself through its methodical and routine nature. The motivation for lifelogging extends beyond the sporadic capturing of moments to a deliberate effort to systematically record and preserve lifetime memories.[7] From the definition itself, the Lifelog data can be in any format such as texts from diaries, health sensor data or even digital footprint from application use. However, in this project, I will be focusing on visual lifelogging which consists of images and videos.[8]

Examples of existing wearable cameras specifics that are used for visual lifelogging are shown in Fig 2.1.[8] However, in recent years with rapid advancement in the miniaturization of consumer products such as smartphones, camera sensors and storage devices, companies like Meta[9] and Snapchat also released their versions of

FIGURE 2.1: Evolution of wearable camera technology. From left to right: Mann (1998), GoPro (2002), SenseCam (2005), Narrative Clip (2013), reproduced from [8]

FIGURE 2.2: Commercially available consumer-level stereo cameras. From left to right: Snapchat Spectacles 3[11], Kandao Qoocam Ego [16], iPhone 15 Pro/Pro Max Spatial Video[17]

smart glasses, similar to Google Glass[10] released years prior. Although the main appeal is for the smart features for convenience, this form factor and placement of the camera is perfect for visual lifelogging aimed for VR userbase especially as already proven by Snapchat Spectacles 3[11] which can capture 3D photos and videos at 60fps.

2.2 Stereo Camera

Stereo camera refers to a camera with stereoscopic imaging capabilities, this is simply achieved using 2 cameras aligned side-by-side at a distance approximately the same as average human inter-pupillary distance (IPD) to mimic human vision[12], [13]. This gives the content viewed proper depth definition as in real life. This idea can also be extrapolated to higher Field Of View (FOV) content such as 180/360 degrees content but requires either a more complicated setup, computation or both[14], [15]. A higher resolution is also needed as the content is stretched into a larger area [4], this in turn increase storage and processing requirement.

Examples of consumer-level stereo cameras commercially available to buy right now as shown in Fig 2.2 are Snapchat Spectacles 3 [11], Kandao QooCam Ego [16] which includes an integrated stereo/VR viewer and technically, iPhone 15 Pro/Pro Max [18] with version iOS 17.2 and above [17] which had Spatial Video recording enabled. This move is obviously to entice consumers using Apple products to buy their upcoming Apple Vision Pro VR HMD [2] which will make stereo camera and content more mainstream.

FIGURE 2.3: Example of VR HMDs. Meta Quest 3[1] and Apple Vision Pro[2]

2.3 Virtual Reality (VR)

Virtual Reality (VR) transcends the conventional by offering a spatial computing platform, a virtual 3D environment, and an immersive emulation of real life. The emotional attachment fostered by VR content stems from its heightened realism, creating a sense of physical presence within the virtual space.[4] This emotional resonance distinguishes VR content from traditional media, providing a compelling reason for its adoption in content consumption including lifelog data.

Contrary to fully immersive Free Viewpoint Video (FVV) [19] or scene reconstruction [20], the adoption of the SBS format is grounded in the current limitations of VR Head-Mounted Display (HMD) hardware. While FVV offers unparalleled immersion and depth perception through 6DoF [21], practical constraints dictate a compromise. This decision is further supported by the spatial video capabilities of the iPhone 15 Pro, which, despite being 1080p at 60fps, aligns with the current standards for windowed style viewing. The emphasis here is not on resolution but on perceived Pixel Per Degree (PPD) which is higher as the virtual window/screen is further, and lower FOV to stretch out. Choosing windowed viewing over panoramic alternatives (180/360 degrees) acknowledges the balance between realism and current technological constraints.

A poignant illustration of this concept is evident in the Black Mirror episode which explores the immersive preservation of memories in an eye-camera format.[22] This format aligns with human visual perception, eliminating the necessity for constant 3 Degrees of Freedom (3DoF) as the human gaze is not constantly omnidirectional. Although a higher FOV would enhance the lifelogging experience, it remains cost-prohibitive at present (full human FOV is approximately 220 degrees)[12].

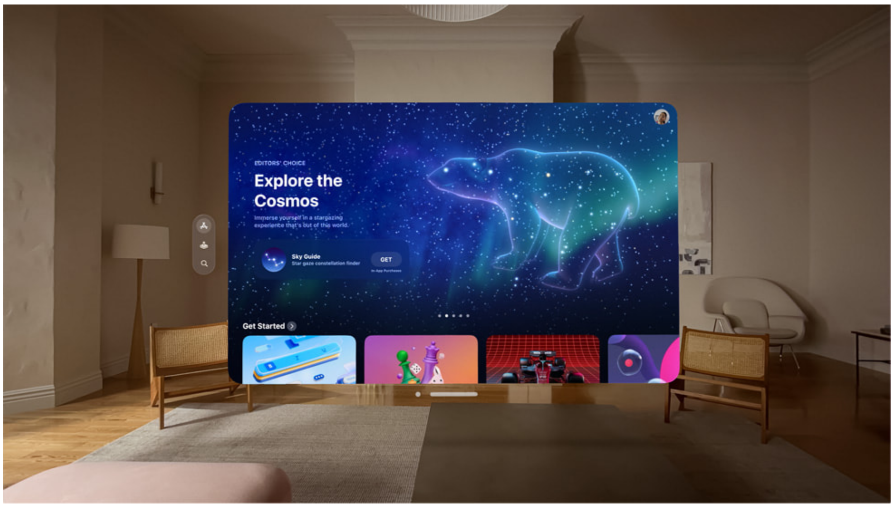

FIGURE 2.4: Apple's visionOS innovative gesture and eye tracking UI navigation, making full use of 3D space for windows elements with variable sizes.[2]

2.4 User Interface in VR

Integrating VR software for content browsing amplifies the lifelogging experience. Leveraging the full potential of 6DoF through innovative UI design, expanded screen space, [23] and interactive features such as timeline shelves, this approach redefines how users engage with their memories. The incorporation of hand tracking and eye tracking further enhances the intuitive and immersive nature of content navigation within the VR environment as shown in visionOS demo of Apple Vision Pro in Figure 2.4.

