

High-speed moving object detection and analysis using event camera

Final Year Project Report

SUPERVISED BY:

LINDSAY KLEEMAN

JONATHAN LI

Monash University

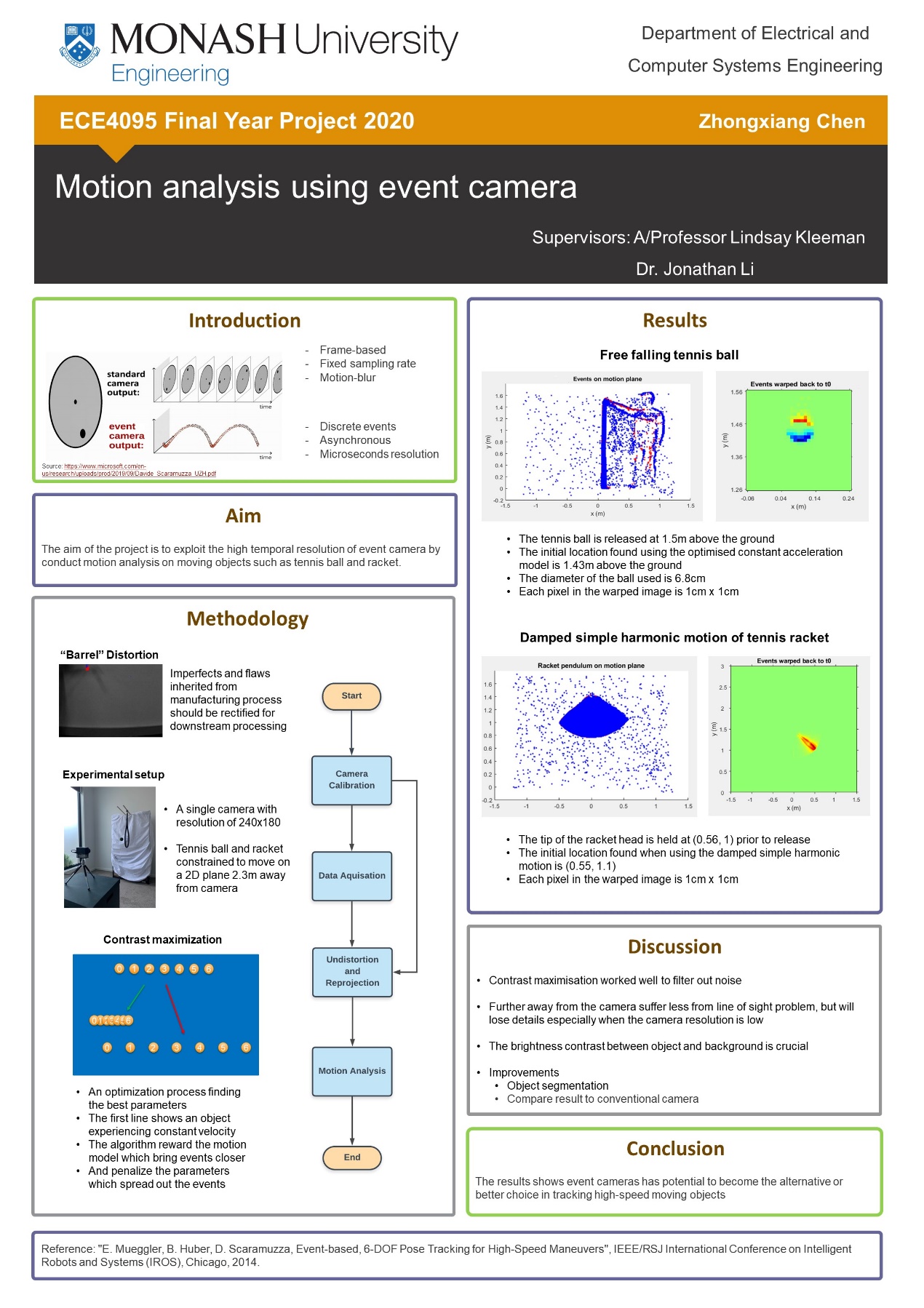
Department of Electrical and Computer Systems Engineering

Zhongxiang Chen

27321959

# Significant Contributions Page

* Designed and implemented hardware and software experimental setup for data acquisition using event camera.
* Performed and evaluated motion analysis results using contrast maximisation via two different approaches.



# Executive Summary

The aim of this project is to exploit the high temporal resolution of an event camera to track and perform motion analysis on high-speed moving objects. The project can be divided into three phases, the hardware and software implementation for data collection, the simple motion analysis of tennis ball and racket separately, and investigation of ball-racket collision.

Contrast maximisation is used as the primary algorithm because it uses all information provided by each event entry. And the results showed that the trajectory of freefalling tennis ball, damped pendulum motion of the racket are successfully reconstructed. In addition, the motion analysis with multiple objects moving simultaneously requires further studies.

The results of the project indicate that event cameras could become the alternative to traditional high-speed cameras in tacking moving objects.

# Acknowledgement

I would like to express my sincere thanks to Dr. Linkdsay Kleeman, for providing the opportunity to work on this project and guidance on design and implementation of the hardware and software.

Thanks to Dr. Jonathan Li for providing support both as the coordinator and supervisor.

A special thanks to Timo Stoffregen, who helped me to be familiarized with the device.

Thanks to everyone who had helped me with excellent advices and inspirations throughout the year.

