

TABLE TENNIS EVENT DETECTION AND CLASSIFICATION
FROM MONOCULAR VIDEO SEQUENCES

A Doctoral Thesis

By

Kevin M. Oldham

Submitted in partial fulfilment of the requirements
for the award of Doctor of Philosophy of Loughborough University

copyright Kevin M. Oldham 2015

Abstract

It is well understood that multiple video cameras and computer vision (CV) technology can be used in sport for match officiating, statistics and player performance analysis. A review of the literature reveals a number of existing solutions, both commercial and theoretical, within this domain. However, these solutions are expensive and often complex in their installation. The hypothesis for this research states that by considering only *changes* in ball motion, automatic event classification is achievable with low-cost monocular video recording devices, without the need for 3-dimensional (3D) positional ball data and representation. The focus of this research is a rigorous empirical study of low cost *single* consumer-grade video camera solutions applied to table tennis, confirming that monocular CV based detected ball location data contains sufficient information to enable key match-play events to be recognised and measured. In total a library of 276 event-based video sequences, using a range of recording hardware, were produced for this research.

The research has four key considerations: i) an investigation into an effective recording environment with minimum configuration and calibration, ii) the selection and optimisation of a CV algorithm to detect the ball from the resulting single source video data, iii) validation of the accuracy of the 2-dimensional (2D) CV data for motion change detection, and iv) the data requirements and processing techniques necessary to automatically detect changes in ball motion and match those to match-play events. Throughout the thesis, table tennis has been chosen as the example sport for observational and experimental analysis since it offers a number of specific CV challenges due to the relatively high ball speed (in excess of 100kph) and small ball size (40mm in diameter). Furthermore, the inherent rules of table tennis show potential for a monocular based event classification vision system.

As the initial stage, a proposed optimum location and configuration of the single camera is defined. Next, the selection of a CV algorithm is critical in obtaining usable ball motion data. It is shown in this research that segmentation processes vary in their ball detection capabilities and location outputs, which ultimately affects the ability of automated event detection and decision making solutions. Therefore, a comparison of CV algorithms is necessary to establish confidence in the accuracy of the derived location of the ball. As part of the research, a CV software environment has been developed to allow robust, repeatable and direct comparisons between different CV algorithms. An event based method of evaluating the success of a CV algorithm is proposed. Comparison of CV algorithms is made against the novel Efficacy Metric Set (EMS), producing a measurable Relative Effi-

cacy Index (REI). Within the context of this low cost, single camera ball trajectory and event investigation, experimental results provided show that the Horn-Schunck Optical Flow algorithm, with a REI of 163.5 is the most successful method when compared to a discrete selection of CV detection and extraction techniques gathered from the literature review. Furthermore, evidence based data from the REI also suggests switching to the Canny edge detector (a REI of 186.4) for segmentation of the ball when in close proximity to the net.

In addition to and in support of the data generated from the CV software environment, a novel method is presented for producing simultaneous data from 3D marker based recordings, reduced to 2D and compared directly to the CV output to establish comparative time-resolved data for the ball location. It is proposed here that a continuous scale factor, based on the known dimensions of the ball, is incorporated at every frame. Using this method, comparison results show a mean accuracy of 3.01mm when applied to a selection of nineteen video sequences and events. This tolerance is within 10% of the diameter of the ball and accountable by the limits of image resolution.

Further experimental results demonstrate the ability to identify a number of match-play events from a monocular image sequence using a combination of the suggested optimum algorithm and ball motion analysis methods. The results show a promising application of 2D based CV processing to match-play event classification with an overall success rate of 95.9%. The majority of failures occur when the ball, during returns and services, is partially occluded by either the player or racket, due to the inherent problem of using a monocular recording device. Finally, the thesis proposes further research and extensions for developing and implementing monocular based CV processing of motion based event analysis and classification in a wider range of applications.

Keywords

Ball tracking, sport, CV, segmentation, table tennis, ping pong, feature extraction, Kalman Filter, object detection, Circle Hough Transform, image segmentation, automated scoring, automated coaching, Gaussian Mixture Model, optical flow, edge detection, monocular, video, event classification.

List of abbreviations

2D	2-dimension
3D	3-dimension
ANNs	artificial neural networks
BB	bounding box
CCD	charged coupled device
CHT	Circular Hough Transform
CMOS	complimentary metal-oxide semiconductor
CSV	comma separated values
CV	computer vision
dpi	dots per inch
ED	equivalent diameter
EMS	Efficacy Metric Set
EVH	Event Activity Hotspots
FoV	field of view
FPS	frames per second
FSM	Finite-state Machine
GLT	goal-line technology
GMM	Gaussian Mixture Model
HSV	hue-saturation-value
kph	kilometres per hour
MPEG	Moving Picture Experts Group
mph	miles per hour
ms⁻¹	metres per second
ms⁻²	metres per second squared
OF	Optical Flow
OOI	object of interest
ppi	pixels per inch
px	pixel
REI	Relative Efficacy Index
RoI	Region of Interest
rpm	revolutions per minute

rps	revolutions per second
TPT	two-pass thresholding
TRASE	Table Tennis Recording Analysis Software Environment

Acknowledgements

I wish to express my deepest thanks to everyone that helped and supported me throughout my Ph.D and when compiling this thesis. I want to thank my supervisors, Prof. Eran Edirisinghe, Prof. Paul Chung and Dr. Ben Halkon, for giving me the opportunity to pursue postgraduate studies and for their continued support and advice. I also wish to express my gratitude to everyone at the Loughborough University's Sports Technology Institute who assisted with the laboratory installations and unwittingly became player subjects; your input has been invaluable.

To the memory of my parents.

Contents

Abstract	i
Keywords.....	iii
List of abbreviations	iv
Acknowledgements.....	vi
Contents	vii
List of tables	xii
List of figures	xiii
1. Introduction.....	1
1.1. Background and motivation.....	1
1.2. Context.....	2
1.2.1. Sport and video analysis.....	2
1.2.2. Table tennis and CV.....	3
1.2.3. Deficiencies of current solutions.....	4
1.3. Research hypothesis	5
1.4. Research aim and objectives.....	5
1.5. Research challenges	5
1.6. Contribution highlights	7
1.7. Thesis structure.....	8
2. Literature review	10
2.1. Introduction	10
2.2. Literature search rationale.....	11
2.3. CV and table tennis	11
2.3.1. Table tennis ball tracking and detection.....	12
2.3.2. Umpiring.....	13
2.3.3. Coaching.....	14

2.4.	CV and other sports	14
2.4.1.	Cricket	14
2.4.2.	Basketball	15
2.4.3.	Baseball	16
2.4.4.	Tennis	17
2.4.5.	Football (Soccer)	18
2.5.	Commercial CV systems	19
2.6.	Non-CV table tennis technologies.....	22
2.7.	CV algorithms	23
2.7.1.	Low level feature extraction (edge detection).....	23
2.7.2.	Foreground extraction	24
2.7.3.	Object detector	24
2.7.4.	Tracking	25
2.8.	Literature review findings	25
2.9.	Summary	27
3.	Design methodology and schedule	28
3.1.	Introduction	28
3.2.	Preliminary video sequence capture design considerations	29
3.2.1.	Philosophies	29
3.2.2.	Technical constraints.....	30
3.2.3.	Ball imaging considerations	31
3.2.4.	Camera setting considerations.....	34
3.3.	Apparatus	37
3.4.	Lab design and layout	40
3.5.	Detecting changes in ball motion/trajectory	43
3.6.	Schedule	46
3.7.	Summary	52

4.	Computer vision processing	54
4.1.	Introduction	54
4.2.	Software design.....	54
4.3.	Workflow.....	59
4.4.	TRASE Findings	65
4.5.	Qualitative comparison.....	92
4.6.	Qualitative summary.....	101
4.7.	Conclusion	103
5.	CV algorithm quantitative comparison: A case study	105
5.1.	Introduction	105
5.2.	A novel CV comparison methodology: The Efficacy Metric Set (EMS)	107
5.3.	TRASE algorithm processing.....	113
5.4.	EMS analysis	116
5.5.	REI Result.....	121
5.6.	Conclusion	123
6.	Determination of ball position and dynamics	124
6.1.	Introduction	124
6.2.	Estimating ball location.....	125
6.3.	Comparing CV and Vicon.....	126
6.3.1.	Scale of reference	126
6.3.2.	Data transformation and smoothing	129
6.3.3.	Measuring the difference.....	132
6.4.	Chalk dust experiment	134
6.5.	Position results.....	135
6.6.	Calculating ball dynamics	140
6.6.1.	Velocity.....	140
6.6.2.	Acceleration	141

6.6.3. Bearing	142
6.7. Dynamics results	143
6.8. Conclusion	145
7. Event classification	147
7.1. Introduction	147
7.2. The proposed classifier	147
7.2.1. Event states	150
7.2.2. Event characteristics	151
7.2.3. Event motion thresholds	159
7.3. Classification results.....	160
7.4. Classification for officiating	161
7.4.1. Service height (apogee).....	162
7.4.2. Table yield	165
7.4.3. Centre line	165
7.5. Application to umpiring and coaching	168
7.6. Conclusion	169
8. Conclusion and further research	171
8.1. Introduction	171
8.2. Discussion.....	172
8.3. Contribution highlights	174
8.4. Further research and improvement.....	175
8.4.1. The EMS as a measure of predictability.....	175
8.4.2. Automatic algorithm switching	175
8.4.3. Let detection	176
8.4.4. Human vs. computer officiating.....	177
8.4.5. Other sports	177
8.4.6. Machine learning based classification	177