Other than that, the existence of a window or virtual screen to host the content will cause less motion sickness as users have virtual anchors to the virtual space [24], [25]. Lower FOV helps as the experience will be just like watching movies in 3D cinema[26]–[28].

The combination of lifelogging and VR represents a novel and niche subset within both domains. As technology advances and user preferences evolve, this integrated approach is poised to become more mainstream, particularly in the realm of personal videos, memories, and photos. The synergy between lifelogging and VR anticipates a future where individuals seamlessly capture, relive, and share their most cherished moments in a more immersive and engaging manner.

Chapter 3

Technical Progress

The project is divided into 3 distinct parts, following the MoSCoW requirement framework (M: Must-haves, S: Should-haves, C: Could-have, W: Wont-have).

3.1 MoSCoW Requirement

3.1.1 Hardware Development

- **M:** Develop an open-source, modular stereo video camera system with the Raspberry Pi Pico microcontroller. Ensure it is low-cost and easily accessible.
- **S:** Consider additional features or improvements based on feasibility, such as using an onboard rechargeable Li-Po battery circuit instead of a power bank to power it.
- **C:** Explore advanced features like wireless connectivity or additional sensor integration, time permitting and if resources allow.
- **W:** Exclude features or components that are deemed impractical or beyond the scope of the project, such as higher resolution or different video format (180/360).

3.1.2 Video Processing Pipeline

- **M:** Implement a video processing pipeline for transforming mono stills and video into stereoscopic Side-By-Side (SBS) format. Synchronize audio files for an immersive surround soundscape.
- **S:** Explore additional video processing features, such as metadata tagging through object and scene detection using existing libraries and tools.
- **C:** Automated video stabilization or advanced filtering options, based on available resources and time constraints.
- **W:** Exclude overly complex video processing tasks that may hinder the project timeline or exceed available resources, such as 3D depth reconstruction.

3.1.3 VR Software Application

- **M:** Develop an intuitive VR software application for seamless file browsing and content viewing. Ensure compatibility with the stereo video format.
- **S:** Implement innovative UI designs for enhanced interaction within the VR environment.
- **C:** Explore the integration of hand tracking, and eye tracking, if resources and time allow.
- **W:** Exclude overly ambitious features that may compromise the core functionality or extend the project beyond feasible timelines, such as a personal AI assistant.

3.2 The Big Picture

The algorithm in the first part of the project, 3.3 runs onboard the microcontroller, the second part, 3.4 is designed to run on a machine running Windows 10 while the last one 3.5 has options to run on any OpenXR [29] capable devices, however, PCVR support is prioritized. Figure 3.1 shows a flowchart representing the flow of different parts of the project.

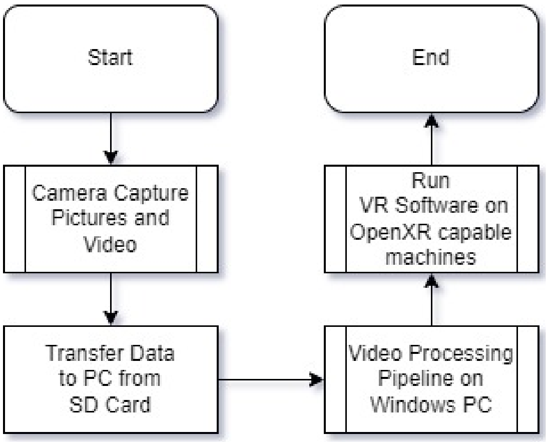

FIGURE 3.1: Overall flow of the project

3.3 Hardware System

3.3.1 Cost Analysis

The budget provided for the project is £150, and the aim is to get all components under £100, making it a low-cost solution compared to alternatives that usually cost around £300 or more[11]. The most expensive part, as expected, is the camera modules, which are around £35 each. However, they boast a 5MP sensor and are capable of taking 1080p60fps video, justifying their cost. The exact component details and costs can be seen in Appendix A.

3.3.2 Components

After comparing costs, availability, ease of use, and hardware constraints, the choice of the microcontroller is Raspberry Pi Pico due to its low cost, cohesive documentation, compatible hardware, and, most importantly, buffer size, which is important to get high enough resolution images/video to prevent motion sickness.

To achieve an immersive experience, audio is also an important variable. Thus, the use of 2 independent electret microphone modules is added to work in tandem with the camera. This, in theory, should make it possible to achieve stereo sound channels for each ear. An SD card extension board is also added to host the SD card that holds all the data.

For the complete system, one of the MCUs will act as the main MCU, it also has a status LED to show when it is recording or taking pictures. It sends signals through GP18 to the other two mono camera systems when the button is pressed to synchronise the capture timing. Trying to make it power-efficient might prove challenging and needs to deal with additional circuitry, so for now, the system will be powered with a power bank to the micro-USB on the main Pico board. The circuit schematics for mono camera setup and complete system can be seen in Figure 3.2.

