[Date]

12ShiroY

AI Computer Vision

OCR Computer Science Project

Table of Contents

[Analysis 1](#_Toc517781149)

[The Problem 1](#_Toc517781150)

[The Stakeholders 1](#_Toc517781151)

[Existing Solutions 1](#_Toc517781152)

[The Essential Features 1](#_Toc517781153)

[Potential Limitations 2](#_Toc517781154)

[Hardware and Software Requirements 2](#_Toc517781155)

[Success Criteria 2](#_Toc517781156)

[Design 4](#_Toc517781157)

[Decomposition 4](#_Toc517781158)

[Structure 4](#_Toc517781159)

[Algorithms 4](#_Toc517781160)

[Usability 4](#_Toc517781161)

[Variables, Data Structures and Classes 4](#_Toc517781162)

[Iterative Development 4](#_Toc517781163)

[Further Data 4](#_Toc517781164)

[Developing 4](#_Toc517781165)

[Iterative Developments 4](#_Toc517781166)

[Prototypes 4](#_Toc517781167)

[Structures 4](#_Toc517781168)

[Examples of Code (Annotated) 4](#_Toc517781169)

[Variable Names 4](#_Toc517781170)

[Validations 4](#_Toc517781171)

[Review 4](#_Toc517781172)

[Evaluation 4](#_Toc517781173)

[Success Criteria 4](#_Toc517781174)

[Development In Future 4](#_Toc517781175)

[Future Improvements 4](#_Toc517781176)

[Usability 4](#_Toc517781177)

[Bibliography 4](#_Toc517781178)

# Analysis

## The Problem

The program I wish to investigate in the program, is the potential abilities of computer vision, to complete tasks which could be applied to the real world. In this case, different types of sports. Above all preference, I would like to create an AI with the ability to find an object in its field of view, and track it, as it travels through its field of view.

In order to do this, I will need to take into consideration the constantly changing light environment, bounce detection and true/false positive ball candidates. I will most likely use a bright luminous yellow tennis ball (as seen on the left), as this has a very clear and bright colour and a very specific spherical shape, making the object unique, allowing me to decrease the room for error in the computer vision by reducing the chances of detecting false positives. I will do this by linking a Pi-Camera to a raspberry pi, and creating a program, using public libraries online, to achieve this.

I believe this can attract a wider range of people to become interested in computer science, and possibly pursuit this as a career. As this product targets people who are interested in computer science and/ or sports, this can allow people who may normally never consider computer science as a subject of study (or possibly even career) to at least think about this.

I will attempt to make this program as accurate and efficient as possible, recording each prototype as I develop this.

## The Stakeholders 1.1

Currently, my stakeholders are people who are interested in computer vision, who would like to see this similar idea be used to track any ball within sports. This could make life much easier for camera crews in sports in the future, and much more cost effective as this eliminates a significant amount of skilled human work, which could be replaced by artificial intelligence. My stakeholder age range is between the ages of 14 and 40, as this can range from students interested in sports and/ or film.

My main stakeholder is Tom Smith, who is a 15-year-old male, with an interest in both computer science and Sports. He would like to be able to have a camera, which tracks his tennis ball, while he is playing. Tom would also like to have this possibly implemented with an ability to record himself playing tennis, and possibly create content for a platform such as YouTube. The convenience of having a stakeholder similar to Tom is that he can provide very quick and accurate feedback frequently, allowing me to understand the possibility for improvement. This will allow me to build upon multiple prototypes and have a relevant member of the public to test my product from a separate view to me. This could allow me to make my hardware and/or software more ergonomic to the user.

This was the response from my stakeholder, Tom Smith, when I asked him 10 questions about the project.

1. What age range do you think this project is suitable for?  
   I would like it to be suitable for 15-16 year olds like me. Either gender
2. Do you find anything confusing about the idea for the project?  
   No.
3. Do you find anything confusing about the hardware?  
   No
4. Do you find anything confusing about the software?  
   I am not sure yet what the software would run on - that is for you to decide. As long as it can track the ball, I do not really mind.
5. Would you buy this product from a store?  
   Yeah, if it was relatively cheap and looked high quality. It'd be creating YouTube videos etc.
6. What would you want this product to be able to do?  
   It must be able to track a ball, hopefully a tennis ball, so I guess it should sense where the ball is in the field of view and track it. If it could learn from its mistakes that'd be great.
7. Do you see any potential application of this in real life?  
   Yeah, I see great potential for this project in the sports industry and potentially in the film industry, as this could save extreme amounts of money for companies spent on human resources.
8. How simple do you think that the user interface needs to be?  
   The user interface should be simple to use and must include things such as begin tracking; display a message when the ball is located in the field of view.
9. Would you like to use a voice operated system for this product or not?  
   That would be awesome!