Contents

[Significant Contributions Page I](#_Toc42989865)

[Executive Summary III](#_Toc42989866)

[Acknowledgement IV](#_Toc42989867)

[1.0 Introduction 1](#_Toc42989868)

[2.0 Literature review 2](#_Toc42989869)

[3.0 Overview 3](#_Toc42989870)

[3.1 Experimental Setup 3](#_Toc42989871)

[3.2 Hardware 4](#_Toc42989872)

[3.2.1 DAViS 240 C 4](#_Toc42989873)

[3.2.2 Tennis ball and racket 4](#_Toc42989874)

[3.2.3 Pivoting mechanism 5](#_Toc42989875)

[3.3 Software 6](#_Toc42989876)

[3.3.1 Camera driver 6](#_Toc42989877)

[3.3.2 Operating System 6](#_Toc42989878)

[3.3.3 MATLAB 6](#_Toc42989879)

[3.4 Project Tasks 6](#_Toc42989880)

[4.0 Detailed discussion 7](#_Toc42989881)

[4.1 Camera calibration 7](#_Toc42989882)

[4.2 Data acquisition 8](#_Toc42989883)

[4.3 Free fall of a tennis ball 9](#_Toc42989884)

[Data acquisition 9](#_Toc42989885)

[Undistortion and reprojection 10](#_Toc42989886)

[Motion Analysis 10](#_Toc42989887)

[Contrast maximisation 11](#_Toc42989888)

[4.4 Racket tracking 13](#_Toc42989889)

[Data acquisition 13](#_Toc42989890)

[Motion Analysis 14](#_Toc42989891)

[4.5 Advanced motion analysis 16](#_Toc42989892)

[5.0 Conclusions and Future Work 17](#_Toc42989893)

[References 18](#_Toc42989894)

Table of Figures

Figure 3.1. ................................................................................................................................ 3

Figure 3.2 ................................................................................................................................. 4

Figure 3.3 ................................................................................................................................. 4

Figure 4.1 ................................................................................................................................. 7

Figure 4.2 ................................................................................................................................. 8

Figure 4.3 ................................................................................................................................. 9

Figure 4.4 ............................................................................................................................... 10

Figure 4.5 ............................................................................................................................... 11

Figure 4.6 ............................................................................................................................... 12

Figure 4.7 ............................................................................................................................... 12

Figure 4.8 ............................................................................................................................... 13

Figure 4.9 ............................................................................................................................... 14

Figure 4.10.............................................................................................................................. 15

Figure 4.11 ............................................................................................................................. 16

# 1.0 Introduction

Event cameras, such as the such as dynamic vision sensor (DVS) [1] and dynamic and active-pixel vision sensor (DAVIS) had attracted attentions from vison and robotics research. Each pixel in the camera works independently and asynchronously which lead to an increase in the temporal resolution and greatly reduced bandwidth and power consumption. These benefits are more prominent in the tracking and motion analysis of high-speed moving objects.

The widely accepted Hawk-Eye system in sports is designed using multiple high-performance cameras which are synchronised to track the trajectory of the target object. For example, in tennis the system is used to assist the judgement of “in” or “out” of the landing. However, due to the limitation of the frame rate of the cameras, the trajectory of the ball between frames will be interpolated which causes potential errors to the reconstruction. Furthermore, tennis ball serves are usually extremely fast, which means the distance travelled by the ball between the frames are harder to be measured accurately.

The aim of the project is to exploit the high temporal resolution of an event camera to track and perform motion analysis on high-speed moving objects. Various trajectories of the tennis ball and racket are recorded and used for analysis and evaluation.

# 2.0 Literature review

Traditional cameras capture information in a scene by taking frames with a fixed time interval, whereas the event cameras collect the asynchronous information generate at each individual pixel locations where light intensity changed above a threshold. The resulted benefits including low latency, high dynamic range, and lower power consumption (0.67%) comparing to the traditional cameras [2]. The traditional approach of tracking an object in a video footage involves estimating the position of the object in each frame, then fitting a motion model to the estimated positions and are likely to encounter motion blur when relative velocity between the tennis ball and camera is large. A similar approach could be implemented with event camera: aggregate events with fixed time intervals or aggregate fixed number of events for each frame and then conduct motion analysis. The disadvantage of this approach is the temporal information carried by each event would be discarded [3][4][5]. Therefore, contrast maximization was used where all events will be warped according to an estimated model and a contrast function will be used to measure how well events agree with the predicted model [2]. The model parameters would be found by maximising the contrast using non-linear reward functions.

An existing problem in tennis playing is to prevent ‘tennis elbow’ occurring due to the low frequency vibrations transferred to the joint via the racket handle when hitting the ball [6]. Research has shown that hitting the tennis ball with different location of the stringbed will result in substantially difference in the amount of resultant force [7]. The resultant force is minimised when the tennis ball hit the region called “sweet spot” which is commonly located at the midpoint of longitudinal axis between the tip and the throat [6]. Also, hitting the ball within the sweet pot will give the maximum ball speed and power (rebound).

The event camera might be used as tool to collect useful statistical information for training and injury prevention in tennis because of its low latency and potential low cost when mass produced.

# 3.0 Overview

The main objective of the project is to perform motion analysis on the dataset collected using the DAViS 240C camera. The following sections describe the overview of experimental setup and the hardware and software considerations taken for the project.