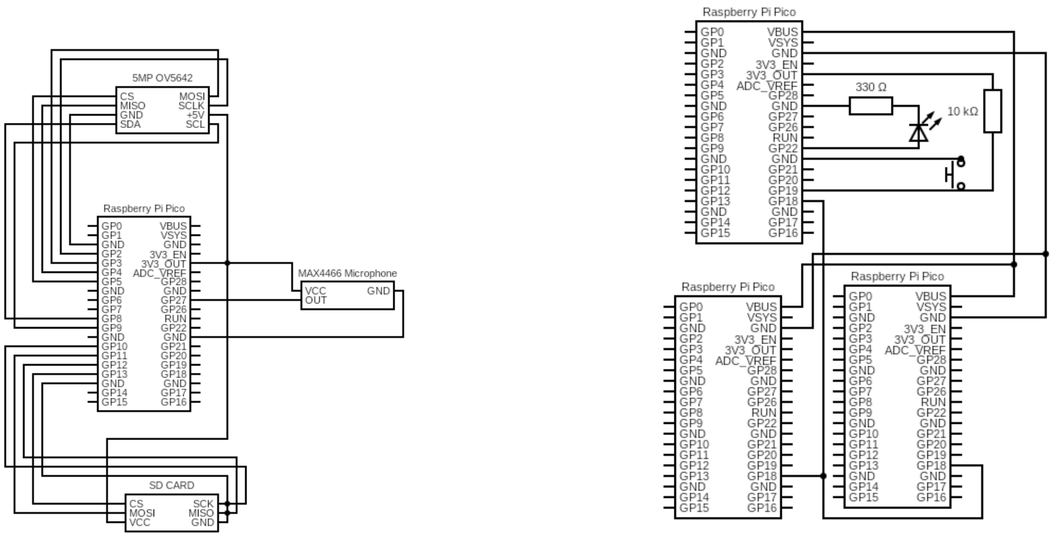

Figure 3.2: Mono Camera and Complete System Prototype Circuit

3.3.3 POV Considerations

The initial draft design is to have both cameras mounted on the side of eyeglasses, akin to Ray-Ban Meta and Snapchat Spectacles glasses. However, after getting the camera module and other components, it is deemed too unwieldy to fit the electronics into a small form factor. The main reason for this choice is to capture the footage as close as how humans see the world, and mounting the camera close to the eye would achieve that compared to a chest mount design.

To compromise, a head-mounted design is chosen, where the POV is higher than eye level, but the camera movement/rotation will still follow the head movement and should give a realistic enough POV as seen in Figure 3.3.

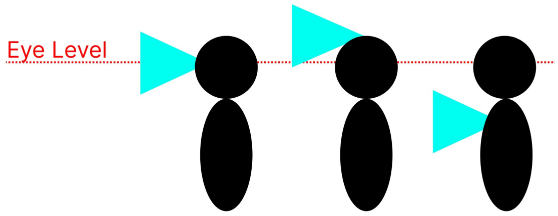

FIGURE 3.3: Designs and POV Example. Left to right: Glasses, Head-Mounted and Chest-Mounted

3.3.4 Onboard Embedded Software Algorithm

In normal lifelogging mode, the cameras would take stereo still images on a timed interval, which will be decided through trial and error and optimization depending on how much storage the images occupy and the power efficiency of the algorithms. However, to get a more immersive experience, video is also needed, and a button can be used to manually start and stop recording, with an LED being an indicator when it's recording. The saved files should be aptly named for easier processing later, in a standardized DATE_NUMBER format with L or R prefixes for the left and right camera respectively. The proposed algorithm can be seen in Figure 3.4

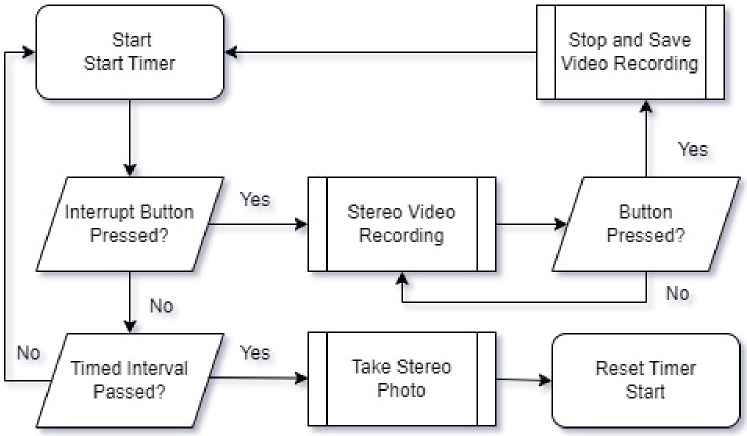

FIGURE 3.4: Flow Chart for Onboard Embedded Software Algorithm

3.4 Video Processing Pipeline

3.4.1 Justification

The justification behind implementing a pipelined process for video processing is to offload the work to a more capable machine, this stems from the acknowledgement that attempting to execute this intricate task directly on the Raspberry Pi Pico would introduce unnecessary complications exceeding its memory and processing power limitations. This is particularly crucial for the generation of Side-By-Side (SBS)/3D format from two mono/2D stills using FFmpeg. Additionally, the inclusion of a custom metadata tagging system for object and scene detection is pivotal. This customized metadata enhances the overall VR experience by enabling advanced filtering and indexing, thereby optimizing the video and image browsing experience within the VR application.

3.4.2 FFmpeg for stitching

The utilization of FFmpeg [30] for mono to stereo stitching serves as a cornerstone in the pre-processing pipeline. FFMPEG, a powerful multimedia processing tool, efficiently transforms mono/2D videos and images into the desired stereoscopic Side-By-Side (SBS)/3D format as seen in Fig 3.5. This process is instrumental in creating a lifelike and immersive visual experience for VR content, aligning with the project's goal of democratizing VR content creation. Refer to Appendix F for the automated FFMPEG PowerShell script.

FIGURE 3.5: Examples of Mono (left) and Stereo (right) images that are stitched from two mono stills

3.4.3 Metadata tagging

Incorporating object and scene detection algorithms further enhances the pre-processing pipeline. Utilizing well-established computer vision techniques, this step automatically identifies and tags objects and scenes within the videos and images. This integration not only enriches the visual content but also establishes the groundwork for sophisticated filtering and indexing capabilities within the VR application. The custom metadata tagging component allows for more refined filtering options, enabling a personalized categorization of lifelogging data. This bespoke metadata layer optimizes the VR content browsing experience, ensuring users can effortlessly locate and revisit specific moments within their immersive collection. At the simplest, this will just be .json files corresponding to each media file.