## Existing Solutions 1.1

Hawk-Eye, developed by Paul Hawkins, is a computer system owned by Sony, used in many sports to visually track the trajectory of the ball and display a profile of its statistically most likely path as a moving image. This piece of technology is used for sports such as tennis, football, cricket and more. In summary, this is six very high performance cameras used to create a 3D representation of the ball’s trajectory, frame by frame. This then allows the use of complex mathematics to predict the movement of the ball based on its trail. It is not perfect, but it is pretty close, as Hawk-Eye is accurate to 3.6 millimetres.

### In-Depth Breakdown of Hawk-Eye

All information learnt, and screen shots used are from this video (“Visual tracking of a tennis ball” - <https://youtu.be/iRlWw8GD0xc>) on the 13/11/2018.

Assuming that ball candidates (what has been detected as a potential tennis ball) in each frame have already been detected, tracking the tennis ball is then broken down into which candidates, are object-originated (true positives), and which are clutter-originated (false positives).   
The candidates can be plotted in a row-column-time 3D space. The objective of this is to recover the class labels of the candidates. The approach used here is “a layered data association scheme.” A candidate triplet (three frames containing the potential tennis ball in each of them, very close to one another) is selected from the beginning of the sequence, which has an extremely high probability of continuing with only true positives. A dynamic model is then fitted, and then optimised recursively until convergence (the other candidates (near the triplet) seem to line up). This optimised model is called a “tracklet”. As a sliding window moves, a sequence of tracklets are generated. Simultaneously, a graph is constructed, where each node is a tracklet, and the edge difference between two nodes is defined according to the “compatibility” of the two tracklets. The graph is then sectioned into sub-graphs, and the optimal path in each graph is found. The desired data association result is then contained in the optimal paths. Next, interpolation and hit/bounce detection is used to create the final product, the tennis ball tracking computer vision.



## The Essential Features 1.1

The essential features of this project consist of a program with the ability to detect a tennis ball in its field of view, and then track this tennis ball as it moves across the room. Another key feature is for the program to be able to detect when the tennis ball bounces off any surface, which should be detected via an unnatural change of direction of the tennis ball.

## Potential Limitations 1.1

The price of this project is a major limitation as in the worst case scenario; I may have to spend over £100 on this project. However, I will try to avoid this by using few expensive components such as the raspberrypi and the picamera; self-made alternate hardware such as a homemade Google AIY kit. Another limitation is the complexity of the coding necessary to complete this task. This project requires an extremely high knowledge of python which I do not yet possess and this will mean that I will have to put in extra time at home to build my knowledge of this. Another possible limitation for the project would be if there was a flaw with the hardware (e.g. response time or frame rate of camera, faulty pins on the raspberry pi etc.). This could potentially limit the processing speed and overall, limit the ability of the camera to track the ball in the air. I will try to avoid this by keeping track of the computer usage throughout the evolution of the project, making sure nothing gets overloaded or potentially break, which could spawn other limitations such as price instantly.