9. Bibliography.....	179
10. Appendices.....	190
A. Event state table	190
B. Defining a FAIL.....	197
C. Core camera settings	200
D. Scholarly contributions	203

List of tables

Table 3-1 Munsell colour standards for approved table tennis balls	32
Table 3-2 Camera sensor dimensions	39
Table 3-3 Summary of schedule.....	46
Table 4-1 Shape features	61
Table 4-2 Kalman Filter missing frame data.....	75
Table 4-3 Comparison of ball diameter calculations.....	82
Table 4-4 Example reflection data for 600-P-10 (OF-LK-100-600-10).....	87
Table 4-5 Observed ball and reflection pairs	89
Table 4-6 Reflection sequence in 700-P-12	89
Table 4-7 Feature priorities for ball Isolation	91
Table 4-8 Refined feature priorities for ball Isolation.....	92
Table 4-9 Horn-Schunck vs Lucas-Kanade in key frames from 700-P-10.....	100
Table 4-10 Initial qualitiative assessment	102
Table 5-1 Summary of CV algorithm efficacy metrics for sports video sequences.....	109
Table 5-2 Computer vision algorithms selected for comparison.....	115
Table 5-3 Relative Efficacy Index Results	121
Table 5-4 REI for Δ UCDF	122
Table 6-1 % Δ Ball image diameter from table centre	129
Table 6-2 Ball characteristics for 2D position validation.....	129
Table 6-3 Butterworth filter coefficients	131
Table 6-4 Data transformation sample output	132
Table 6-5 Video data for position validation.....	135
Table 6-6 CV position difference relative to Vicon	136
Table 6-7 Summary results for the Δ displacement between CV and Vicon	138
Table 6-8 Summary of bounce distances (600-P-12)	139
Table 7-1 Event threshold values.....	159
Table 7-2 Summary of classification results.....	160
Table 10-1 Core camera settings.....	200
Table 10-2 Video sequence to camera configuration ID cross reference	201

List of figures

Figure 3-1 Siemens Star Test Chart	35
Figure 3-2 Lab with Loughborough University's Sports Technology Institute	38
Figure 3-3 Candidate camera positioning	40
Figure 3-4 Optimum camera position	41
Figure 3-5 Optimum FoV (a) and umpire position (b) (Table Tennis England 2010)	42
Figure 3-6 Cameras Co-located	42
Figure 3-7 Vicon configuration.....	44
Figure 3-8 Modified table tennis balls as Vicon markers showing two types of material	45
Figure 3-9 Location of markers around the table	48
Figure 3-10 Using Foresight™ Launch Monitor to measure speed	49
Figure 3-11 Ball annotated with spin markers	49
Figure 3-12 Calibration rig in situ	51
Figure 3-13 Calibration rig measurement zone	51
Figure 4-1 TRASE User Interface	57
Figure 4-2 Comparison Processing modules	58
Figure 4-3 Event Analysis	59
Figure 4-4 The effect of binary thresholding (100-P-03 frame 50 T=0.9)	66
Figure 4-5 Shape variation when thresholding the ball in consecutive frames.....	67
Figure 4-6 Binary 100-C-14 frame 20 T=0.5	67
Figure 4-7 Binary 9999-U-01 frame 50 T=0.7	68
Figure 4-8 Determining the binary threshold	70
Figure 4-9 The effect of contrast on binary threshold	71
Figure 4-10 Collision (let) induced segmented regions.....	72
Figure 4-11 100-P-03 Processed using Kalman Filter	73
Figure 4-12 Ball location using Kalman Filter	74
Figure 4-13 Representation of the ball shape when detected at low shutter speeds.....	76
Figure 4-14 Variations in ball area across the frame	77
Figure 4-15 Variations in ball diameter across the frame.....	78
Figure 4-16 Variations in ball characteristic indices.....	79
Figure 4-17 Area pre and post let event	80

Figure 4-18 Diameter measurements pre and post let event	81
Figure 4-19 Algorithm y-values for 100-P-6.....	83
Figure 4-20 Algorithm x-values for 100-P-6.....	83
Figure 4-21 100-P-06 x-values between frames 30 and 40	84
Figure 4-22 Maximum difference in ball location between algorithms (100-P-06).....	85
Figure 4-23 Issues of reflection as observed in TRASE.....	86
Figure 4-24 Pre-filtering of TRASE output including reflections	87
Figure 4-25 Comparison of characteristic filtering	91
Figure 4-26 Edge detector results	93
Figure 4-27 GMM applied to 100-P-06	97
Figure 4-28 Applying Optical Flow (H-S) to detect the ball.....	98
Figure 4-29 Optical Flow applied to 9999-U-01	99
Figure 5-1 Overview of the TRASE processing stages	114
Figure 5-2 Initial frame detection ($\mu\Delta\text{IFD}$)	117
Figure 5-3 % Failure Initial frame detection (% ΔIFD)	117
Figure 5-4 Final frame detection ($\mu\Delta\text{FFD}$)	117
Figure 5-5 $\Delta\text{FFD FAIL}(\%)$	117
Figure 5-6 μ Multiple false candidate detection (MFD)	117
Figure 5-7 MFD FAIL (%)	117
Figure 5-8 μ Total candidate detection (ΔTCD).....	118
Figure 5-9 $\Delta\text{TCD FAIL}(\%)$	118
Figure 5-10 False frame detection (% FFD)	118
Figure 5-11 FFD FAIL (%)	118
Figure 5-12 Non-detection (% ND)	118
Figure 5-13 ND FAIL (%)	118
Figure 5-14 μ Undetected collision and detection frames (% UCDF)	119
Figure 5-15 FAIL UCDF (%).....	119
Figure 5-16 μ Undetected ball trap frames (ΔUBTF).....	119
Figure 5-17 FAIL UBTF (%)	119
Figure 5-18 μ Undetected bounce frames (ΔUBF)	119
Figure 5-19 FAIL $\Delta\text{UBF}(\%)$	119
Figure 5-20 μ Undetected during hit (ΔUDH).....	120
Figure 5-21 FAIL $\Delta\text{UDH}(\%)$	120

Figure 5-22 μ MBE (%).....	120
Figure 5-23 Relative $\mu\Delta$ ball diameter (Δ MBD)	120
Figure 5-24 Relative μ processing speed (PS)	120
Figure 6-1 Fixed rig locations	127
Figure 6-2 Rig based ball metric analysis	128
Figure 6-3 Euclidean compared to Manhattan distances	133
Figure 6-4 Chalk dust experiment	134
Figure 6-5 Chalk dust contrast in video sequence 600-P-12.....	135
Figure 6-6 Δ displacement between CV and Vicon (Euclidean)	138
Figure 6-7 Comparison of (a) non-smoothed and (b) Butterworth smoothed data.....	139
Figure 6-8 Velocity calculation	141
Figure 6-9 Quadrant reference for bearing analysis	142
Figure 6-10 Comparison of CV and Vicon Velocity for 700-P-17	143
Figure 6-11 Comparison of CV and Vicon acceleration for 700-P-17.....	144
Figure 6-12 Comparison of CV and Vicon bearing data for 700-P-17	144
Figure 6-13 Statistical comparison of CV and Vicon motion characteristics (n=48)	145
Figure 7-1 Match-playing zones	149
Figure 7-2 Analysis of bounce dynamics in TRASE	152
Figure 7-3 Bounce event characteristics	152
Figure 7-4 Return sequence	153
Figure 7-5 Return event characteristics	154
Figure 7-6 Ball boundary merging with the foreground and background	154
Figure 7-7 Net (let) frame sequence	155
Figure 7-8 Let event characteristics	156
Figure 7-9 Net collision from 500-P-02 showing typical occlusion	157
Figure 7-10 Net collision characteristics	158
Figure 7-11 100-P-06 net collision bearing and trajectory	159
Figure 7-12 Apogee calculation.....	163
Figure 7-13 Centre line location	166
Figure 7-14 Occlusion of the doubles line.....	167
Figure 7-15 Doubles line correction factor (ϵ)	168
Figure 8-1 Net motion detection	176
Figure 10-1 Original unprocessed frame containing a single ball	198

Figure 10-2 700-P-10 Failed detection using Zerocross and CHT	198
Figure 10-3 Failed detection using Zerocross and CHT	199

1. Introduction

1.1. Background and motivation

Over recent years there has been an increasing demand for the use of analytics in sport (Fry & Ohlmann, 2012). From players and officials to broadcasters and sports fans, the need for automated information relating to performance, accuracy and match-play tactics is becoming a key asset to success. Image analysis techniques for segmentation and detection are progressing from long established theoretical algorithms to practical, commercial installations of real-time processing systems. The outputs of such installations are often represented in the form of match statistics, or as overlay graphics tracing the dynamics of objects of interest (OOI) for specific match events. A number of globally-popular, professional sports leagues are using sophisticated, hardware intensive ball detection systems, such as “Hawk-Eye” (Hawk-Eye Innovations Ltd, 2013), within stadia to provide umpire decision support systems; this technology is also extending into non-ball sports, including athletics (Hart, 2014). In parallel, there is an increase in miniaturisation, portability and availability of image recording technologies for low-cost consumer oriented cameras, mobile phones and other portable electronic devices.

Bringing these two trends together would integrate consumer level cameras with the ability to perform automated Computer Vision (CV) based match statistics, umpire support and performance analysis for use by clubs, coaches, school teams or individuals across a wider range of less commercial sports. However, incorporating low-cost hardware in the design and build of such CV software requires efficiency in every element of the build, from awareness of hardware constraints to the optimal selection of CV and image manipulation techniques.