3.5 Video Player VR App

3.5.1 Game Engine – Godot 4.2

The selection of the Godot 4.2 game engine[31] for the development of the video player VR app is rooted in the project's commitment to Free and Open Source Software (FOSS) principles. Unlike more established game engines such as Unity or Unreal Engine, Godot aligns with the FOSS ethos, making it the ideal choice for this project. Despite having fewer documentation and examples compared to its counterparts, this presented an opportunity for active participation in its development. Consultation with the Godot community, particularly through their Discord channel, guided the decision-making process, revealing that utilizing shaders is the most straightforward approach for rendering stereo Side-By-Side (SBS) video playback.

3.5.2 SBS Video Projection

In the implementation of stereo video playback using the Godot 4.2 game engine, shaders play a crucial role in rendering. The process is straightforward, aligning with the Side-By-Side (SBS) format of the video. The left half is rendered for the left camera eye, and the right half for the right eye. The gdshaders code in Appendix G utilized adheres to this logic, ensuring an efficient rendering process.

Moreover, the user interface (UI) for the video player VR app is designed within a 3D space, a common practice for VR applications and games. However, a challenge arises as the video player operates as a 2D screen within this 3D environment. Initial attempts to integrate the shader script within the same scene as the videostreamplayer node, following a tutorial by Malcolm Nixon on YouTube, faced difficulties. Subsequent experimentation and development efforts proved inconclusive.

Fortunately, after seeking guidance from the Discord community, Malcolm Nixon, the tutorial's author, provided a pivotal solution[32]. The revised method involves instantiating a 2D screen videostreamplayer node in the main scene, where the shader is then applied. This modification successfully addresses the challenges encountered during the development process. An example of the app running is depicted in Figure 3.6.

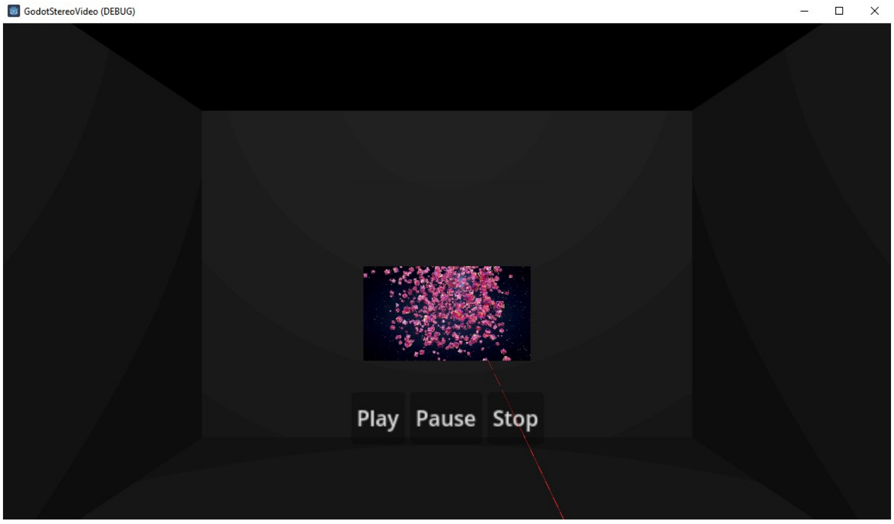

FIGURE 3.6: Images of App Running

3.5.3 File Browsing

While file browsing functionality has not been fully implemented, conceptualization has begun. Proposed ideas include leveraging metadata tagging for specific searches and employing bookshelves or 3D objects as interfaces for navigating between months or weeks, deviating from the conventional 2D screen approach to enhance interactivity.

To optimize browsing efficiency, the screen space may be expanded to resemble an ultra-wide monitor or more, enabling users to view a greater number of files in a single window. This approach capitalizes on the immersive capabilities of VR, utilizing the 360-degree view and 6 Degrees of Freedom (6DoF). Additionally, the implementation of the depth axis remains contingent on contextual considerations and future developments.

Chapter 4

Plan of Remaining Work

This project consists of three distinct components, simplified to their basic elements to demonstrate a minimum viable product prototype. Therefore, meticulous planning and initial design drafting are essential, and this has been executed thoroughly, as evidenced. However, the challenging phase lies in translating the bulk of the design into a functional solution, which is where my focus will be directed in the upcoming phases.

As observed in the Gantt Chart in Appendix B and C, the proposed timeline has encountered some deviations, primarily due to unforeseen risks such as the microcontroller breaking due to a short circuit during user testing and instability in the connections of the wires provided by manufacturers, thereby causing delays in testing. This had been recorded in the risk assessment in Appendix D. Nevertheless, all three main components underwent preliminary testing, and initial work was successfully accomplished.

Appendix C displays the executed Gantt chart timeline alongside the planned one, distinguished by varying colour densities. Through this comparison, I recognized the necessity of refining the specificity of tasks in my Gantt chart for future reference.

Appendix A

Cost Table

Budget for 3rd Year IP = £150

Supplier: The Pi Hut

Refer Table A.1 for cost breakdown by components.