Figure 3.1 Overview of the experimental setup

## 3.1 Experimental Setup

The flowchart in Figure 3.1 shows the high-level overview of the project. It shows all the processes necessary from camera calibration throughout to the motion analysis of tennis ball or/and racket.

The focus of the camera shall be adjusted at the beginning of a recording session and its intrinsic characteristics shall also be measured and noted for compensation of any flaws within the camera.

The asynchronous events and image frames sent from the camera will then be recorded into rosbag log files. The event arrays are then extracted from the ragbags into text file (txt) for computation and analysis in MATLAB.

The events are firstly undistorted and projected to the motion plane where the objects are moving on. Different motion models will be implemented and used for analysing the motion experienced by the objects.

## 3.2 Hardware

### 3.2.1 DAViS 240 C

This is the primary hardware device used in the project. The power consumption of this device is 180/145 mA at 5VDC with/without advanced photo system (APS). It comes with a single USB 2.0 I/O port which is utilised as both the power source and communication port. The array size is 240 x 180 in pixels and 4.44 x 3.33 in mm.

The original dimension of this device is H 56 x W 62 x D 28 mm, which is mounted to a 3D case with dimension of H 63 x W 67 x D 31 mm. The case is then attached to a tripod stand at the bottom.

Figure 3.2 DAViS 240 C

### 3.2.2 Tennis ball and racket

A generic tennis racket with the following specifications: a total length of 68.6 cm from the top of the head to the bottom of the handle, a head size of 729 cm2 and unstrung weight of 253 g. The majority parts of the racket have a thickness smaller than 4 cm.

Figure 3.3 Tennis ball and racket, painted in black

A generic tennis ball with diameter in range of 6.54–6.86 cm (2.57–2.70 inches), the sizes of the tennis balls used in the project will be continuously measured.

### 3.2.3 Pivoting mechanism

Two wooden rod of diameter 16mm are secured on the top of a shelf by using saddle clips. The top of the shelf is 1465mm above the ground and the rods are placed 106 mm away centre to centre. The racket is then fixed on to one of the rods using a saddle clip and strings while the ball is hanged from the other rod and lies on the same height as the centre of the racket.