Considering that there is substantial literature available for the detection, segmentation and analysis of OOI dynamics using CV in sport, there is little empirical evidence for methods of comparison, evaluation and validation of such algorithms when applied to ball sports. Fewer still consider the computationally challenging application of table tennis, a global sport popularised by its accessibility and simplicity of concept. As such, the motivation of this thesis is to provide a unique and rigorous scientific foundation and approach for CV algorithm evaluation, comparison and implementation in sport, through the specifics of experimentally based data analysis when applied to monocular ball detection video sequences in table tennis. The result is a novel versatile comparator index, with 2-

dimensional ball dynamics validation and robust event classifier. An event is defined here as a significant change in table tennis ball motion which affects game statistics, performance analysis data, or officiating decisions.

1.2. Context

1.2.1. Sport and video analysis

Sport presents a challenging and competitive environment within which the possibilities of new technologies are continually evolving. The application of CV based video analysis technologies is one particular area, which has gained more recognition in recent years for its automated processing and immediate feedback of sports analytics. It is being implemented across an increasingly wide range of sports by top athletes, their coaches and broadcasting companies as (a) the need to find incremental improvements becomes increasingly critical, (b) referee and umpiring decisions become identified as the cause for an individual or team failing (Leveaux, 2010) and (c) sports fans' desire for insight into technique and tactics increases. Central to many sport's use of CV techniques is determining the location of the ball (or other OOI) at any given moment, by analysing image sequences recorded using multiple cameras to model the 3D trajectory path and then accurately determining its position in relation to a boundary, player or other part of the environment. This has become part of both the umpire decision support system and the spectator's excitement.

The accuracy of all of these systems is critical to their successful implementation, with acceptance into sport under debate. Some of the most popular solutions undergo years of testing and validation before being introduced into the game – and even then are under continual scrutiny, evaluation and development (Hawkins, 2013). Systems such as "Hawk-Eye" (Hawk-Eye Innovations Ltd, 2013) are applied across several sports with individually calibrated and validated installations. The relatively high cost and complexity of commercially available CV systems is largely due to this need for a high degree of accuracy. At the professional level of sport, incorrect officiating can cause substantial losses to teams and their businesses. However, no solution can ever be 100% accurate in every event. Tolerances occur through environmental variations, calibration processes, software design and human interpretation; acceptable tolerances are defined, understood and accepted prior to implementation, such as in FIFA's accuracy requirement of $\pm 1.5\text{cm}$ for goal-line technology (GLT) (FIFA, 2014).

For many sports, automated CV systems have not been implemented due to either a) the lack of investment, b) the physically prohibitive and intrusive challenges of installation, or c) approval by the sport's governing bodies. Only highly commercialised professional global sports leagues have the resources to invest in developments for OOI detection and analysis technology. However, reducing the complexity of necessary detection hardware and simplifying the requirements of the CV software would bring CV based OOI detection solutions to a wider use. Simplification can come from monocular (2D) based analysis of consumer-level video records made by amateur players, coaches or clubs to enable mass performance analysis, umpire decision support and automatic scoring.

1.2.2. Table tennis and CV

Real-time video analysis and OOI detection is common in globally recognised professional sports. Yet table tennis, one of the most popular sports in the world, generally lacks the exploration and investment in CV based ball detection technologies. Outside of human based match-play, CV has been regularly incorporated in the technical challenges of enabling robots to play table tennis, as suggested by Zhang et al. (2010). Although the fundamentals of the processing technology may be re-used, the requirements and constraints between robot design and non-intrusive OOI detection for performance analysis and umpire decision support is very different.

In the most recent Olympics (London 2012) high speed cameras were used to slow down match-play for the benefit of the spectator, but there had been no attempt to enhance this technology to provide data to the players, coaches or umpires. The benefits this could bring to the sport are considerable. Take, for example, a young player practising to improve speed, direction, point of impact with the table or service legality. An additional application would be for a coach having to observe and develop multiple players at once, perhaps even remotely using online applications. Even during a club tournament, having access to umpire decision support systems offering impartial data and analysis would be of benefit. CV based outputs from low-cost monocular video sequences, carefully positioned alongside the umpire, would enable data to be immediately available and of sufficient accuracy to be of benefit. It is in this domain where the focus of this research resides.

From this research perspective, table tennis is ideally suited to stress test algorithms and develop CV comparison methodologies. The sophisticated match dynamics of a 40 mm diameter ball travelling at speeds up to 31.25 ms^{-1} (Rusdorf & Brunnett, 2005) with spin-induced curved trajectories, high

frequency collisions and occlusions and played across a table only 2.74 m in length at a raised plane 76 cm from the ground makes this a challenging prospect for a CV project.

1.2.3. Deficiencies of current solutions

There are several high-end commercial solutions for the detection and analysis of ball movements in several sports but none of them have been designed to specifically look at the challenges of table tennis. Table tennis has a number of factors which, when combined, make ball detection using current systems challenging. Not only is the playing area a fraction of that of other sports, but the ball is significantly smaller and the interval times between shots and events are measured in fractions of a second. Additionally, the physical environment of table tennis generally prohibits fixed, permanent installations and the lack of global investment into the sport has negated the possibility of established solutions being installed within specifically designed arenas. These factors lend themselves to a solution with non-invasive CV technologies using low-cost monocular recording hardware.

At the time of writing, handheld computing devices, originally designed for mobile communications, are ever expanding into personal sport, health and fitness monitoring and feedback devices. As innovations in semiconductor, photographic and electronic sensor technology continue, the potential to combine these devices with useful CV applications is viable. However, with sports and activity applications the trend for compact computing technologies is in either (a) healthcare, with digital biomarkers being recorded integrated directly to track subject activity (Earl, 2015), or (b) intrusively modifying sport equipment with sensors to gather motion data, such as the Adidas miCoach (Adidas, 2014). Yet not all sports equipment can be modified with sensor based technology; modifying a table tennis ball, where the ball weighs only a few grams, adds extra weight and alters the dynamics of flight substantially.

Occasionally, academic research has considered theoretical aspects of CV in table tennis, or focussed on a very specific solution to a single event in the game, often by the useful work of Wong (2009). No research to date has investigated requirements, issues and solutions through experimentally based, repeatable analysis, to offer a data-driven comparison and justification of low-cost CV algorithm selection and implementation in table tennis.

1.3. Research hypothesis

The hypothesis for this research states that low-cost monocular table tennis video recordings can be used for automatic event classification by detecting changes in ball motion using CV processing, without the need for 3D modelling and representation. Player performance can be evaluated by measuring the ball motion using image characteristics of the ball itself. This approach provides a viable and accessible solution for the mass application of match officiator support, automatic scoring systems and player performance assessment, in table tennis and other sports.

1.4. Research aim and objectives

The aim of the proposed research reported in the following thesis is to test the hypothesis by means of practical, robust experimentation and results analysis. A thorough review of recording configurations will be made, together with a process to test and rank CV algorithms for use in table tennis using 2D image data. This comparison process will measure the ability of a selection of algorithms to automatically detect a table tennis ball during match-play. Once a viable CV algorithm has been identified, further assessments of automated scoring, player performance assessment and umpiring decision support can be made. Central to the research is the investigation of methods to validate the detection of change in ball motion; sole reliance on outputs from CV data is not sufficient evidence of proof of detection and therefore the investigation and implementation of non-CV detection is required to run in parallel during the experiments. This work aims to provide a foundation for CV algorithm evaluation and selection through rigorous and robust computer science techniques when applied to table tennis. A comparison with other sports is to be discussed throughout, assessing if there are cross-over properties between differing sports which would make the application of this technology suitable for a broader base.

1.5. Research challenges

Simply detecting a ball during a table tennis match is a challenge. As discussed, the use of fixed, high-end camera installations for table tennis is restricted by the portability of the tables and the non-dedicated environments within which the game is played. The tables do not require a permanent court, ground or field and so multi-purpose sports halls, open venues and arenas designed for a variety of activities are used. There are no rigid fixings in relation to the table to which a camera can be attached. The image background within the match scene is unpredictable, and the lighting is sub-

optimal. The fast ball speed along with the spin-induced curved trajectories, being played across relatively short distances and along a raised plane is unusual in sport. This makes the application of CV for ball detection, without heavy investment in arenas and recording technologies, both a complex and an extreme test of current CV algorithms.

Use of a single camera means the image data is lacking spatial content. Table tennis is not played on a flat 2D surface. Complex, 3D ball trajectories caused by the light ball and violent spin generated by a player off their racket is an essence of the game. A truly accurate CV based detection of the table tennis ball's position in 3D space traditionally requires preconfigured, calibrated, multi-camera installations. Yet, according to the hypothesis of this research, knowing the precise location of the ball in 3D space is not necessarily the correct question. Instead, the challenge is to interrogate the *relative* motion of the ball from a video sequence, detecting deviations from a predicted path and analysing the data through an event classifier. Given sufficient reliability in determining the location of the ball as it moves across the camera's sensor, subtle movements of the ball in a 2D projection provide inherent information from which an event assertion can be made. The challenges in doing so are a) detecting the initial frame containing the ball, b) obtaining sufficient data points, c) designing experiments to validate fit-for-purpose precision of the location data points, d) positioning of the camera to avoid critical occlusions and finally e) determining a suitable unit of measure. With this approach the research aims to remove the reliance on multi-camera installation. Using only a single recording device, the need to be able to measure distances in standard units of a ball's trajectory then becomes an additional challenge. This research will investigate the use of the known and fixed sizes of the ball and table as the references for scale; the benefits and challenges of such an approach will be documented throughout.