## Hardware and Software Requirements 1.1

|  |  |  |
| --- | --- | --- |
| Number of Requirement | Hardware Requirements | Software Requirements |
|  | A raspberry pi – most likely a raspberry Pi zero as I am working on a limited budget but ideally, a Raspberry Pi 3 B+, as it has a variety of essential upgrades in comparison to the cheaper alternative. I would also need to buy essential parts to use the raspberry pi such as their power block, a micro SD card and potentially, a case (to prevent any damage being done to the raspberry pi). The features of the Pi Zero: BCM 2835 SOC @ 1GHz, 512MB of RAM, micro-SD, mini-HDMI, micro-B USB for data, micro-B USB for power, CSI camera connector (needs adaptor cable for an extra £5), Unpopulated 40-pin GPIO connector, Compatible with existing HAT add-ons, Dimensions: 65mm x 30mm x 5mm  The Features of the Pi B+: 1.4GHz 64-bit quad-core ARM Cortex-A53 CPU (BCM2837), 1GB RAM (LPDDR2 SDRAM), On-board wireless LAN - dual-band 802.11 b/g/n/ac (CYW43455), On-board Bluetooth 4.2 HS low-energy (BLE) (CYW43455), 4 x USB 2.0 ports, 300Mbit/s ethernet, 40 GPIO pins, Full size HDMI 1.3a port, Combined 3.5mm analog audio and composite video jack, Camera interface (CSI), Display interface (DSI), microSD slot, VideoCore IV multimedia/3D graphics core @ 400MHz/300MHz | AI – an evolutionary AI which learns how to move the car around the track, when it crosses the finish line and how to decrease the time taken to complete a lap. This AI will do this using computer vision and measuring the outputs via ammeter and voltmeter. |
|  | (possibly) Voice Recognition kit for Google Cloud Speech – this consists of 170 tie point mini breadboard, 10 male to male 20cm breadboard cables, MAX9812 amplified microphone, NPN transistor, red LED and two resistors (1k and 220ohm). | A UI – easily understandable and useable user interface, suitable for all the essential criteria of the stakeholder |
|  | The Raspberry Pi Camera -  “Raspberry Pi v2.1 8 MP 1080p Camera Module” is the perfect camera for me, however it is on the expensive side of the discussion, however the benefits of its features may outweigh the con of the price. It records in 1080p with 60 fps, meaning that the potential limitation of lack of frames is most likely eliminated here, making prediction of the path (of the ball travelling through the air) will be more accurate. It also has a wide angle lens, meaning that it can track a tennis ball significantly further in the environment. It is worth around £24 on Amazon.  “Raspberry Pi 3 2 model B B+ A+ Mini Camera Video Module 5MP 1080p OV5647 Sensor with 15 Pin FPC Cable + Pi Zero Ribbon Cable 15cm” is significantly cheaper than the v2.1 (only £11) however, it is limited to 30 fps and a regular flat lens. This could mean that it would be a waste of money for me to buy this product, as in my circumstance, I will definitely need more than 30 fps. | Faulty computer vision – if the computer vision cannot see where the ball is and how it travels through the air. This could cause major issues with my project and could mean that I would have to start again on working with finding true positives and false positives. |
|  | Currently no other hardware requirements in 1.1 | Currently no other software requirements in 1.1 |

## Success Criteria 1.1

|  |  |  |
| --- | --- | --- |
| Criteria | Explanation of Criteria | How Criteria is Met |
| 1) The program needs to understand where the ball is in the field of view | The program must be able to recognise the ball if it is in its field of view, and if it is, where bouts it is in each frame. | I will implement this by allowing the program to understand the HSV colour of the ball, and how it changes in differently lit environments. |
| 2) The program needs to understand when the ball bounces | The program must be able to understand whether the ball bounces and recognise when this occurs. | I will implement this by allowing the problem to recognise, when the path of the ball changes in a way that is considered not natural (caused by gravity) allowing it to then again predict the path of the ball quickly after the bounce. |
| 3) The program needs to track the ball | The program needs to be able to track the ball in its field of view. | I will implement this by using probability, after eliminating the false positives. This will be done by observing the location of the ball in each frame, and predicting the path of it through the field of view. |
| 4) The program needs to eliminate false positives | The program needs to be able to distinguish between false positives and true positives via probability, and eliminate the false positives. | I will implement this by using probability of potential ball candidates in each frame (the distance between the potential balls in each frame) to allow me to understand which candidates are true positives. |
| 5) The whole project needs to be suitable for both genders | The project needs to be simple, easy to use and suitable for both genders | I will implement this by keeping by project as easy to understand as possible for someone without any knowledge of computing. I shall also keep the colours and style based around brown, grey and black (general “boxy” and “homemade” aesthetic). |
| 6) The project needs to be suitable for ages between 15-16 | The project needs to be ergonomic for users and suitable for a wide variety of children/ teenagers of different ages | I will do this by keeping my product slick, easy to use and exciting. This will be achieved by not undermining the potential which this technology can achieve. |
| 7) (possibly) The program will use voice command | The program needs to use voice command in order for the user to turn the program on | I will do this by using the google AIY kit which comes with a microphone and basic voice commands already preinstalled. These can be altered to be specified to my specific situation |

# Design

## Structure

## Decomposition

My project breaks down into 2 main sections, software and hardware. These then break down further as shown in the hierarchy diagram above. This shows the basic “barebones” concept of my project, which shows how I could potentially dedicate time to each section, building it up to form the combined final product.