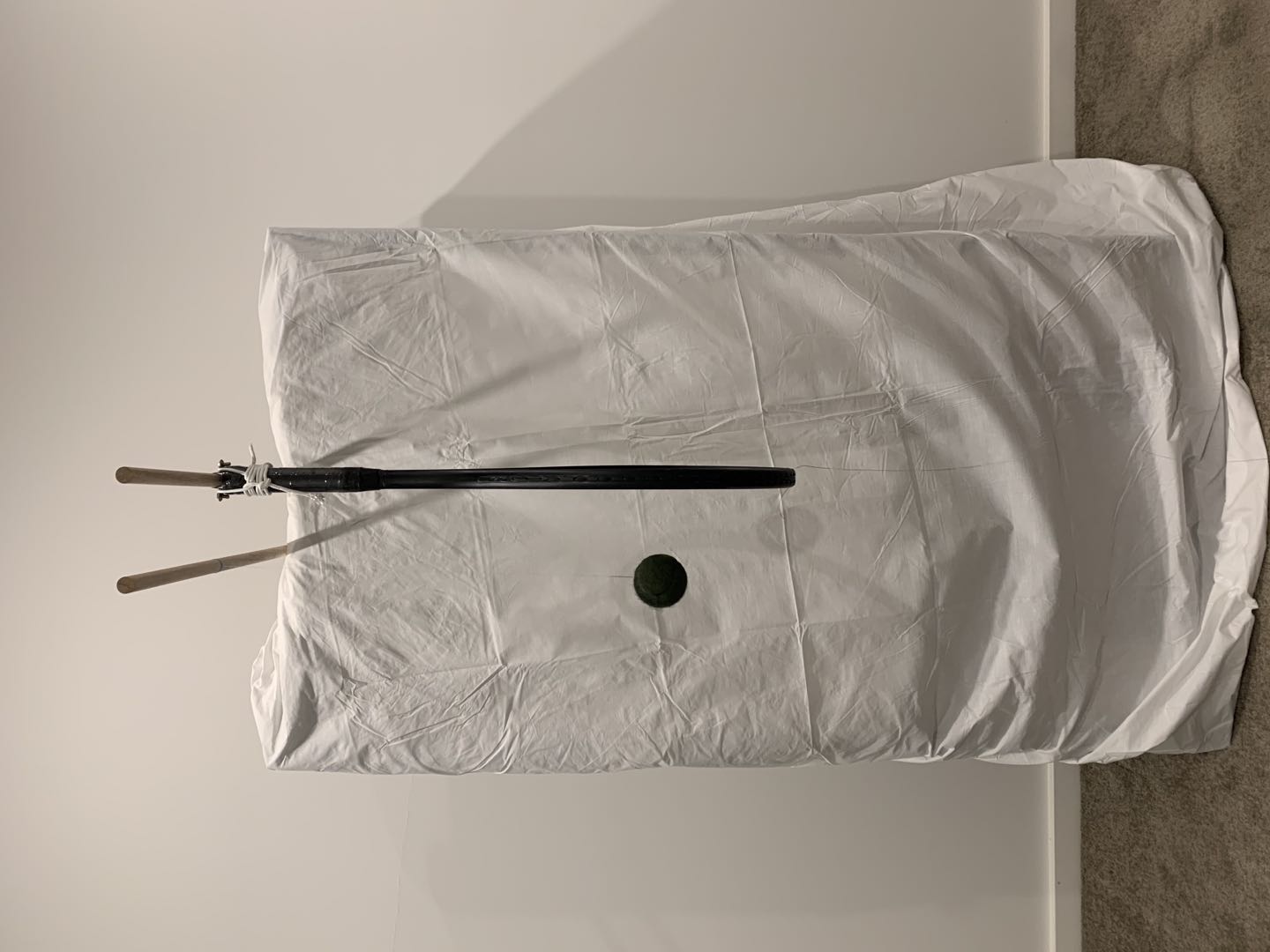


Figure 3.4 Hardware setup

As shown in Figure 3.4 above, the tip of the racket is 760 mm above the ground and the centre of the tennis ball is 950 mm above the ground.

## 3.3 Software

### 3.3.1 Camera driver

The rpg\_dvs\_ros paclage was used as the driver for the DAViS 240 C camera. The driver provides a GUI for rendering the output of the camera including both the image and events in C++. The driver also allows playback of pre-recorded datasets in rosbag file type.

### 3.3.2 Operating System

The rpg\_dvs\_ros is built and tested on Ubuntu only along with ROS (Robot Operating System) dependencies. Therefore, all the software components are developed under Ubuntu 16.04 LTS.

### 3.3.3 MATLAB

MATLAB contains powerful processing and optimisation toolboxes which satisfy the needs of all stages of the project.

## 3.4 Project Tasks

The structure of this project has three main phases which are listed below:

First is hardware and software development required for data acquisition.

Second, simple motion study which includes the motion analysis of a freefalling tennis ball and the motion analysis of rotation around a fixed axis of a tennis racket

Lastly, the dynamics of the tennis ball and racket during collision will be investigated.

# 4.0 Detailed discussion

## 4.1 Camera calibration

The events captured by the event camera are send to ROS as a vector with four dimensions.

*e* = [ *t, x, y, p* ], where *t* is the timestamp in seconds, (*x, y*) is the location of the event in pixels in horizontal and vertical directions respectively, and *p* is the polarity of the event indicating whether the brightness was increased or decreased.

All cameras suffer from flaws caused during manufacturing process. Therefore, camera calibration is important to reduce the effect those imperfects from causing errors in downstream computation and calculations. As seen in Figure 4.1, the curved line in the bottom of the camera is a straight line in real world. Thus, the camera suffers from positive radial distortion, which is also known as “barrel” distortion. The skew and shifting of the camera are not observable and will be determined according to the calibration results.

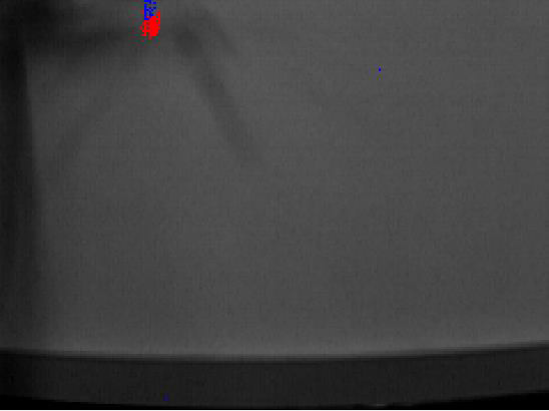


Figure 4.1 A snapshot of the recorded image by event camera

Two camera models are involved in the camera calibration process: the pinhole camera model [8] and lens distortion [9]. The pinhole camera model assumes a fixed focus length of a camera and the camera projects real world points on to an image plane. The camera matrix shall be found to describe all intrinsic parameters of the camera including the actual pixel location of principal point on image plane, the x and y-directional focal length, and the skew ration. The distortion coefficients describe the deformation caused by the lens of the camera which include the radial and tangential distortions.

The *camera\_calibration* in ROS package is used for camera calibration with the following steps:

1. Prepare a 7x9 checkerboard with 20x20 mm square size.
2. Start the camera driver
3. Link the image output of the event camera to the calibration code
4. Move the checkerboard around the scene, perform motions in as many degrees as possible while having the checkerboard facing the camera
5. Save the calibration results for the rectification process.

## 4.2 Data acquisition

The event camera is mounted at a stationary location with both the x and z axes horizontally aligned, and y axis vertically aligned. As shown in Figure 4.2 below.



Figure 4.2 Experimental setup for data acquisition

Throughout the project, the location of camera lens will be regarded as the optic centre of the camera. The elevation to ground was measured to be 0.84m and the distance from the lens to the image plane is 2.34m.

The RPG\_ROS\_DVS package [10][11][12] was utilised to drive the event camera and store the frames and events generated in to rosbag log files (rosbag). The event arrays are then extracted from the ragbags into text file (txt) for computation and analysis in MATLAB.

The recording is started prior to the desired motion of the objects. Therefore, unnecessary motions such as moving the objects to the desired location will be removed before computation.

## 4.3 Free fall of a tennis ball

The free fall of a tennis ball was selected to be tested first for its well understood motion model. A falling object due to gravity will experience a constant acceleration at 9.81 m/s^2 at the surface of Earth.

### Data acquisition

The experiment is setup as described in section 4.2, where the height of the camera is 0.84m above the ground and a tennis ball is released at the height of 1.5m above the ground on a motion plane 2.34m away from the camera. Figure 4.3 below shows the moment prior to the release.