This is evidence based research. It is to take theoretical applications of CV based software and demonstrate their suitability using video recordings made specifically for this research. The design of the experiments must consider variations in player performance, equipment and environment and are to be detailed for repeatability. Any conclusion of the selection of one (or more) CV algorithms must be supported with an evidence based review; this will be in the format of a direct comparison of CV algorithms against a statistically large data set.

1.6. Contribution highlights

As brief highlights, the research presented claims to contribute to the area of CV in sport with the presentation, analysis and discovery and of the following contributions and findings:

1. A comprehensive root analysis of CV processing capabilities and evidence based optimal CV application, using empirical data from monocular recordings of table tennis match play events. The result is a CV algorithm review and selection framework extendable to other ball sports (iv).
2. A database of reusable, designed, event based monocular table tennis video sequences. A software environment within which to directly and automatically compare multiple CV algorithms against this video library (vi & vii).
3. Discovery of variations in the measurements of ball motion detection between different CV algorithms (ii).
4. A rigorous comparison of CV algorithms against a novel event based measurement, the Efficacy Metric Set (EMS). The creation of a unique, unbiased, event based comparison index, the Relative Efficacy Index (REI). Application of the index additionally includes the justification for automatic switching of CV algorithms for specific events (ii).
5. An approach for the non-intrusive validation of CV-based detection of changes in ball motion using measurements of velocity, acceleration and bearing (i).
6. Feasibility studies for using the table tennis ball as the continually optimised scale of reference to measure real world distances from 2D images (v).
7. A novel method of unsupervised table tennis (or similar ball sports) event classification based on 2D ball motion data and the sequence of expected events (iii).

A list of existing and planned publications as a direct result of this research can be found in Appendix D. References to these publications are highlighted in parenthesis above.

1.7. Thesis structure

The thesis presented here initially discusses the background to the sport of table tennis and how CV is currently a new topic within the game. A literature review in Chapter 2 broadens the discussion to look at how other sports use CV based techniques (both commercial solutions and research based concepts) to detect a ball and, where appropriate, describes the benefits this can bring. As appropriate, a reference to a potential use of a technology with table tennis is described. The literature review then covers, in detail, current research of table tennis and CV: where it has progressed, what has been discovered and which techniques have been suggested so far. Finally, the literature review is summarised with an outline of current gaps in the knowledge.

This research is primarily experimental and as such Chapter 3 details the method of designing and performing experiments with table tennis and a variety of recording hardware to collect data used for subsequent analysis. A description of expected problems in recording and analysing outputs is discussed, together with their mitigation through experiment design. The ideal camera location and configuration for monocular sequences completes this chapter.

Chapter 4 describes in detail the system (TRASE) for comparing different CV techniques using the source video sequences from Chapter 3; it covers the decisions made for generating suitable data for use in CV algorithm comparison, ball dynamics analysis and event classification described in subsequent chapters. A comparison of selected CV based ball detection algorithms is made in Chapter 5 using a novel event based index. This identifies the algorithm offering the greater success in ball detection, to be used in subsequent analysis. This is not intended to compare all possible CV algorithms (this is outside the scope of this research) but does compare several high profile candidates identified in the literature review. This also offers a baseline for evaluating any CV algorithm for use in sport with similar characteristics to table tennis. A qualitative as well as a quantitative appraisal of the algorithms is made. The outputs reveal a variance in the location of the ball, depending on which algorithm is used. A proposal for an optimum algorithm is assessed in subsequent chapters. Chapter 6 provides a method to validate ball location in a table tennis match-play environment. The chapter concludes by presenting a case study of applying motion data to estimate the trajectory of the ball between frames (its speed, acceleration and bearing). This output is then used in Chapter 7 to classify the ball's motion into a number of discrete events used for player performance improvement, automatic scoring and match officiating. Chapter 8 concludes with a summary of results and findings, with a detailed report of future work and further research. Finally, there are appendices featur-

ing a detailed definition of key table tennis events, the definition of a failed detection used in the CV algorithm comparison and a summary of the scholarly contributions.

2. Literature review

2.1. Introduction

The practical applications of CV in table tennis have not been widely researched. The fundamental appraisal of low-cost CV solutions for ball detection intended for umpiring, coaching and performance assessment within the sport is not available. There are a number of articles covering concepts, theories and ideas of how a ball may be segmented from a table tennis video sequence, but no rigorous, experimental analysis has been made. The literature review presented here details these areas. This chapter aims to provide the most current assessment of the potential for ball detection, identification, tracking, event classification and coaching performance analysis within table tennis.

The review identifies existing CV algorithms and techniques for table tennis and, where relevant, other sports which would benefit from this research. CV algorithms which have not been evaluated in this domain are also included for comparison. To this extent this review has a focus on five key areas:

1. Ball detection and identification, including multi-ball detection where implemented
2. Tracking solutions and their application
3. Event classification and match statistics
4. Post-event analysis for coaching and skills improvement
5. Umpiring for referee decision support and rules adherence

The review begins with a broad outline of a range of commercial solutions for object detection currently applied across sport generally. For initiating CV based strategies within the table tennis domain, understanding the wider available technical capabilities, hardware investments and successful technologies is pertinent. Following this, a review of current research in CV and table tennis is detailed in Section 2.3, including a discussion of the potential benefits when applied to the game. Section 2.4 performs a post-review of CV in sports other than table tennis, to assess if there are learn-

ings from other domains which can be applied. The findings of the literature review are presented in section 2.6. Throughout the literature review, necessary designs in hardware and software are highlighted, with discussion on the suitability of applying the technology to table tennis.

2.2. Literature search rationale

This research domain intersects several disciplines across computer science, information science and sports technology. The application of CV to sport is well recognised and developing as a mature technology. However, evidence of its use within table tennis is not prominent. Furthermore, most of the available solutions tend towards its use within robotics and do not offer a detailed analysis of algorithm selection, review, complexities and areas for improvement. Finally, sports technology is making step changes to the performance of athletes and sports people. The sports community is becoming used to measuring an athlete's work rate, examining player motion to improve techniques, or using game theory to improve strategic abilities. Yet for table tennis the application of CV to for performance analysis is not apparent.

As such the rationale for this literature review is to examine current research in the detection of a ball from video sequences and also investigates any work in using this data for event analysis and player performance assessment. This involves a breadth and depth assessment of both academic research and commercial solutions, when using CV and other detection technologies, not only for table tennis, but also across many different sports. This is essential for the root analysis of how CV has been applied and its further potential in the sport. Where CV algorithms are presented, these are critically appraised for the purpose of table tennis. A direction towards a fully evaluated, optimised and refined CV solution applied within a practical, real world table tennis environment is then progressed.

2.3. CV and table tennis

There is no ready-made solution for detecting a ball in match-play table tennis. However, there is some preliminary research for the partial application of CV to ball detection, either for the ultimate goal of officiating, or for the technical challenge and necessary CV research. A summary of available documented research in this area is presented below.

2.3.1. Table tennis ball tracking and detection

Mezaris et al. (2004) proposed a method of image segmentation which could be incorporated into video encoding technology, such as MPEG-4. The algorithm applies image segmentation to the initial frame and then tracking through the remaining frames, with novel algorithms for each stage. The claim is that this is suitable for fast moving and newly appearing and disappearing regions. Table tennis is used as one example, although full description of the table tennis footage and environment are not explained. This would require further investigation for the suitability of this particular application.

In Wang's article (1998) describing the Spatial Segmentation Algorithm, the author claims to successfully segment a table tennis image taken from a video sequence into homogenous regions based on intensity. Although admitting slow processing rates would make a commercial, or near real-time application unforgiving, using more modern technologies could allow the algorithm to be developed further. What can be seen in the images provided is that the image background is clear of complex, confusing and distracting surfaces and colours. It does not detail the source or quality of the video nor does it highlight strengths and weaknesses during specific match-play events. This makes it difficult to assess the success, or make a retrospective comparison, of the algorithm.

Early work by Wong (2008), focussing on examining the table tennis service only, confirms that further investigation needs to be made in ball tracking. In this article there is the suggestion that Artificial Neural Networks (ANNs) have the potential to identify the ball and it has investigated techniques to reduce processing time for each consecutive frame, by using prediction algorithms estimating the neighbourhood within which the ball may exist. Further development of this approach suggests using frame based object segmentation together with spatial and temporal clues (K. C. P. Wong & Dooley, 2010). Both two-pass thresholding (TPT) and clustering have been considered as the initial image processing step, with image thresholding being the agreed algorithm to implement, based on computational efficiency and having a single ball. The article suggests that the ball must be visible during the serve and so camera positioning is critical for success. Partial ball obscurity is managed through prediction techniques, assessing trajectory and motion to confirm the candidate object is the ball. The experiments were conducted on three video sequences of varying quality, finding that to achieve 100% detection rate a minimum pixel resolution of 720 x 576 at a frame rate of 100 Hz was required (compared to 72% for 352x240 pixels at 30 Hz). All of these experiments were con-

ducted using MatLab (The MathWorks Inc., n.d.-b) on an Intel Core Duo 2.2 GHz (Intel Corporation, 2009).

It is encouraging that the ball can be successfully segmented with, under the conditions provided, a high detection rate. However, Wong acknowledges that the work here is not able to perform real time analysis due to the user being required to input initial ball size, colour and location for each video sequence. There are also no details of the camera angle, or an evaluation and comparison of tracking and prediction algorithms across all events in a typical table tennis match. The study was focussed on the serve, leaving many more aspects of event detection, officiating and performance review to be investigated. These contributions by Wong have been important to this research domain, having pushed the understanding of CV in table tennis and demonstrated its fundamental problems, complexities and future potential.