### Computer Vision

This is the most important part of the program itself, as without the ability for my program to detect a tennis ball candidate in its field of view, none of the other parts of the program can work as intended.

First of all, I will need to create a repository for the program to understand the rough shape of the ball and the HSV (Hue Saturation Value). This will allow the program to detect the ball candidates in a constantly changing environment (in terms of lighting). This will work hand in hand with the actual path detection and path prediction for tracking the ball.

### Tracking The Ball

#### Finding Ball Path

This is a very key stage in the object tracking as it will allow the program to determine whether the ball candidate is a true positive or a false positive. This will be done by taking 3 frames as soon as a ball candidate is detected on the screen. If the 3 frames contain 3 ball candidates in very similar position (tracklets), then the program will be able to determine, that this ball candidate is most likely a true positive. This will then be further checked if the ball candidate has actually remained in the same position in the camera’s field of view, if so, then this could possibly be a false positive (however, if not, the ball candidate is almost definitely a true positive). If no true positive is detected, then this will be repeated until one is found. The next part of the program is to use the trajectory of the ball, and the path detected, to determine the most probable predicted path of the ball.

#### Bounce Detection

This is a relatively simple part of the program, as this will require for the program to detect an unnatural change in the ball’s path. By using techniques from finding the ball path and predicting it, we can save a significant amount of time here as this can mainly be done by reusing code. This will be done by using the already detected and predicted path of the ball, and cutting this off (and repeating the whole process again) when an unnatural change in the ball’s trajectory occurs.

### User Interface

For this part of my project, I would like to keep as close to my stakeholder wishes as possible. Tom Smith requested that the program was gender neutral, suitable for a wide age range and simple yet easy to use. I would like for my UI to have a simple black and white colour scheme, with green and red boxes around the tennis ball which would be tracked. This part is very simple and the project doesn’t heavily rely upon the design of the UI.

## Algorithms

For this section, I will have three separate sections of pseudocode. The first section of my pseudocode will be to access the raspberry pi camera and display the video feed. The second section will be for the HSV mask, which will take the input (video feed) from the webcam and separate it into Hue Saturation and Value, helping me find the specific HSV of my tennis ball. The third section of pseudocode is to both detect and track the tennis ball, which would use both the shape of the tennis ball and the HSV. For my situation, the HSV is the main variable, which I will use to track the tennis ball by, as it is very easy to isolate the HSV colour values of a tennis ball.

### Accessing camera and displaying camera feed

#import packages  
import system.packages

#define camera resolution + fps and reference to raw camera capture  
camera = picamera()  
camera.resolution = (x,y)  
camera.framerate = 32  
rawcaprute = PiArray(camera, size =(x,y))

#camera warmup time  
time.to.warmup(1)

#capture frames from camera and loops until we stop running the program  
for frame in camera.capture\_continous(rawcapture, format = ‘bgr’, use\_video\_port= true):

#grab raw image  
image = frame.array

#show frame  
 cv2.img.show(‘frame’, image)

#clear stream and prepare for next frame  
 rawCapture.truncate(0)

#end loop  
break

#destroy all windows and release the video capture  
destroy.all.windows()  
camera.capture.release()

### HSV colour mask

#import packages  
import system.packages

#define camera resolution + fps and reference to raw camera capture  
camera = picamera()  
camera.resolution = (x,y)  
camera.framerate = 32  
rawcaprute = PiArray(camera, size =(x,y))

#open a window to define the upper and lower HSV boundaries  
open.namedwindow(‘HSV’)  
#lower range sliders  
createtrackbar(‘lowHue’, ‘HSV’, trackbar[0], 255, nothing)  
createtrackbar(‘lowSaturation’, ‘HSV’, trackbar[1], 255, nothing)  
createtrackbar(‘lowValue’, ‘HSV’, trackbar[2], 255, nothing)   
#upper range sliders  
createtrackbar(‘highHue’, ‘HSV’, trackbar[3], 255, nothing)  
createtrackbar(‘highSaturation’, ‘HSV’, trackbar[4], 255, nothing)  
createtrackbar(‘highValue’, ‘HSV’, trackbar[5], 255, nothing)