No.	Item	Function	Cost (£)	Quantity
1	Raspberry Pi Pico W	Signal/Control/Main Pico	6.3	1
2	Raspberry Pi Pico	Eye'/Camera Pico	3.9	2
3	5MP ArduCam Camera module, OV5642	Camera Module (SPI)	35	2
4	SD Card SPI Breakout Board	SD Card 'reader' (SPI)	3	2
5	32GB MicroSD Card	Storage	8	2
6	Electret Microphone Amplifier	Audio recorder	6.9	2
7	Breadboard for Pico	Prototyping	4.2	1
	Shipping Fee		3.99	1
			Total	128.09

Table A.1: Cost Breakdown of Components

Appendix B

Planned Gantt Chart

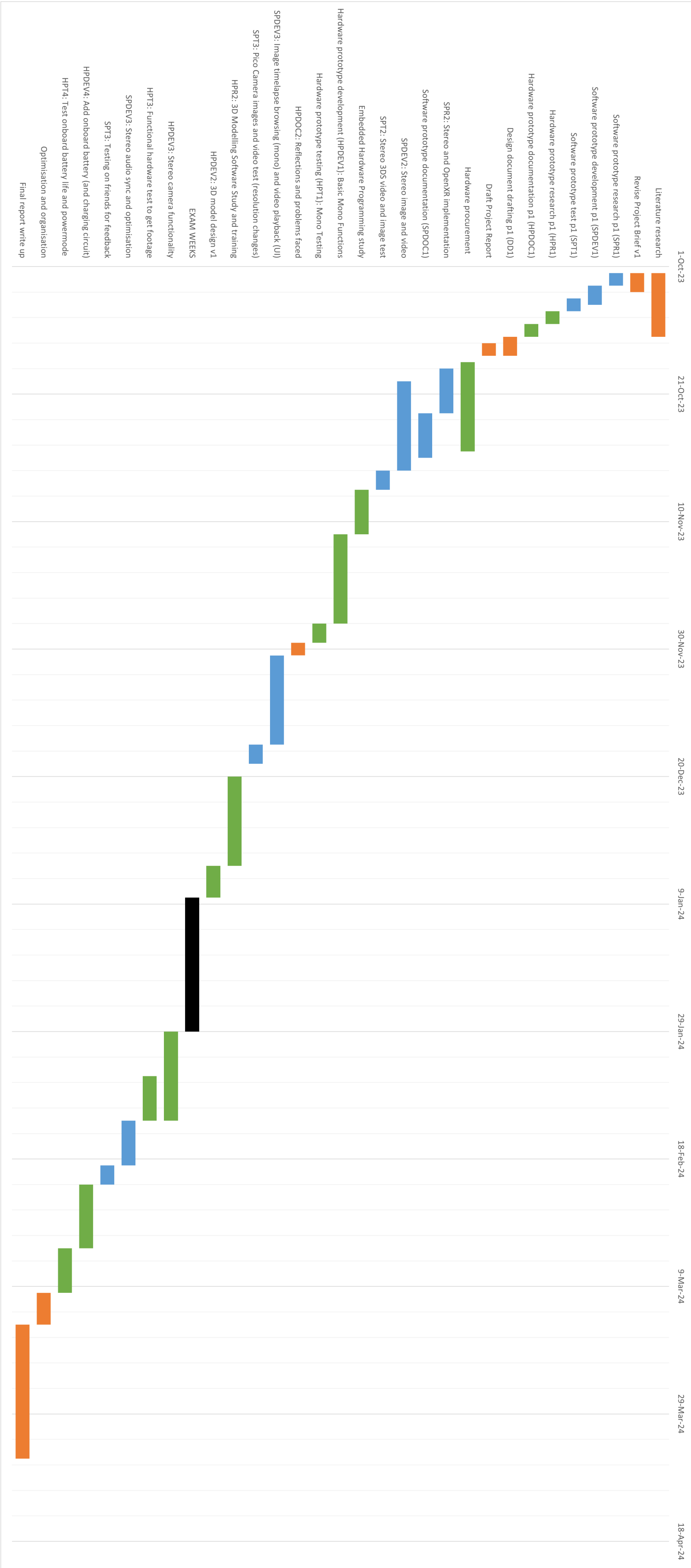

Figure B.1: Planned Gantt Chart

Appendix C

Combined Gantt Chart

Legends:

Orange: Report or writing related task
Blue: Software related task
Green: Hardware related task

Softer colour represents task over expected time or unexpected tasks compared to Appendix B.