2.3.2. Umpiring

The initial work by Wong (2007) is of importance when reviewing the potential for applying CV to table tennis for umpiring. His research has focussed around detecting and successfully tracking the ball particularly during a serve, with the aim of ensuring it complies with the International Table Tennis Federation (ITTF) (The International Table Tennis Federation, 2014) rules. The ball detection methods employed here are focussed around segmentation. Wong does not propose the use of motion and colour based segmentation because, for example, of the multiple overlapping moving objects in the scene and the inability to accurately identify sufficient numbers of colour clusters. Instead, the approach suggested is one of threshold based segmentation. According to Wong, this has the additional benefit of being computationally low resource and easy to implement. Here, the image is made binary with a reference made as to how to evaluate the threshold value; it is not a proposal to do this automatically. Key to this and future officiating implementations, is that the process of segmentation is aimed at *approximately locating* the ball. Because extreme image thresholding can distort the shape, Wong suggests the original image is used to restore the true ball shape.

In later research, Wong (2009) states that the tracking of a table tennis ball, serve analysis and shot recommendation is a new area of study. It goes on to further state that a number of papers relating to table tennis ball tracking exist, but these do not focus on shot umpiring. It is worth noting that the literature review in this chapter supports this view; additionally no research has yet proposed a solution for table tennis shot selection and coaching. In this article, Wong describes a process of officiat-

ing in table tennis service as a novel idea without previous attempts at a solution. Wong continues with a multi stage process to detect the ball from the ITTF photo gallery (ITTF, n.d.), through: 1) threshold the image 2) object evaluation; it concludes that a TPT is an improvement over single thresholding and ANNs are superior to feature filters based on shape descriptors. However, Wong acknowledges that this is evidence based on images and recordings made without influence over recording apparatus design and configuration.

2.3.3. Coaching

There is currently no available literature for automated coaching using CV techniques in table tennis. Being able to detect the location of the ball during match-play or practice has not been extended into providing coaching feedback and support to enable automated and detailed player performance feedback. It is a key area for this research.

2.4. CV and other sports

Confining this review to CV implementations within table tennis alone is not conclusive. A great deal of research has been made with respect to the application of CV to ball detection, tracking and shot analysis across a number of sports. Many of these algorithms are context specific, requiring exacting environmental situations, video quality or sporting event detail. The following articles do not relate directly to table tennis, but should be read for potential implementation within the table tennis domain. Conversely, outputs from the discussion of CV in table tennis could increase potential to other sports.

A link between the applicability of algorithms across sports is investigated; algorithms managing common characteristics, such as number of players, ball speed, size, colour, area and size of play etc., are of relevance. Comparing combinations of algorithms across sports also provides an insight into the sophistication of the problem domain, and offers ideas of how other researchers have solved similar complexities.

2.4.1. Cricket

With its tendency to measure gameplay in great detail, cricket has become one of the leading sport advocates of allowing commercial technology to assist with both increased spectator information

and umpiring decision. Examples range from the ‘Snick-ometer’ (Wood, 2008), to the introduction of Hawk-Eye for assisting with leg-before-wicket decision making and providing the broadcast audience with an immediate analysis of ball trajectories during any given innings. In cricket, Hawk-Eye is primarily for leg-before-wicket detection but it also gathers ball metrics (speed, spin), reaction times and shot trajectory analysis (the waggon wheel) for any given innings (Hawk-Eye Innovations, 2013). The system requires a short delay for processing before being transmitted¹.

For non-commercial implementations, one approach (Velammal & Kumar, 2010) has been to remove noise through a median filter, convert to grey-tone and then to segment each frame using the seeded Region Growing algorithm. Each frame then goes through a process of ‘sieving’ by ball characteristics until the ball is found. The principle being that it is easier to find a ball among a set of candidates than it is within the entire frame. However, no metrics have been provided to demonstrate the success of the approach. No analysis of ball speed and size is provided, no ball occlusion solution and no direct comparison has been made against other segmentation techniques given the same video data inputs.

2.4.2. Basketball

There has been a solution proposed for basketball (Ruiz, 2010) combining the colour histogram of the ball with Circular Hough Transform (CHT) (Duda & Hart, 1972) to extract and detect the ball, then recursively track the ball in 3D using a series of calibrated cameras. The tracking required searching the 3D space around the previously known position by assessing each individual 2D representation of the space created by the individual cameras. This technique was based on video having a resolution of 720x576 pixels at a frame capture rate of 25 Hz. When compared to table tennis the solution, however, did not have the extra complexities of a much faster ball with frequent trajectory deviations, played across a smaller area and having a much smaller diameter. Also the low-cost implementation of a *single* camera for this research limits the applicability of this solution. However, the CV algorithms employed provide an indication for image processing evaluation.

An alternative 2D approach has been suggested by Chakraborty and Mehr (2011), where each video frame is segmented using frame differencing to identify a set of candidate objects. This is then filtered based on size within a range relating to the distance from the camera, shape (allowing for de-

¹ An interesting discussion in recent times was the step away from the Hawk-Eye solution by the International Cricket Council (Holt, 2011) on the grounds of unproven accuracy.

formation of the ball, either physical or apparent due to video artefacts) and compactness, defined as the area of the smallest circle drawn around the candidate ball. Trajectories for all candidates are mapped and the true trajectory of the candidate ball can be identified through its physical motion. During times of occlusion, trajectory interpolation is implemented. Finally the trajectory is superimposed on the original frames. Of importance, the article highlights the problem of calculating the difference of two frames when a fast ball is being identified; the proposed solution is to apply a three-frame differencing technique. An additional edge detection method is used where there are significant discontinuities in intensity values. Test data was created using a standard Canon™ camcorder video having a resolution of resolution of 640x480 pixels with a frame capture rate of 30 Hz. Results suggest an accuracy of 96.22% in detecting the ball and 100% accuracy in tracking. If applied to our table tennis scenario, we may expect good results if the ball size was significant relative to the frame size and the limitations of post processing to track the ball was within our expectations.

2.4.3. Baseball

Using low-cost cameras with strobe techniques for frame capturing is central to work performed by Theobalt et al. (2004) in analysing baseball trajectories. Baseball was chosen for its high speed and with the aim of measuring initial velocity, rotation axis and spin frequency. In these experiments optical markers were placed on the ball, of different colours. The environment is placed in a dark room, with the Olympus™ Camedia C5050 digital camera (Olympus America Inc., n.d.) set to a long exposure and the 20 μ s-long strobe flashes at a frequency of 80 Hz effectively acting as the shutter. OpenCV (Itseez, 2015) is used for colour based background segmentation to create a binary mask. The centre of the ball is found by fitting ellipses to the boundary points and further improved by applying a CHT around the ellipse centre. A spin rate of 1600 revolutions per minute (rpm) is detected with the ball travelling at over 80mph. An admission is made that no ground truth data for the flight of the ball is available and therefore validation is made through physical modelling. Final location accuracy using four cameras for 3D imaging is in the region of 18-15% of the diameter of the baseball.

Although operating in a darkened room is too challenging for the majority of table tennis players, the recording specifications, camera quality and configurations are of value for the research presented here. Of further importance are the detailed descriptions, within the article, of all design parameters for the experiments, allowing for the experiments to be repeated and useful insights for further research gained. This is a key learning when describing CV experiments within table tennis.

2.4.4. Tennis

Much work has been done in applying CV to tennis ball detection and tracking. For example, both Yu et al. (2004) and Yan et al. (2005) present papers on novel ball tracking algorithms. The latter method doesn't state its suitability for lighting issues, or small object of interest (OOI) such as that found in table tennis. It also doesn't describe other algorithms against which it performs better. The former (a trajectory based algorithm) works in broadcast video, reviews the environment of the game and applies various 'sieves' for ball tracking. The authors suggest not applying the more common 'object' based detection algorithms; instead evaluating all moving objects and assessing their trajectories for candidate objects. The approach, they state, is more accurate in doing so, although they offer no direct comparison to other algorithms for reference.

For low quality, off air tennis video Yan et al. (2005) describe a novel method of elliptical object detection and refining of the ball trajectory. They report experimental findings that the algorithm is of sufficiently high quality for tennis annotation. However, this experiment only covered video data from a single tournament (2003 Australian Open) – and as they state, this is insufficient data for ground truth. The algorithm does show promise for table tennis data and, if used in optimised video sequences, could be relevant. All development was made using C++ with the OpenCV (Itseez, 2015) library set, within the CVLab (EPFL, 2013) at Ecole Polytechnique Fédérale de Lausanne.

Yan et al. (2008), aiming to track the ball in broadcast video, describe a problem space for tennis which has many similarities with the research problem presented here for table tennis. With a lack of control of the video hardware and camera positioning, they discuss issues of occlusion behind players, difficulties of tracking the trajectory of the ball when its direction can easily change, the high speed of the ball causing blurring to the point where it can blend into the background especially within the first few frames and multiple false positives detections created by the rackets, the players and background facets. This article also recognises the requirement of knowing the background data prior to segmentation, something which is not always achievable. The authors declare that the ball tracking is too challenging for traditional background segmentation techniques. As a solution, the approach taken is one of offline multiple object tracking in monocular sequences, using background subtraction and Kalman Filters (Kalman 1960), this relies on a very small number of candidate ball objects in each frame. The model tracks all moving objects, frame by frame, and plotted in 3D column-row-time co-ordinates. The most likely track of the ball is then calculated from all the potential objects provided. This concept has potential for further research into the table tennis domain.

