Electronics and Computer Science

Faculty of Engineering and Physical Sciences

University of Southampton

Muhammad Hazimi Bin Yusri

24/11/2023

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

Project supervisor: Tom Blount

Second examiner: <2nd Examiner>

A project progress report submitted for the award of

<MEng Electrical and Electronics Engineering>

Abstract

In the realm of virtual reality (VR) and lifelogging, this project endeavors to overcome barriers of exclusivity and cost by developing an open-source, low-cost, and modular stereo video camera system. Designed to be modular with first design to stay on top of a cap, this system integrates lightweight cameras and a microphone with the Raspberry Pi Pico microcontroller. It offers efficient stereoscopic (3D) video capture and immersive surround sound recording. Complementing the hardware, the project entails the development of lifelogging VR software using the Godot game engine. This includes a side-by-side (SBS) video player and intelligent metadata auto-tagging through scene and object detection. The primary objective is to democratize VR content creation, making it accessible to a broad audience, from VR enthusiasts to content creators, encouraging innovation in VR and lifelogging. Challenges, such as technical complexities and power management, are addressed through rigorous prototyping and optimization, ensuring project success and fostering inclusivity, innovation, and the advancement of VR content creation technology in the field of lifelogging.

Abstract should be written last according to what u wrote to reflect rest of document, progress report abstract is different than end project report abstract.

Typical structure is like inverted pyramid)

- Introduction (problem statement, scope, goals)

- Background of existing technologies (can separate/structure however I want), and its considerations (where most reference comes from are research, academic papers)

dont need to be exactly like what im doing, separate it for example (lifelogging and virtual reality), and talk about how ur design is combination of both on next section

- Design, who its for, what, why? (Can be moscow requirement)

Circuit diagram, codes when its existing in report, acquisition, budget and cost analysis.

Make sure all diagrams/pictures are related to paragraph its in. Dont need to be my images/pictures for reference, but make sure you got the rights for it.

(usually larger and more important one)

- Project management, how u manage time (gantt chart), risk assessment, how do u plan of next half of project.

max 3000 words (not including references)

if enough space

-conclusion of what i do and what i do next

content are more important than heading names

Contents

[Abstract 2](#_Toc151382824)

[Contents 3](#_Toc151382825)

[1 Introduction 4](#_Toc151382826)

[2 Background and report of literature search 5](#_Toc151382827)

[3 Report on Technical Progress 6](#_Toc151382828)

[4 Plan of remaining work 7](#_Toc151382829)

[References 8](#_Toc151382830)

# Introduction

## Problem Statement

The recent proliferation of Virtual Reality (VR) Head Mounted Display (HMD) technologies has ushered in a new era of creativity, offering users immersive experiences. However, the formidable barriers of exclusivity and high costs associated with current solutions for recording stereoscopic/3D/spatial video constrain the full potential of these technologies. This issue assumes critical significance for the broader adoption of VR, as low user retention on the platform is often attributed to the scarcity of exclusive content. The redundancy of viewing conventional 2D content through VR headsets, given the superior displays and audio capabilities of other mainstream devices such as TVs, phones, and tablets, underscores the urgent need for a solution. Empowering the general consumer to effortlessly create their own stereoscopic video content not only enhances the appeal and novelty of VR but also addresses the demand for exclusive and engaging material.

Traditional approaches to VR content creation often prioritize factors like Field of View (FOV) and 3 Degrees of Freedom (3DoF), resulting in 180-degree or 360-degree videos. However, these videos often lack a compelling depth effect, making them visually uninteresting. Moreover, directing such videos becomes challenging as viewers can look in any direction, diminishing the directive control from the content creator. Furthermore, higher FOV requirements necessitate increased resolution to meet acceptable Pixel-Per-Degree (PPD) resolution, leading to elevated hardware costs. This is because higher FOV videos are mapped into larger surfaces, decreasing their perceived resolution.

## Goals

This project endeavors to surmount these challenges by:

1. **Developing an Open-Source, Cost-Efficient, and Modular Stereo Video Camera Hardware System:**
   * Introducing a hardware solution that is accessible to a broader audience, mitigating the cost barrier associated with existing options.
2. **Implementing a Pre-processing Pipeline:**
   * Mixing two 2D videos into the correct stereoscopic Side-By-Side (SBS) format to achieve an authentic depth effect.
   * Synchronizing audio files with the correct channels to create an immersive surround soundscape.
   * Incorporating metadata tagging through object and scene detection, facilitating streamlined browsing and organization of content.
3. **Creating Intuitive VR Software for Content Management:**
   * Designing software that enables seamless file browsing and content viewing within the VR environment.

This project not only addresses the critical gaps in hardware accessibility and the pre-processing pipeline but also represents a significant enhancement to existing methodologies. 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.

# Background and report of literature search

Why lifelogging?

Just old terms for current habits of social media, only difference is lifelogging is more methodical and routine. Also bring a good point for comparison instead of just randomly taking videos/photos every now and then. Inspiration from black mirror episode to ‘preserve’ and record lifetime memories, especially in immersive mode like SBS, just how like we see the world, thus no need 6DoF as we are not looking around us 360 all the time, having better FOV (human fov around 220? Need cite), would be cool but too expensive for now.

Why VR?

VR is virtual reality, spatial computing, space, 3D, emulating real life etc.

More emotional attachment as the content is more immersive/real, feels like you are there physically.

Why windowed instead of 180/380?

Why SBS instead of FVV/reconstruction?

Limitations of current VR HMD hardware for real FVV, which is more immersive and uses 6DoF and have depth but current hardware for VR themselves is not good enough yet, so basic SBS is sufficient enough [Affordable content creation for free-viewpoint video and VR/AR applications - ScienceDirect](https://www.sciencedirect.com/science/article/pii/S1047320318300683), 6.2.1. data format limitations.

This also can be seen by iphone 15 pro adoption of spatial video, which is only 1080p 60fps, which prove 1080p is good enough for now for window style viewing. Only problem is I need good synced 60fps.

Why VR software to browse content?

Can use the 6DoF with innovative UI design to its fullest, can have more screen space for loads of content, have a shelf of timeline etc, so many possibilities. (cite needed, one of paper tom emailed) + can use hand tracking for more intuitive

++ use eye tracking.

Why combine both?  
Novel and niche subset of both, but surely will become more adopted and mainstream in future, especially for personal videos/memories/photos etc.

# Report on Technical Progress

The initial design is to have it mounted on top of a cap, providing a more realistic point-of-view (POV) compared to chest-level strap design, but not constrained with too small of form factor like spectacles design.

IMAGES OF DESIGN 1,2,3 (CHEST, TOP CAP, SPECTACLES)

Main design is aimed towards lifelogger-esque/timelapse algorithm where pictures are taken at timed intervals.

There will be buttons for manual recording with LED light indicator.

The video also includes stereo audio from two individual electret microphone.

# Plan of remaining work

12 pt text for main body

References