#create loop to get HSV values from UI sliders  
while true:   
#return the fps of camera  
 timeCheck = time.time()  
 #now get hsv values  
 lowHue = cv2.getTrackbarPos(‘lowHue’, ‘HSV’)  
 lowSaturation = cv2.getTrackbarPos(‘lowSaturation’, ‘HSV’)  
 lowValue = cv2.getTrackbarPos(‘lowValue’, ‘HSV’)  
 highHue = cv2.getTrackbarPos(‘highHue’, ‘HSV’)  
 highSaturation = cv2.getTrackbarPos(‘highSaturation’, ‘HSV’)  
 highValue = cv2.getTrackbarPos(‘highValue’, ‘HSV’)

#camera warmup time  
time.to.warmup(1)

#capture frames from camera and loops until we stop running the program  
for frame in camera.capture\_continous(rawcapture, format = ‘bgr’, use\_video\_port= true):

#grab raw image  
image = frame.array

#show frame  
 cv2.img.show(‘frame’, image)

# Convert the frame to HSV colour model.  
 frameHSV = cv2.cvtColor(frame, cv2.COLOR\_BGR2HSV)

# HSV values to define a colour range to create a mask from  
colorLow = np.array([lowHue,lowSat,lowVal])  
colorHigh = np.array([highHue,highSat,highVal])  
mask = cv2.inRange(frameHSV, colorLow, colorHigh)  
# Show the first mask  
cv2.imshow('mask-plain', mask)

im2, contours, hierarchy = cv2.findContours(mask, cv2.RETR\_TREE, cv2.CHAIN\_APPROX\_SIMPLE)

contour\_sizes = [(cv2.contourArea(contour), contour) for contour in contours]  
biggest\_contour = max(contour\_sizes, key=lambda x: x[0])[1]

# Show final output image  
cv2.imshow(‘HSV’, frame)

#clear stream and prepare for next frame  
 rawCapture.truncate(0)

k = cv2.waitKey(5) & 0xFF  
if k == 27:

#end first loop  
break  
#print fps  
print('fps - ', 1/(time.time() - timeCheck))

#end second loop  
break

#destroy all windows and release the video capture  
cv2.destroyAllWindows()  
vidCapture.release()

### Detecting and tracking the tennis ball

#import packages  
import system.packages

#define camera resolution + fps and reference to raw camera capture  
camera = picamera()  
camera.resolution = (x,y)  
camera.framerate = 32  
rawcaprute = PiArray(camera, size =(x,y))

#construct the argument parse and parse the arguments  
ap = argparse.ArgumentParser()  
ap.add\_argument("-v", "--video",  
 help="path to the (optional) video file")  
ap.add\_argument("-b", "--buffer", type=int, default=64,  
 help="max buffer size")  
args = vars(ap.parse\_args())

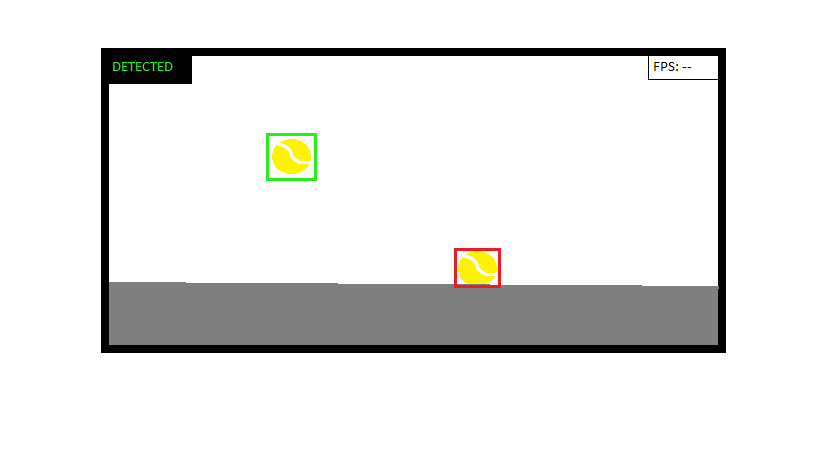
# define the lower and upper boundaries of the colour of the tennis ball in the HSV colour space, then initialize the list of tracked points  
greenLower = (best lower HSV boundaries found for tennis ball)  
greenUpper = (best upper HSV boundaries found for tennis ball)  
pts = deque(maxlen=args["buffer"])

#if a video path was not supplied, grab the reference  
#to the webcam  
if not args.get("video", False):  
 vs = VideoStream(src=0).start()