However, even with this close approximation of the problem space, table tennis has additional complexities of the unusual physics of the flight of the table tennis ball, increased occlusions by players (particularly in doubles) and the speed of the ball in such a confined space (consider the ball moving across the frame in a very short space of time before expecting to be returned).

2.4.5. Football (Soccer)

Pallavi et al. (2008) describe a method of detecting a football from broadcast soccer videos, discovering that the attributes of the ball (size, colour, shape) vary from frame to frame. Their solution was to focus on the ball trajectory as a key component in successful video analysis. Due to the low resolution of the broadcast and the complex backgrounds, they dismiss previous work involving applying a general Hough Transform with a neural network classifier, template matching and Kalman Filter (Kalman 1960) based techniques. Instead, they employ a CHT, then filter false positives using Optical Flow (OF) (Horn & Schunck 1981) velocity, camera motion and background subtraction. Any frame where the ball cannot be detected using CHT is dropped. The method goes on to describe the application for shot classification and of ball trajectory estimation using dynamic programming.

Although not being applied directly to table tennis, the proposal to avoid Kalman filtering, neural networks and template matching due to complex backgrounds and low resolution is worth investigating. There is one major difference however: football camera operators will inevitably move the camera to follow the ball, keeping it towards the centre of the captured image. This makes sense for football as the pitch is relatively large and the camera is generally some distance away. Using this manually operated camera ‘tracking’ allows the ball to be detected by manually creating a region of interest (ROI). With table tennis, the viewing area is greatly reduced, the camera is much closer to the action, and the ball is moving too rapidly to follow accurately. As an indication of the complexity of one particular ball tracking and single event analysis system (the ball crossing goal line) proposed by D’Orazio et al. (2009), four cameras capable of running at a frame capture rate of 200 Hz are placed on the four sides of the goal, with four parallel processing nodes computing the images in real time. This sophistication is beyond the remit of this research, but does highlight the obvious need for a simpler solution for table tennis at club level. What can be taken from this is the further study of CHT for ball detection, with a neural network algorithm for recognition and a custom tracker employed once the ball has been identified.

Yu et al. (2003) have taken the ball detection and tracking processing algorithms used for tennis and also applied it to football with success. The importance of this re-usability mustn't be neglected. An algorithm, which can be incorporated with broadcast quality video (uncontrolled) across at least two major sports, brings tantalising suggestions for extensibility. Whether this could be applied to table tennis, however, is potentially untried and requires further investigation and experimentation.

2.5. Commercial CV systems

CV algorithms are now widely used to track the players, vehicles and/or ball in many sports, often to support umpire decisions and implemented at the request of a team or player to attempt to alter the decision made by an umpire or referee. A literature review of CV and sports technology would not be complete without a brief assessment of systems designed for commercial purposes. Although the details of the algorithms used are often highly protected behind intellectual property concerns, their existence provides some clues as to the hardware and software used, which may offer useful insight into this CV and table tennis research domain. The requirements of such systems also provide indications of the expectations of a CV based solution applied to sports. The majority of commercial CV sports solutions focus on providing referee decision support, performance analysis or broadcast annotations (Yu & Farin, 2005). There is little need for the application to only detect a ball; it must also generate object shape, size, location and sufficient data for motion analysis, which are essential elements for applications in performance review, automatic annotation, coaching and umpiring decision support. Below is a description, with available public knowledge, of key commercial CV solutions relevant to this work.

a) “Hawk-Eye”

“Hawk-Eye” is a CV based product developed since 1999 (Hawk-Eye Innovations Ltd, 2013) and subsequently purchased by Sony in 2011. As previously discussed, it was initially developed for cricket as an officiating system, it is now implemented across sporting activities including tennis, football, snooker and most recently athletics. In all of these sports, Hawk-Eye requires calibrated high speed cameras (six in cricket, seven in football, for example) positioned around the arena, pitch or field. For tennis (problem requirements closely matching those of table tennis), the source data passes through four stages: (1) 2D information from each of the cameras is used by CV processing to detect the centre of the ball, (2) 3D data is calculated using triangulation from all available cameras, (3) this

is repeated for each frame to create a motion sequence, finally (4) the trajectory is superimposed over the court using virtual reality software to determine the location of the bounce of the ball.

Hawk-Eye compensates for ball ‘squashing’ and skidding as it hits the surface of the ground, camera shake (by detecting the lines around the court), sunlight and shadows with a claimed error in tennis of 3.6 mm (Hawk-Eye Innovations Ltd, 2015). Hawk-Eye famously incorrectly modelled the wrong bounce during the Indian Wells 2009 Murray V Ljubicic (Hawk-Eye Innovations Ltd., 2009) tennis match. This process and software limitation was subsequently corrected, but it does show that even with extensive research, investment and clearly defined remit required by the top CV applications, the systems get it wrong. According to the Director of Tennis at Hawk-Eye Innovations, much of the ability to manage variations in lighting, ball colour and shadows is not automatic but requires ‘human intervention’ (Aggas, 2013). The frame rates of each installation vary between sports, with cricket requiring a higher frame rate due to its trajectory predication capabilities. Hawk-Eye is not currently used in table tennis.

b) **Virtual-Eye**

Virtual-Eye is used as a commentary aid rather than a decision making tool developed by Animation Research Ltd., primarily for sailing, cricket, golf and motorsport. For cricket video analysis, the cameras track the ball at 230 frames per second (FPS) (Animation Research Ltd, 2013a) which in turn is locked to the live camera to detect contact between the cricket bat and the ball. As demonstrated in this article, the automatic tracking may fail close to the point of impact, but extrapolation techniques provide an estimate of the predicted contact frame. With cricket, the solution has software to show field positions, scoring, and bowlers pitch maps (Animation Research Ltd, 2013b). Although Virtual Eye does have ball tracking capabilities, in the case of golf it is able to simulate the ball trace with a combination of radar (for speed) and aerial stereo photography (Animation Research Ltd, 2013c).

Virtual-Eye is capable of broadcasting event tracking data via the internet, either live or through re-playing historical data, via its Unity Web Player. Speeds, regular tracking locations, distances between leaders and ground/sea overlays are available to offer additional data to the spectator. The techniques incorporated in Virtual Eye depend on a combination of GPS and CGI technologies and do not incorporate CV technologies. However their visual representations of sport meta-data, in particular for the replaying of historical events such as the Americas Cup (America’s Cup Event Authority

LLC, 2013), offer the potential for successfully providing metrics feedback for coaching support and spectator experience enhancement. Virtual-eye is not currently used in table tennis.

c) **SportVU**

Developed by Chicago based Stats LLC (2013) their player tracking and forecasting 3D video data analysis solution, SportVU (2012), focusses on tracking players and generating game statistics. It calculates player's distance, speed and the relation (distance) to the ball across several sports including football, basketball and American football, with the target customers being media, broadcast companies and also coaching organisations.

Available as two offerings, implementation can either be basic or complex. In the basic option, three small high definition cameras are installed in one location, separated by a few centimetres and software identifies objects in the field of view and extracts 3D positioning data, creating a real-time stream of each object's movements, generating hot spot maps of match activity. The complex solution is based on six high definition cameras installed as triplets in two locations within the venue. In both cases, the cameras can either be fixed or portable, each camera records a third of the field, and all 6 cameras are combined to form a panoramic view. For the sports being analysed, successful ball tracking operates at 25 FPS, player tracking at 15 FPS. A fourth sensor is used to detect new objects, such as substitutions. The system requires an operator to initially identify the players, referee, etc. There are no sports identified by SportsVU which have match-play speed or relative dimensions approaching anything close to that of table tennis (at the time of writing, the fastest National Basketball Association player travelled an average of 4.8 mph (STATS LLC, 2015)).

d) **SportVision**

SportVision (SportVision Inc., 2012) has developed a series of tools and techniques for tracking objects across a range of sports, including baseball, American football, NASCAR motorsport and horse racing. In providing real-times metrics and graphically enhanced images for live events, their customer base has traditionally been media companies but more recently the domestic viewers through portable consumer level computing devices. For SportVision's baseball technology, PITCH/fx, measures release rate, spin, velocity, and movement of the ball's pitch at 60 points during the flight of the ball, using three high speed cameras installed in every stadium. This technology is used for broadcast overlay graphics and for coaching (Smith, 2012), measuring average ball speeds of 93.05

mph within an accuracy of 2.5 cm. When compared to table tennis, the high speed and significant rates of spin of the ball have similar characteristics. However the technology is not used in table tennis.

2.6. Non-CV table tennis technologies

Throughout this literature review several table tennis and CV references have been found which are useful for providing a broader foundation in the early discussion within this area of research. One such article (Okumura, Oku, & Ishikawa, 2011) concerns a novel approach to tracking a table tennis ball by applying a gaze double mirror solution to feed the light from the ball to the camera. The OOI tracking works by thresholding a Hue-Saturation-Value (HSV) (Joblove & Greenberg, 1978) colour image to convert to binary. The location of the object is then sent to the processor to continually adjust the mirror angles, maintaining the object in the centre of the lens. This is repeated every 1 millisecond. Taking the approach to a fixed camera with a moving mirror gaze allows for very fast image capture of a fast moving table tennis ball without the general issues of motion blur. If the frame rate could be reduced to considerably less than 1000Hz then this may allow for a longer shutter speed and hence work in the lower light levels found in non-dedicated table tennis environments. In the accompanying video, it is noticeable that there is no background noise with a flat colour background. There is no evidence of event detection or location measurement. These latter constraints are not discussed by the paper, but of value is the suggested method of thresholding.