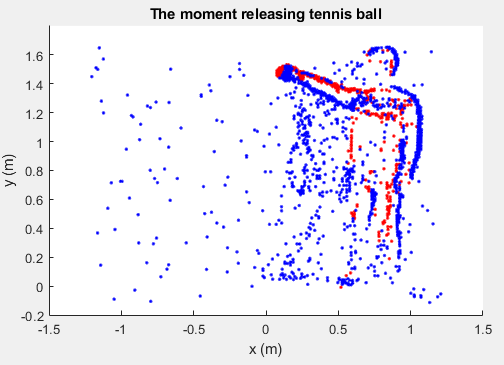


Figure 4.3 The Moment releasing tennis ball

### Undistortion and reprojection

The recorded events are then rectified using the camera calibration information and projected on to the motion plane using the calibration results generated prior to the recording session and measurements taken for the test. Figure 4.4 below shows a selection of the accumulated events after the release in original pixel image and their corresponding real-world coordinates. The ball is found to be released from 1.5m and bounces around 0m relative to the ground, indicating the rectification process was successful.

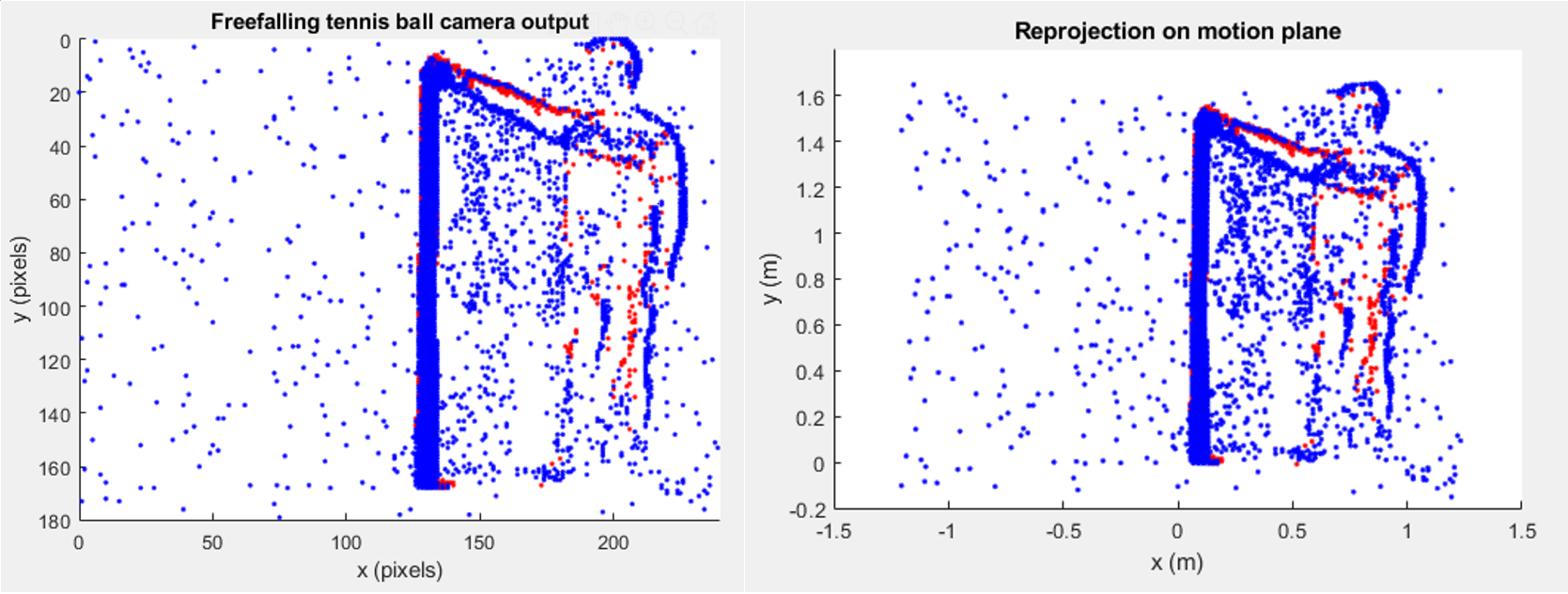


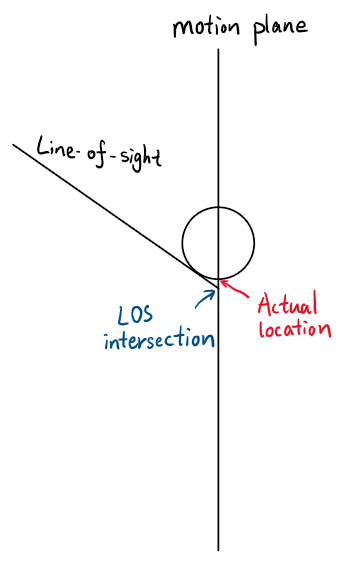
Figure 4.4 The rectification result using free falling tennis ball

### Motion Analysis

The terminal velocity of tennis ball is around 100 km/h, or 28m/s due to air resistance. However, the tennis ball was released from a height of 1.5 meters and the theoretical velocity before hitting the ground can be calculated using the following equation:

Substituting *vi* = 0, *a* = 9.81, *s* = 1.5 into the equation, *vf* is found to be 5.42 m/s.