#Otherwise, grab a reference to the video file  
else:  
 vs = cv2.VideoCapture(args["video"])

#camera warmup time  
time.to.warmup(1)

## Usability



This is currently the way that I want my user interface to look like. As you can see, in the top left corner, there is a box which displays whether or not a tennis ball has been detected in the camera’s field of view. This will be green and state “DETECTED” when the tennis ball is detected and will be red, stating “NOT DETECTED” when the tennis ball is not detected. This is to make it obvious for the user to understand whether or not the object detection is working, similarly, this makes testing a lot easier. The tennis ball is tracked with a bright green box around it, which changes to red when the tennis ball bounces. In the top right corner, there is an fps counter which will mainly help me to see whether the raspberrypi is doing too much work, or is failing in some sort to achieve the necessary frame count. I want my fps count to be around 60.

I’ve tried to make my UI as simple as possible, to make it obvious for the user to understand what’s going on. Also, I have tried to make it suitable for both genders, and for a wide age range, by keeping the UI and the colours very basic, using black, green and red. I will also limit the program to only being able to detect one tennis ball at a time, limiting the chance of false positives being detected simultaneously as true positives.

## Variables, Data Structures and Classes

|  |  |  |
| --- | --- | --- |
| Variable / Data Structure Name | Type | What The Variable Does |
| BallShape | Array | Stores the shape of the ball from a repository of images of it. |
| BallColour | Variable – Integer | Stores variables of the HSV of the ball (Hue (dominant wavelength), Saturation (purity/ shades of the colour) and Value (intensity)) |
| BallCancidate | Record | Stores a record of different combinations of sizes of the ball and HSV values, to eliminate false positives |
| BallCandidateDetected | Variable - Boolean | Returns whether there are any ball candidates in its field of view |
| Bounce | Function | Responsible for sending information to the program that the path of the ball has changed in an unnatural way, meaning that a new path should be predicted now. However, the ball should still be tracked. |
| BounceDetected | Variable - Boolean | Returns whether there are any bounces detected. Does this by detecting whether the ball’s path changes in an unnatural game |
| PathAfterBounce | Function | Detects the path of the ball from 3 frames straight after the bounce, and begins to detect the balls path |
| PredictedPath | Function | From the detected path of the ball, detects ball trajectory and predicts the most probable path of the ball, as it travels through the air and more frames with the path of the ball are found |
| DetectedPath | Method | Uses probability strategies such as a candidate triplet, to create a tracklet, which grows as more frames of the ball are detected. |
| FalsePositive | Variable – String | Returns the value of ball candidate which has been detected as not a true positive, allowing the program not to include this in the tracking of ball candidate. |
| TruePositive | Variable – String | Returns a value of the ball candidate which has been detected as not a false positive, allowing the program to display this ball in the output. |

## Test Data

For this section of my project, I intend to create some ideas for testing my first prototype. Once the first prototype is completed, I will fill in this test table with findings and evidence via screenshots.

|  |  |  |
| --- | --- | --- |
| Test | Expected Outcome | Actual Outcome |
| Is the ball detected | Yes the ball is detected |  |
| Is a message given when the ball is detected | Yes the box in the top left of the form states “DETECTED” in green when the ball is detected |  |
| Is a message given when the ball is not detected | Yes the box in the top left of the form states “NOT DETECTED” in red when the ball is not detected |  |
| Is the ball tracked with a green box | Yes the ball is tracked with a green box |  |
| Does the box change to red when the ball bounces | Yes the box which tracks the ball changes to red when the ball bounces |  |
| Does the box change the colour back to green after the bounce is detected | Yes the box which tracks the ball changes back to green when after the bounce is detected |  |
| Does the fps counter display the fps | Yes the fps counter works effectively and accurately |  |
| Is there screen tearing while the ball is being tracked | No there is no screen tearing while the ball is being tracked, the camera works perfectly fine and the image doesn’t seem “jagged” at any point during the tracking |  |
| Is there screen tearing while the ball is not being tracked | No there is no screen tearing while the ball is not being tracked, the camera works perfectly fine and the image doesn’t seem “jagged” at any point |  |

## Iterative Development

## Further Data

# Developing

## Iterative Developments

## Prototypes

## Structures

## Examples of Code (Annotated)

## Validations

## Review

# Evaluation

## Success Criteria

## Potential in Future

## Future Improvements

## Usability

# Bibliography