A sophisticated humanoid table tennis robot (Sun et al., 2009) developed by researchers at the Zhejiang University tracks a ball using binocular vision. The tracking algorithm is based on learned trajectory patterns, recording using a Photron Fastcam (Photron USA, 2014) operating at 2000 Hz to capture the trajectory. Given this knowledge base, the ball's flight trajectory is calculated using just a few points captured by the robot's binocular vision. These binocular cameras operate at a much slower 120 Hz, send the images to the processor and calculate the trajectory within 50-100 milliseconds, with an accuracy of 2.5 cm. Applying this tracking knowledge base technique is out of the realms of all but the most ardent and resourceful table tennis organisation, but does demonstrate the potential for knowledgebase application and real-time trajectory estimations.

2.7. CV algorithms

Whilst reviewing the current literature in this research area, the amount of image processing algorithms, their definitions and variations in CV terminology, is not insignificant. When compared to many other scientific disciplines, this area of research is young; the common language has yet to become established, the semantics are shifting and much of the real-world evaluation, validation and implementation lies ahead. Even defining CV is not straightforward. It has been described as ‘an enterprise that uses statistical methods to disentangle data using models constructed with the aid of geometry, physics, and learning theory.’ (Ponce & Forsyth, 2012a). CV is therefore not restricted to pure computer science. It incorporates physics and mathematics, and also psychology and biology when analogous to the understanding of human vision (Pingle, 1969). Furthermore, recognising that symbolic information extracted from CV processing cannot be separated from the image creation process is essential to realising the capabilities and limitations of a CV solution. Knowledge of the recording hardware and configuration is therefore necessary to accurately interpret the data extracted; this is especially true with image sequences. This makes it an exciting scientific area of research, but also challenging to consolidate. As such, this section provides an essential summary of CV algorithms as discussed in the application of table tennis and other sports within this thesis.

2.7.1. Low level feature extraction (edge detection)

Nixon et al. (2008) define low level feature extraction using edge detectors as ‘basic features that can be extracted automatically from an image without any shape information’. By examining visual cues within the raw image data image, such as texture, colour or luminance, inferences are made for the boundaries surrounding distinct objects. Several algorithms have been developed and, over time, improved. Most consider the grey level profile of an image to detect ‘significant, local change in image intensity.’ (Morris 2004). The earliest of these is the Roberts operator; simple to design, yet is sensitive to noise. The Prewitt (Prewitt, 1970) and Sobel Operator (Ballard & Brown, 1982) detection algorithms, accordingly, are designed to account for noise by incorporating a degree of pixel based smoothing. The Canny (J Canny, 1986) and Laplacian of Gaussian (LoG) (Marr & Hildreth, 1980) algorithms improve upon this by either applying a Gaussian template to computationally smooth the image or by double differentiation to detect zero crossing, respectively.

These final two methods are considered to have become ‘optimal edge detectors’ (Morris 2004). However, no assumptions have been made at the outset of this research as to which edge detector

would perform most accurately. Without prior investigation of edge detector performance comparisons in table tennis, it is not known, which would be the most suitable (if any) at describing edges within the image, or perhaps if an edge detector alone would be suitable for the broader aims of ball detection and event classification.

2.7.2. Foreground extraction

There has been no thorough research to evaluate the success of foreground extraction, when applied to table tennis. As such, it is suggested here that at least two foreground extraction algorithms are considered for evaluation. The first of these is the well understood and widely implemented CV Gaussian Mixture Model (GMM) (Stauffer & Grimson, 1999) as presented by Kaewtrakulpong et al. (2001). For GMM based segmentation the image is first converted to grey-tone and each pixel in the image is modelled as a mixture of Gaussians. The Gaussian distributions are then evaluated to determine whether the pixel is to be a part of the background, or a part of a moving foreground object. A foreground mask is then created for each frame. This method is also selected for its well established use in outdoor tracking with varying lighting conditions. Its implementation within table tennis has not been documented and so this is considered a novel implementation. With its broad use in many other applications, however, the GMM algorithm provides a benchmark for other algorithms to be compared.

The second method, Optical Flow (OF) analysis, is derived from the biological visual systems of mapping continuous flow of the effects of visible light fluctuations across the retina (Ballard & Brown, 1982). In CV this is modelled as the mapping of pixels in continuous frame data to ‘instantaneous’ velocities for each pixel across the image field. It is worth considering that OF algorithms require constant intensity throughout the motion of the ball (O’Donovan, 2005). For evaluation and comparison, two common OF implementations, the Horn-Schunck (Horn & Schunck 1981) and Lucas-Kanade (Lucas & Kanade, 1981), are of primary interest.

2.7.3. Object detector

The low level feature extractors and foreground extractors are able to segment the image based on homogenous regions (detected boundaries are ‘filled’ to define the area within the region). This alone will not detect the desired objects within the image. For this, the images must be further processed to detect the ball based on known features, such as colour, shape and location. In current

research within table tennis, two primary methods are most common: (a) a simple CHT (Duda & Hart, 1973) to detect circular regions, and (b) a feature filter based on shape properties, such as eccentricity. Feature processing alone may produce more than one candidate object to evaluate.

2.7.4. Tracking

Once an object is detected, tracking it in a video sequence by the prediction of the likely location in future frames has the potential to improve the processing speed and quality of outputs. Tracking generally falls into three categories: (1) detection, (2) matching (3) region of interest analysis, and (4) Kalman filtering (Kalman 1960).

Tracking by detection is useful for instances where only one object in each frame is detected, as instability increases with false positive locations. Reliance is then placed on the quality of the detector, where combinations of features of the desired object are unique within the image. Matching computes a summary appearance of a domain area within the image and attempts to find the best match within subsequent frames. Flow-based matching searches for a transformation of a set of previous pixels in the known domain to a good match in the new domain. Using flow-based matching allows the generation of robust motion models (Ponce & Forsyth, 2012b) and is applied in the Optical Flow extractor described earlier. Region of interest considers an area around the detected object and calculates and, by extrapolating for frame rate and speed of the object, computes the location of the object's bounding box in the next frame. Finally the Kalman Filter exploits predictable dynamical information to determine the approximate future search domain of a moving object, given an initial object state and observations. If the motion of the object is predictable, there is not the requirement for a high frame rate or a slow moving object. The location of the object in the next frame could be far away from the current frame and its location may still be determined. This is an improvement on region of interest or matching, where success is improved when the object has not moved substantially between frames. However, for Kalman Filtering to be of benefit, the motion must be predictable.

2.8. Literature review findings

Literature searches for CV applied directly to table tennis produce few results. The available research material in this domain has proposed several methods for detecting a ball from an image. However, the proposals are not conclusive, lack detail and application, and do not investigate the

implementation possibilities as shown by CV in other sports. These gaps in the current knowledge are summarised below.

- a. There has been no thorough analysis or comparison of CV algorithms for the detection of the motion of a table tennis ball during match-play. Subsequently there is no comparative ranking index to measure the success and suitability of a chosen algorithm.
- b. The application of CV in a low-cost, monocular recording environment, with the aim of performance analysis, coaching and officiating, is minimal.
- c. There has been no full assessment of table tennis match-play event classification. The service is the only event which has been assessed to date, with only theoretical proposals for measuring its legality. An understanding and compilation of the inherent challenges and issues arising from CV based ball detection in table tennis, throughout all events in general match-play, does not exist.
- d. There is no detailed description of an optimised single camera configuration and positioning for the 2D based approach of ball detection using CV. The videos used to date are not fully documented, with no reference to the camera hardware specification, positioning or lighting requirements.
- e. There has been no attempt to validate the accuracy of ball detection and its motion in 2D table tennis image sequences.

This research fills these knowledge gaps by detailing rigorous experimental designs for the recording of table tennis video sequences, from which a direct comparison of CV algorithms can be made. It also proposes a method of validating detection of motion change during key match-play events using 2D monocular data. Furthermore, this research presents solutions for managing occlusions, reflections, and false detections across a range of carefully documented image sequences, recorded within laboratory conditions. Finally, a solution for table tennis event classification using monocular recording devices is tested. This report, with the fully maintained literature review, highlights these (and other) gaps in knowledge within the CV community, with sufficient detail to ensure the findings, solutions and proposals by the author are unique and novel.

2.9. Summary

The application of CV technologies applied to many sport is of great interest and investment, from the spectator, player performance and coaching perspective. There are many commercial products available for the major worldwide sports, many of which produce automated analysis and statistics on match events; others are there to supplement umpiring decisions – to assist in an unbiased, objective and reasoned decision making process. Existing commercial technologies could, in theory, also be applied to table tennis yet, for many reasons discussed, they are not implemented due to investment requirements and the portable nature of the playing area.

Whilst there is a wealth of knowledge for ball detection in sports generally, academic research investigating ball detection and segmentation with regards to table tennis is confined to a handful of articles focussing on a small number of specifics, generally using uncontrolled video sequences gathered from public internet sites. This literature review has found no research into systematic experimental observations for table tennis ball detection using CV processing of documented image sequences. There is no CV algorithm comparison methodology – and hence no standard ranking capabilities. Furthermore, there is no documented evidence of the validation of event detection using 2D monocular image sequences, based on the ball dynamics of speed, accelerations and the detection of sudden changes in direction due to a let, bounce and return. With all of these areas of interest relevant to a successful, low-cost, implementation of CV technologies for officiating, automated scoring and performance assessment implementations, a gap in the current academic literature is presented.