Since the tennis ball is released at a relatively low height and the velocity prior to hitting the ground is 5.42 m/s which is much smaller than the terminal velocity, the effect of air resistance is removed from the modelling and analysis.



The diameter of a tennis ball is around 6.7cm which is smaller than 5% of the absolute distance between the tennis ball and the optic centre when the tennis ball is falling at 2.34 meters away from the optic centre. Therefore, any events generated on the intersections between the camera’s line of sight and the surface of the tennis ball is assumed to occur on the motion plane as shown in Figure 4.5.

Figure 4.5 The LOS intersection is assumed to be the actual location

### Contrast maximisation

Conventional computer vision algorithms are not readily applicable to asynchronous events generated by the dynamic vision camera. The frames of a conventional camera capture the scene with a certain timestamp. The instantaneous location of a moving object is then determined for that timestamp. The continuous motion model of the object can then be found by interpolating between the frames.

Contrast maximisation is a method which can exploit the asynchronous events by warping events back to their original position according to a motion model. The process could be done for arbitrary number of events and the motion model will be evaluated after the warping. The overlapping of events indicates a good prediction of the motion and a sparse warped image means the model requires further optimization. Therefore, it is intuitive to use a non-linear reward function which rewards the overlapping of events and penalise the model for spreading out the events.

Two methods are used to analyse the free fall of a tennis ball. The first works as a sliding window with fixed time length and step size while the second uses all the events provided to find a single optimized model to fit all events. The implementation and results will be further discussed in the following subsections.

#### Constant velocity reward function

This method assumes that the velocity of the tennis ball will not vary significantly over a small amount of time. The motion model is simplified to including only two parameters, vx and vy. vy is the velocity in y direction and vx is the velocity in x direction. The acceleration in y direction experienced by the tennis ball was estimated to be 9.86 ms-2 using the peak values in vy. It is noticeable when vy is close to 0m/s, the approximation is worse which is understandable as there are less events when the ball is moving slow.

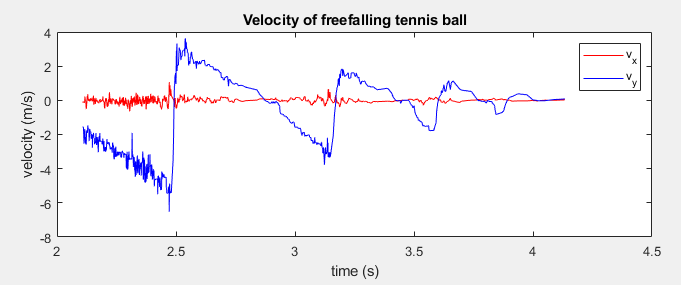


Figure 4.6 Results of using a constant velocity motion model for small sections of events

#### Constant acceleration reward function

This motion model assumes the tennis ball is experiencing a constant acceleration due to gravity. Three parameters are used for this model including *ay,* the acceleration in y direction, *tr*, the time of release, and vx, which takes into the account of initial horizontal velocity passed to the ball when it is released. Only the events prior to the first impact is used because the motion model will fail when the ball hit the ground.

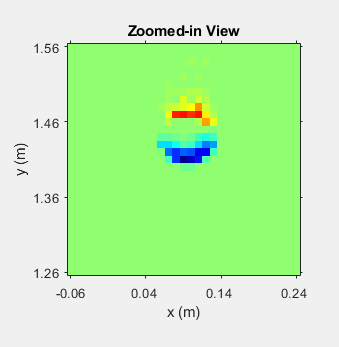
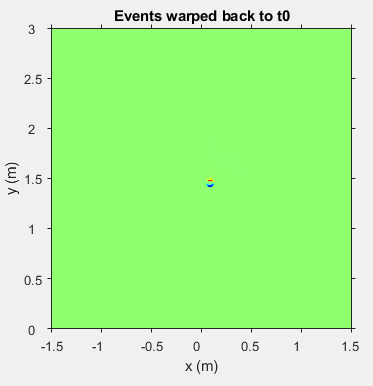


Figure 4.7 (L) Warping events back to their initial position; (R) A closer look

The performance is evaluated using the following measurements. Firstly, *ay* is found to be 9.96ms-2 which is close to theoretical acceleration of 9.81ms-2 due to gravity on Earth. Secondly, the initial position of the tennis ball found in the warped image is 1.43 m which is 0.07m away from the known 1.5m height. Thirdly, the diameter of ball in the warped image was 7cm which is 0.2cm larger than the measurement of the ball. Overall, the motion analysis of a free-falling tennis ball is accurately done using event camera.

## 4.4 Racket tracking

A similar experimental environment is used for racket tracking. The position of the midpoint of the rotational axis pass through the racket handle is measured in camera coordinate system. The racket is released from an elevated position and allowed to pivot freely.

### Data acquisition

The experiment is setup as described in section 4.3, where the height of the camera is 84 cm above the ground and a tennis racket is mounted on a rod 1.47m above the ground on a motion plane 234 cm away from the camera. Figure 4.8 below shows the moment prior to the release.