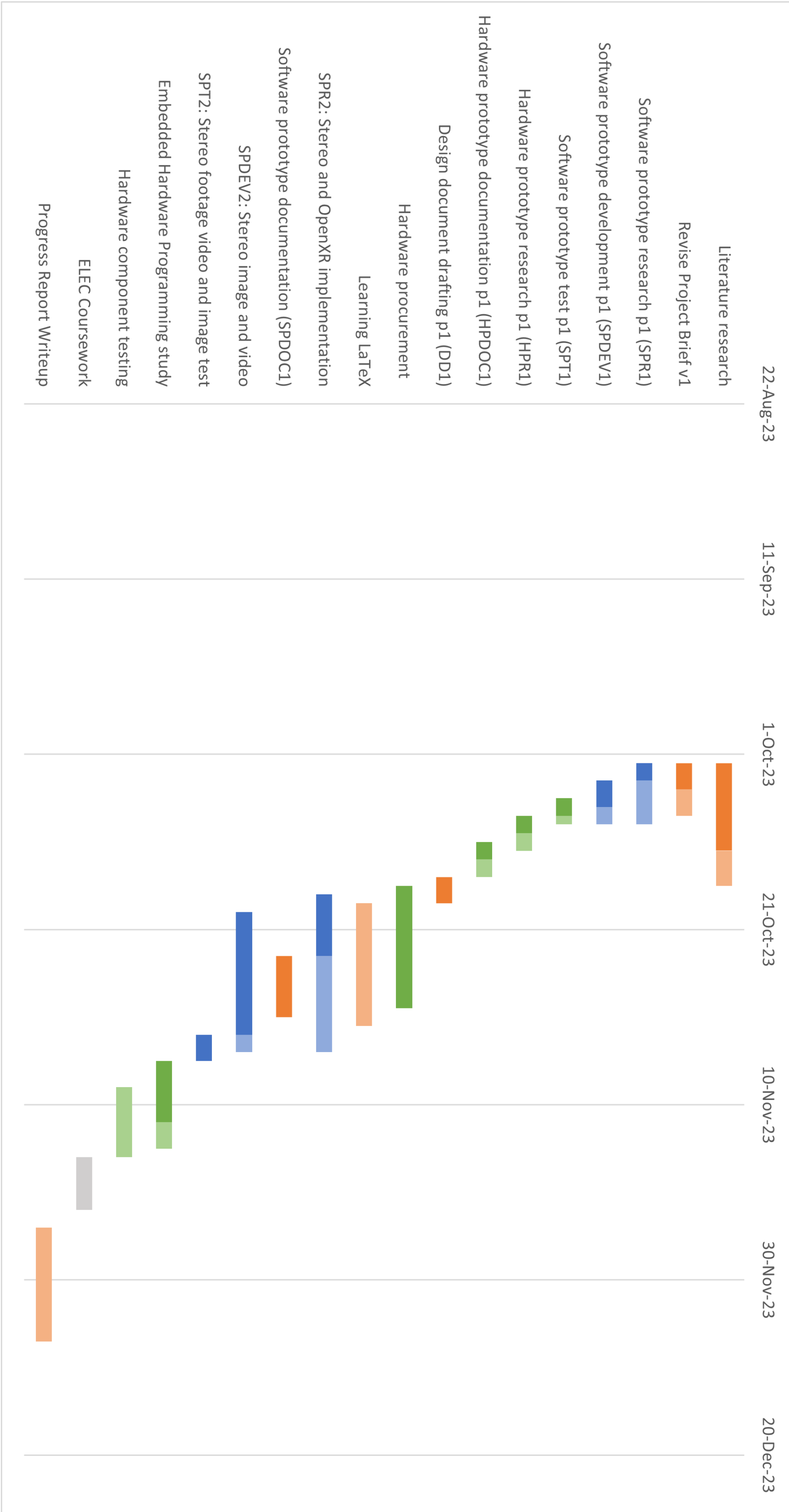

Figure C.1: Combined Gantt Chart

Appendix D

Risk Assessment

Risk event	Likelihood (1-5)	Impact (1-5)	Risk Exposure (1-25)	Mitigative Action	Alternative Action	Occurs? (Yes/No/TBD)	Actions taken
Fitting electronics into a spectacle frame clip-on enclosure is too technically complex and difficult.	4	4	16	Conduct thorough prototyping and testing of the case design early in the project to identify and address integration issues.	Explore alternative enclosure designs or materials that are more manageable in terms of complexity.	Yes	Instead of a glass frame design, a head-mounted design is chosen
Processing power and onboard memory of Raspberry Pi Pico might not be enough for the complexity of this project.	4	5	20	Optimize code and data storage to work efficiently within the constraints of the Raspberry Pi Pico boards.	Consider upgrading to more powerful microcontrollers if necessary.	TBD	
Powering three Raspberry Pi Pico devices with their electronics from a single supply may lead to suboptimal or non-functional performance.	5	3	15	Calculate power consumption and explore energy-efficient solutions. Implement features such as power management and hot-swappable batteries to address power challenges.	Use separate power supplies for each Raspberry Pi Pico to ensure stable and independent power sources.	TBD	

TABLE D.1: Risk Management Table (Part 1)

Developing VR-specific lifelogging software with features like immersive user interface elements, scene/object detection, and metadata auto-tagging may introduce technical complexities, potentially causing project delays.	5	2	10	Simplifying the software scope, such as omitting scene detection and reducing metadata features, can help mitigate technical complexities and minimize project delays.	Utilize pre-existing VR software or libraries to simplify the software development process while still achieving project goals.	TBD	
Implementing stereo content within a VR game engine may prove technically challenging and time-consuming.	5	5	25	Plan for potential difficulties in stereo implementation and be prepared to use existing applications for viewing SBS content as an alternative.	Explore VR game engine plugins or assets specifically designed for SBS video playback, simplifying the integration and reducing technical complexities.	No	Consulting with GodotXRTools developer results in a working stereo video player prototype.
NEW: Components breakdown or not working perfectly due to technical error or manufacturing defects	5	5	25	Be careful when testing components using a multimeter, especially when turned on.	Have enough budget left to buy spare components, or use other available components from the university.	Yes	

TABLE D.2: Risk Management Table (Part 2)

Appendix E

Mono Camera Testing

As mentioned earlier, unintentional shorting occurred on one of the Raspberry Pi Pico microcontrollers during debugging, attributed to a rusty jumper wire. After cleaning, functionality was restored, though not flawlessly.

Following the guidelines from the PICO SPI CAM GitHub Repository example project, both cameras underwent testing. However, only the low resolution appears operational; adjusting resolution settings yielded no change. This issue is suspected to stem from an unstable connection due to a defective jumper cable. To address this, the plan is to proceed with the prototype design for the complete system, incorporating soldering for a more reliable connection.

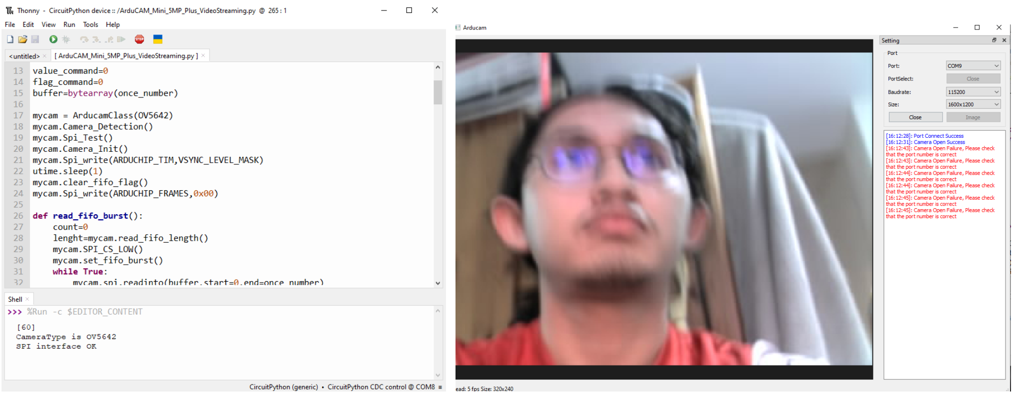

Figure E.1: LEFT: MicroPython software example flashed using Thonny IDE
RIGHT: Unstable Port Connection shown in red

The focal length also plays a crucial role in maintaining subject focus, accommodating near or distant objects. However, experimentation revealed that using the default length produced the best results for the intended application.

Figure E.2: Blurry and Clear Pictures

Appendix F

FFMPEG

Using following example script to stitch two videos horizontally is easy. However there will be hundreds or thousands of files to be processed thus automation is needed.

Using a simple powershell script (.ps1) on Windows as follows solve the problem:

```
    # Get the current directory of the script
$scriptDirectory = Split-Path -Parent $MyInvocation.MyCommand.Path

# Get all left input files matching the LDATE_NUMBER format
$leftFiles = Get-ChildItem -Path $scriptDirectory -Filter "L*.mp4"

foreach ($leftFile in $leftFiles) {
    # Extract DATE_TIME part from the left input file name
    $dateTime = [System.IO.Path]::GetFileNameWithoutExtension($leftFile.Name) \\
    -replace "^L"

    # Construct the corresponding right input file
    $rightFileName = "R$($dateTime).mp4"
    $rightFile = Join-Path -Path $scriptDirectory -ChildPath $rightFileName

    $outputFileName = "SBS$($dateTime).mp4"

    if (Test-Path $rightFile) {
        # Stack left and right files horizontally into the output file
        ffmpeg -i $leftFile -i $rightFile -filter_complex \\
        "[0:v][1:v]hstack=inputs=2[v]" \\
        -map "[v]" -codec:a copy $outputFileName
    }
}
```

Appendix G

Shaders Code

The following is the .gdshaders code used for the SBS stereo effect:

```
shader_type spatial;
render_mode unshaded;

uniform sampler2D movie : source_color;

void vertex() {
  UV = vec2(UV.x * 0.5, UV.y);
  if (VIEW_INDEX == VIEW_RIGHT) {
    UV.x += 0.5;
  }
}

void fragment() {
  ALBEDO = texture(movie, UV).rgb;
}
```

More shaders options can be found in Godot Shaders documentation.

Bibliography

[1] *Meta Quest 3: New Mixed Reality VR Headset*. [Online]. Available: https://www.meta.com/gb/quest/quest-3/.

[2] *Introducing Apple Vision Pro: Apple's first spatial computer*, Jun. 2023. [Online]. Available: https://www.apple.com/newsroom/2023/06/introducing-apple-vision-pro/.

[3] M. Park, *3 keys to improving user retention in virtual reality*, Oct. 2017. [Online]. Available: https://venturebeat.com/games/3-keys-to-improving-user-retention-in-virtual-reality/.

[4] V. Weis, *3DOF, 6DOF, RoomScale VR, 360 Video and Everything In Between — Blog — Packet39*, 2018. [Online]. Available: https://packet39.com/blog/3dof-6dof-roomscale-vr-360-video-and-everything-in-between/.

[5] E. Callenbach, ": The Five C's of Cinematography: Motion Picture Filming Techniques Simplified . Joseph V. Mascelli.," *Film Quarterly*, vol. 19, no. 4, 1966, ISSN: 0015-1386. DOI: 10.1525/fq.1966.19.4.04a00320.

[6] N. Dodgson, "Autostereoscopic 3D displays," *Computer*, vol. 38, no. 8, pp. 31–36, Aug. 2005, ISSN: 0018-9162. DOI: 10.1109/MC.2005.252. [Online]. Available: http://ieeexplore.ieee.org/document/1492263/.

[7] C. Gurrin, A. F. Smeaton, and A. R. Doherty, "LifeLogging: Personal Big Data," *Foundations and Trends® in Information Retrieval*, vol. 8, no. 1, pp. 1–125, 2014, ISSN: 1554-0669. DOI: 10.1561/1500000033. [Online]. Available: http://www.nowpublishers.com/articles/foundations-and-trends-in-information-retrieval/INR-033.

[8] M. Dimiccoli and P. Radeva, “Visual Lifelogging in the Era of Outstanding Digitization,” *Digital Presentation and Preservation of Cultural and Scientific Heritage*, vol. 5, pp. 59–64, Sep. 2015, ISSN: 2535-0366. DOI: 10.55630/dipp.2015.5.4. [Online]. Available: https://dipp.math.bas.bg/dipp/article/view/dipp.2015.5.4.

[9] E. Waisberg, J. Ong, M. Masalkhi, *et al.*, “Meta smart glasses—large language models and the future for assistive glasses for individuals with vision impairments,” *Eye*, 2023, ISSN: 1476-5454. DOI: 10.1038/s41433-023-02842-z. [Online]. Available: https://doi.org/10.1038/s41433-023-02842-z.

[10] A. Orchard, M. O’Gorman, C. La Vecchia, and J. Lajoie, “Augmented Reality Smart Glasses in Focus: A User Group Report,” in *CHI Conference on Human Factors in Computing Systems Extended Abstracts*, New York, NY, USA: ACM, Apr. 2022, pp. 1–7, ISBN: 9781450391566. DOI: 10.1145/3491101.3503565. [Online]. Available: https://dl.acm.org/doi/10.1145/3491101.3503565.

[11] *Spectacles by Snap Inc. Spectacles 3*, 2023. [Online]. Available: https://www.spectacles.com/uk/shop/spectacles-3.