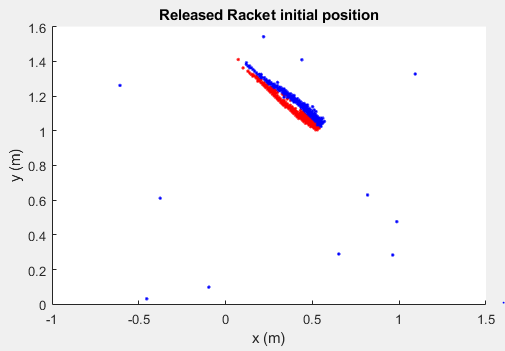


Figure 4.8 The moment releasing tennis racket

### Motion Analysis

The racket is fixed on the pivoting mechanism with the orientation which the camera can only see one side of the racket. The pendulum motion of the racket can be modelled using simple harmonic motion (SHM) especially when the racket is released at a small angle where the small angle approximation is accurate. The general expression of the angular velocity of SHM is

,

where A is the maximum amplitude, is the oscillation frequency, and decides the initial velocity at t=0.

Because friction force is present between the surface of the rod and the metal clip, therefore, a damping factor is added to the expression

Two methods are used to analyse the pendulum motion of a tennis racket. The first works as a sliding window with fixed time length and step size while the second uses all the events provided to find a single optimized model to fit all events. The implementation and results will be further discussed in the following subsections.

#### Constant angular velocity

This method assumes that the angular velocity of the tennis ball will not vary significantly over a small amount of time. The motion model is simplified to include only three parameters, va, *xp and yp*. va is the angular velocity, *xp and yp* are the coordinates of the pivot point on the motion plane. The result can be viewed in Figure 4.9 below, the oscillation frequency and damping factor were estimated to be 4.53 and 0.14 respectively.

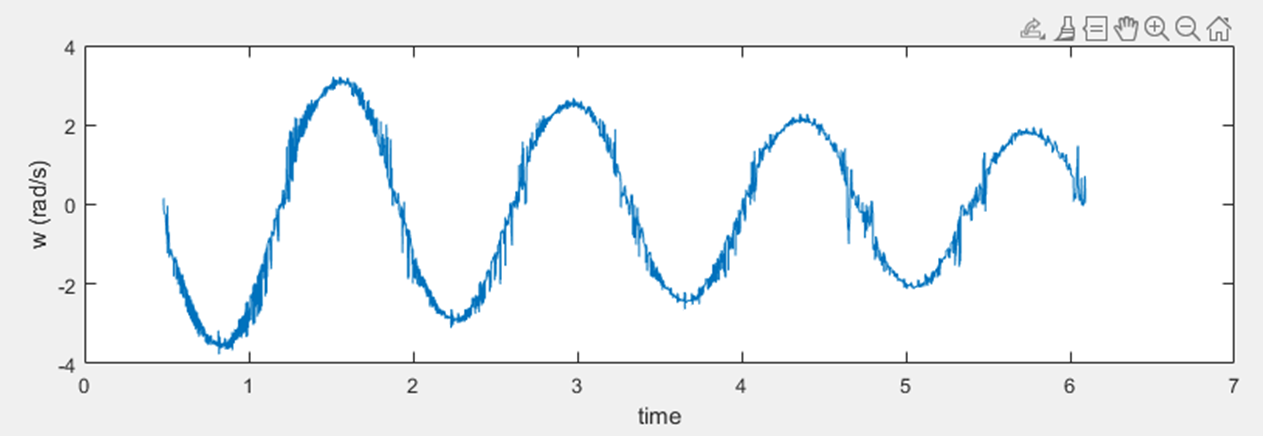


Figure 4.9 Results of using a constant angular velocity motion model for SHM of racket

#### Simple harmonic motion with damping

This motion model assumes the tennis racket underwent pendulum motion which is modelled as a damped simple harmonic motion. Five parameters are used for this model including *w*, the oscillation frequency, , a constant which shift the starting time of the motion, *d*, the damping factor and (*xp, yp*), which are the coordinates of the pivot point on the motion plane. The estimation from constant angular velocity method is used as the initial guess for this model.

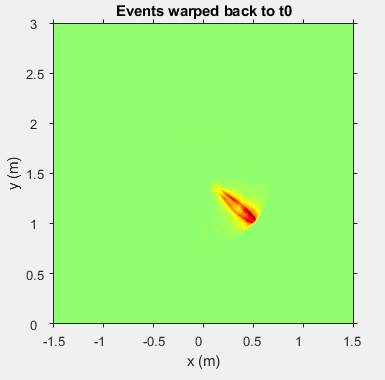


Figure 4.10 Warping events back to their initial position

Figure 4.10 above shows the estimated initial position of the racket by warping all events to t0, which denotes the moment tennis racket is released. The result is also very promising, the tip of the racket was measured to be 1.1 m above the ground and 0.55m away from the z-axis, while the actual values are 1.0m above the ground and 0.55m away from the z-axis . The width of the racket in the warped image is larger then expected, which could be caused by two factors, one being the pivoting mechanism is not perfect and the racket is slightly tilted, the other being the width of the pendulum is ignored when warping the events which would cause some error in the calculation.

## 4.5 Advanced motion analysis

This part of the project works on fitting two motion models for a tennis ball and tennis racket which moves simultaneously within the scene. The tennis ball is stationary initially and the racket is released at a known position. The collision will cause the tennis ball to start moving and slow the racket down.