[12] D. Marr and T. Poggio, “A computational theory of human stereo vision,” *Proceedings of the Royal Society of London. Series B. Biological Sciences*, vol. 204, no. 1156, pp. 301–328, May 1979, ISSN: 0080-4649. DOI: 10.1098/rspb.1979.0029. [Online]. Available: https://royalsocietypublishing.org/doi/10.1098/rspb.1979.0029.

[13] M. S. Banks, J. C. A. Read, R. S. Allison, and S. J. Watt, “Stereoscopy and the Human Visual System,” *SMPTE Motion Imaging Journal*, vol. 121, no. 4, pp. 24–43, May 2012, ISSN: 1545-0279. DOI: 10.5594/j18173. [Online]. Available: https://ieeexplore.ieee.org/document/7308394/.

[14] K. Matzen, M. F. Cohen, B. Evans, J. Kopf, and R. Szeliski, “Low-cost 360 stereo photography and video capture,” *ACM Transactions on Graphics*, vol. 36, no. 4, pp. 1–12, Aug. 2017, ISSN: 0730-0301. DOI: 10.1145/3072959.3073645. [Online]. Available: https://dl.acm.org/doi/10.1145/3072959.3073645.

[15] R. Aggarwal, A. Vohra, and A. M. Namboodiri, “Panoramic Stereo Videos with a Single Camera,” in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, vol. 2016-December, IEEE, Jun. 2016,

pp. 3755–3763, ISBN: 978-1-4673-8851-1. DOI: 10.1109/CVPR.2016.408. [Online]. Available: http://ieeexplore.ieee.org/document/7780777/.

[16] *Kandao QooCam Ego 3D Camera*. [Online]. Available: https://www.kandaovr.com/qoocam-ego/.

[17] Umar Shakir, *The iPhone 15 Pro is getting spatial video capture in iOS 17.2*, Nov. 2023. [Online]. Available: https://www.theverge.com/2023/11/9/23954310/apple-iphone-15-ios-17-beta-spatial-video-vision-pro.

[18] *Apple unveils iPhone 15 Pro and iPhone 15 Pro Max*, Sep. 2023. [Online]. Available: https://www.apple.com/newsroom/2023/09/apple-unveils-iphone-15-pro-and-iphone-15-pro-max/.

[19] M. Mühlhausen, M. Kappel, M. Kassubeck, *et al.*, "Immersive Free-Viewpoint Panorama Rendering from Omnidirectional Stereo Video," *Computer Graphics Forum*, vol. 42, no. 6, 2023, ISSN: 14678659. DOI: 10.1111/cgf.14796.

[20] R. Pagés, K. Amplianitis, D. Monaghan, J. Ondřej, and A. Smolić, "Affordable content creation for free-viewpoint video and VR/AR applications," *Journal of Visual Communication and Image Representation*, vol. 53, 2018, ISSN: 10959076. DOI: 10.1016/j.jvcir.2018.03.012.

[21] A. Serrano, I. Kim, Z. Chen, *et al.*, "Motion parallax for 360° RGBD video," *IEEE Transactions on Visualization and Computer Graphics*, vol. 25, no. 5, 2019, ISSN: 19410506. DOI: 10.1109/TVCG.2019.2898757.

[22] J. Armstrong, *"The Entire History of You"*, 2011.

[23] A. Duane, B. P. Jónsson, and C. Gurrin, "VRLE: Lifelog Interaction Prototype in Virtual Reality: Lifelog Search Challenge at ACM ICMR 2020," in *LSC 2020 - Proceedings of the 3rd Annual Workshop on the Lifelog Search Challenge*, 2020. DOI: 10.1145/3379172.3391716.

[24] H. Chen, R. Shi, D. Monteiro, N. Baghaei, and H. N. Liang, "VRCockpit: Mitigating Simulator Sickness in VR Games Using Multiple Egocentric 2D View Frames," in *IEEE Conference on Computatonal Intelligence and Games, CIG*, vol. 2022-August, 2022. DOI: 10.1109/CoG51982.2022.9893678.

[25] S. H. Park and G. C. Lee, "Full-immersion virtual reality: Adverse effects related to static balance," *Neuroscience Letters*, vol. 733, 2020, ISSN: 18727972. DOI: 10.1016/j.neulet.2020.134974.

[26] K. Lim, J. Lee, K. Won, N. Kala, and T. Lee, "A novel method for VR sickness reduction based on dynamic field of view processing," *Virtual Reality*, vol. 25, no. 2, 2021, ISSN: 14349957. DOI: 10.1007/s10055-020-00457-3.

[27] A. S. Fernandes and S. K. Feiner, "Combating VR sickness through subtle dynamic field-of-view modification," in *2016 IEEE Symposium on 3D User Interfaces, 3DUI 2016 - Proceedings*, 2016. DOI: 10.1109/3DUI.2016.7460053.

[28] O. M. Zarka and Z. J. Shah, "Virtual Reality Cinema: A Study," *IJRAR-International Journal of Research and Analytical Reviews*, vol. 3, no. 2, 2016.

[29] *OpenXR is a royalty-free, open standard that provides high-performance access to Augmented Reality (AR) and Virtual Reality (VR)—collectively known as XR—platforms and devices.* [Online]. Available: https://www.khronos.org/OpenXR/.

[30] *FFMPEG A complete, cross-platform solution to record, convert and stream audio and video.* [Online]. Available: https://ffmpeg.org/.

[31] *Godot 4.2 Game Engine Docs.* [Online]. Available: https://docs.godotengine.org/en/4.2/.

[32] Malcolmnixon, *Malcolmnixon / godot-stereo-video Github Repository*, Nov. 2023. [Online]. Available: https://github.com/Malcolmnixon/godot-stereo-video.