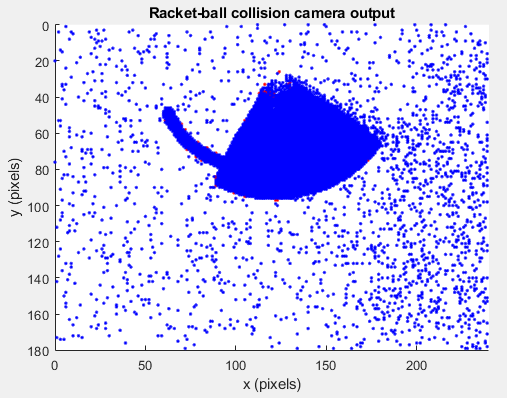
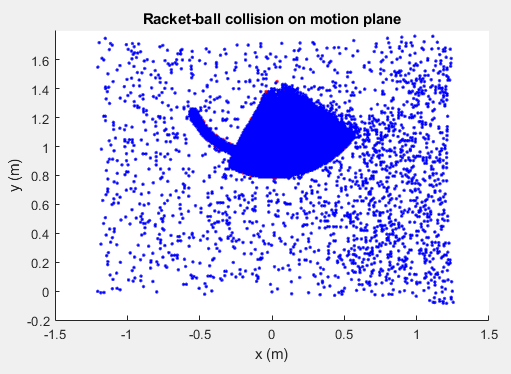


Figure 4.11 The rectification result using free falling tennis ball

The segmentation of the events was not successfully implemented thus no meaningful results can be shown. However, if more time is available one of the following approaches might solve the problem:

* Monitoring the loss of incoming events and compare that to a predefined model with no tennis ball presence, there should be a noticeable change in the loss when collision occurs.
* Using the image channel of the camera to perform crude segmentation as the exact locations of the ball and racket can be easily found using conventional vison algorithms, especially when the object is moving slowly

# 5.0 Conclusions and Future Work

The results of this project showed that event cameras have great potential to be used in object tracking and motion analysis. The contrast maximisation algorithm made it possible to exploit the high temporal resolution of the sensor and performed well to reconstruct the trajectory of the moving objects despite the limitations of hardware and software.

The following aspect of the project can be improved if more time and resources are available:

* If more than one camera could be used or the resolution of the camera is high enough, then the objects can be allowed to move with more degrees of freedom instead of limited to a 2-D plane.
* Perform the same motion analysis using the image output of the event camera or a high framerate conventional camera, then compare the results.
* Design and implement an algorithm to classify the newly arrived events to the object it belongs to simplify the motion analysis of multiple objects simultaneously.

The code repository could be accessed at <https://github.com/zche0011/FYP_event_camera>.

The video is also uploaded at <https://youtu.be/NMnG5PfeDrA>.

# References

1. P. Lichtsteiner, C. Posch and T. Delbruck, "A 128 X 128 120db 30mw asynchronous vision sensor that responds to relative intensity change," 2006 IEEE International Solid State Circuits Conference - Digest of Technical Papers, San Francisco, CA, 2006, pp. 2060-2069.
2. G. Gallego, H. Rebecq and D. Scaramuzza, "A Unifying Contrast Maximization Framework for Event Cameras, with Applications to Motion, Depth, and Optical Flow Estimation," 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, 2018, pp. 3867-3876.
3. C. Sulzbachner and J. Kogler, "A Load Balancing Approach for Silicon Retina Based Asynchronous Temporal Data Processing," 2010 36th EUROMICRO Conference on Software Engineering and Advanced Applications, Lille, 2010, pp. 431-435.
4. J. Kogler, C. Sulzbachner, M. Humenberger, and F. Eibensteiner. Address-event based stereo vision with bio-inspired silicon retina imagers. In Advances in Theory and Applications of Stereo Vision, 2011, pp. 165-188.
5. A. I. Maqueda, A. Loquercio, G. Gallego, N. García and D. Scaramuzza, "Event-Based Vision Meets Deep Learning on Steering Prediction for Self-Driving Cars," 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, 2018, pp. 5419-5427.
6. H. Brody, R. Cross and C. Lindsey, The physics and technology of tennis, United States Racquet Strings Association, Racquet Tech Publishing, California USA, 2002
7. L. Li, S. H. Yang, C. Hwang, and Y. S. Kim, “Effects of string tension and impact location on tennis playing,” Journal of Mechanical Science and Technology, vol.23, 2009, pp. 2990–2997.
8. Z. Zhang, "A flexible new technique for camera calibration," in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 22, no. 11, pp. 1330-1334, Nov. 2000.
9. J. Heikkila and O. Silven, "A four-step camera calibration procedure with implicit image correction," Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition, San Juan, Puerto Rico, USA, 1997, pp. 1106-1112.
10. E. Mueggler, B. Huber and D. Scaramuzza, "Event-based, 6-DOF pose tracking for high-speed maneuvers," 2014 IEEE/RSJ International Conference on Intelligent Robots and Systems, Chicago, IL, 2014, pp. 2761-2768.
11. P. Lichtsteiner, C. Posch and T. Delbruck, "A 128 ×128 120 dB 15 µs Latency Asynchronous Temporal Contrast Vision Sensor," in IEEE Journal of Solid-State Circuits, vol. 43, no. 2, pp. 566-576, Feb. 2008.
12. C. Brandli, R. Berner, M. Yang, S. Liu and T. Delbruck, "A 240 × 180 130 dB 3 µs Latency Global Shutter Spatiotemporal Vision Sensor," in IEEE Journal of Solid-State Circuits, vol. 49, no. 10, pp. 2333-2341, Oct. 2014.