3. Design methodology and schedule

3.1. Introduction

Despite notable advances in the application of Computer Vision (CV) in sport, no substantial evidence has been found of thorough experimental and observational research of CV in table tennis. As such, presented here are the experimental design considerations and practical undertakings necessary for generating table tennis event source data used in CV optimisation, ranking, performance review and officiating. Each experiment is designed to record a specific key event (or combination of events) within a typical table tennis match. The recordings can then be processed in an isolated and controlled method to detect the ball, its location, speed, acceleration and bearing. The resulting outputs provide the data necessary for robust analysis of CV based event detection and classification.

Previous work attempting table tennis ball detection has often used video sequences captured from the Internet, for example from YouTube. The use of online image and video libraries for source data is to be discouraged due to image degradation caused by undocumented image and video compression processing during file upload. Exchangeable Image File Format (Exif) data defined by Electronics and Information Technology Industries Association (2014) is also at risk of being lost during any potential file format conversion. In the work of Wong (2007) videos from The Umpires & Referees Committee of the International Table Tennis Federation (ITTF) (2002) were used as the source. In this case, the *potential* for CV algorithms and areas of further investigation can be evaluated, but as key video parameters (such as camera hardware and location, shutter speed, ball size etc.) are not provided, a critical algorithm appraisal for repeated, non-subjective use, is not viable. To improve upon this work, it is suggested here that all video recording settings are documented, along with the measured location of the recorder relative to the table. As such, all video data used in this research is from video sequences created specifically for this project. Details of the recording schedule, experiment design consideration and specific camera configurations are documented.

To fully evaluate the CV processing requirements, a number of table tennis simulated match-play events have been captured on video using multiple recording devices, with each generating source data of differing quality. Each video recording was assigned to a series, with each series designed to capture a specific set of match-play events. As required, a video recording was divided into a num-

ber of individual sequences. Each video sequence was given a unique identity prefixed by the series identifier. There were eight series in total, captured over a period from June 2012 to April 2014. Each recording session was incremental in event complexity, based on the previous outputs, findings and data gaps. Outputs from each session additionally provided test data for the bespoke software workbench (TRASE), designed specifically for CV comparison and motion analysis in this research.

3.2. Preliminary video sequence capture design considerations

3.2.1. Philosophies

In testing the hypothesis in Section 1.3, that events can be classified using only changes in motion data created from monocular recording devices, importance was placed on the origination of the video sequences. From the outset, these experiments adhered to the following three design constraints and philosophies.

- a. **Always consider the match-play environment.** Match-play events were recorded in environments that are typical of those in which table tennis is played. Environmental conditions should not unnecessarily be sanitised of all CV processing complexities; this is avoidance of the challenges of the research. For deeper analysis and proof of outcomes, isolation of environmental characteristics was made as necessary.
- b. **Experimental, repeatable comparator data.** Videos sequences were recorded consistently within their given series, reducing variables to the point where the only variable in any direct comparison is the CV algorithm itself. Several recordings were made of the same event (e.g. a let) with no change in camera configurations. If an environmental factor, such as the background, was changed, the subsequent recording was placed into a new group.
- c. **Record and document.** Any measurable variations in recording hardware location, lighting configuration, exposure configuration, selection of apparatus and player performance were recorded.

With no previously reported work to use as a benchmark, a series of pilot studies were undertaken to reduce the potentially broad scope of experiment design. These initial studies create the domain space from which key configurations, layouts and event parameters were defined. To establish suitable frame rates, shutter speeds and lighting positions, a more flexible approach was required from

these pilot tests. As each pilot study is completed, each recording specification defines a recording session with each session designed to be successively incremental in complexity. It must be noted that, as with all experimental research, not every conceivable table tennis event can be re-created in a laboratory environment for analysis. However, this incremental approach allows for initial indications and proposition statements to be made against the most common types of scenario where CV may add most benefit, whilst allowing future research plans to be documented.

3.2.2. Technical constraints

In approaching the design of the experiments there were a number of technical possibilities based on current literature in other sports which may, in theory, offer useful results in this domain. However, when assessed for their suitability using pilot studies and/or simple experimentation, the following were removed from consideration early on from the experiment methodology.

- a. **Source professional videos.** A review of CV in table tennis could be made using professionally made coaching or broadcast videos, such as those used by the ITTF. However, there is no control over camera positioning, lighting conditions, ball colour, exposure settings or resolution. They are not designed for source data in experimental CV analysis.
- b. **Computer graphics model.** One approach would be to not record the ball in a real match environment, but generate images using physical modelling and computer graphics. Although this would give the ability to validate the location of the ball with a high degree of accuracy, it could falsify the CV processing. Natural image artefacts, such as background noise, resolution, indoor lighting and motion blur would be complex to model. Simulating ball motion due to spin, speed and atmospheric variations would also be open to design complexities beyond the current scope.
- c. **Artificially reduced quality.** Using post-recording image processing techniques to reduce the quality of the high-end recordings, by blurring images or superimposing foreground obstacles, would create imaginary CV application challenges and is therefore of little value. However, managing and adjusting the recording configurations *at source* to reduce overall contrast, frame rate and pixel resolution is of interest. This includes configuring a high specification camera to run at lower frames rates and/or shutter speeds than it is capable of, or reducing the image resolution at the time of image capture, for example, matching that of low-cost recording hardware.

- d. **Strobe analysis.** It has been shown in previous work that using strobe lighting creates simultaneity between camera devices, capturing a motion at any given moment. For elements of ball analysis, such as developing algorithms to measure spin rate, this may be suitable. However this approach would mean the players would be working in the dark – a challenge for even the most adept professionals. Additionally this does not allow for the testing of realistic lighting conditions and evaluation of the camera functionality using low-cost recording hardware. Therefore this has not been included.
- e. **Optical markers.** In a study by Theobalt et al. (2004) applied markers to a baseball for use with strobe analysis and some consideration was given to using markers with standard lighting and camera technology. For the work presented here, early attempts were made to apply markers in standard lighting to a table tennis ball. However the high ball speeds, small ball size, glare, reflection and motion blur made the markers very difficult to detect, unless the camera was placed close to the point of action.

Implementing any of the above technologies and techniques either adds unknown variables into the experiments, or it requires an environment in which the game could not be played to ITTF guidelines. For these reasons, it is a proposal of this study that they are not used when evaluating and comparing CV algorithms for table tennis ball detection.

3.2.3. Ball imaging considerations

Every sport has known characteristics, e.g. ball diameter, playing surface texture and dimensions, which can be used as image processing features to filter the object of interest (OOI) within a given image. When applying image processing to detect the ball in table tennis for example, there are four particularly important ball characteristics: colour, size, speed and shape. Each of these characteristics is a consideration when determining recording hardware settings and configuration. The factors are described in detail as follows:

a. Colour

Official table tennis balls may be manufactured in either white or orange and in matte only (ITTF, 2011). Measured and calculated according to the CIE Lab system (Hoffmann, 2013), the differences in **L**, **a** and **b** (black/white value, green/red value, blue/yellow value) when compared to the Munsell colour standards (Munsell Color, 2013) for white and orange are given in Table 3-1 below.

Table 3-1 Munsell colour standards for approved table tennis balls¹

Characteristic	White	Orange
Munsell notation	N9.5/ M	7,5YR 8/10M
L	-21 ≤ ΔL	-10 ≤ ΔL
a	-6 ≤ Δa ≤ +5	-15 ≤ Δa ≤ +10
b	-12 ≤ Δb ≤ +5	-15 ≤ Δb

When processing the image, a background close in colour to that of the ball, will cause ball boundary detection difficulties as the ball transits those areas. In general, a white table tennis ball does not contrast well against a complex background (non-distinct from overall ambience). This is due to the prominence of white in images, not only from object colour but also from image saturation. Even though the surface of the ball is designed to be of a matte material or coating, they have a tendency to appear over saturated in the image during full image exposure balancing, due to the limited dynamic range of image sensors (Hu, Gallo, Pulli, & Sun, 2013)². Saturation additionally affects the colour presentation of the ball, increases the apparent diameter and distorts its shape. Experiments presented here have shown white balls appearing yellow when in lower light zones and orange balls appearing white when highly lit. As such, relatively simple extraction techniques using the ball's colour features are prone to a variety of uncertain outcomes. This is supported by the work of Wong (2009), where detecting the ball based on the colour and k-means clustering (Kanungo et al., 2002) processing failed to provide conclusive results.

b. Size

The standard table tennis ball diameter is 40mm, having increased in February 2000 from 38mm in an attempt to reduce the speed of play (Takeuchi, T., Kobayashi, Y., Hiruta, S., & Yuza, 2002). Relative to the typical size of the field of view (FoV), particularly when aiming to capture the entire area of play within one single camera, the ball is small. With a standard frame aspect ratio of 4:3 and assuming a rectilinear³ camera lens positioned to just view the full length of the table (2740 mm), the recorded frame will cover an area of 5,630,700mm². If the FoV were to be widened to include two 1000 mm player zones, the area would increase by approximately 300%. In comparison the square

¹ For a description of the Munsell colour notation see “A Colour Notation” by Albert Henry Munsell (Munsell, 1905)

² Although not included in this research, recent developments in HDR exposure may improve this (potential for performance degradation would need to be assessed).

³ For lenses with barrel distortion, the objects at the edge of the image will appear smaller than those at the centre and will curve towards the centre. The effect increases the total FoV area, which should be minimised by careful lens selection.

formed by the bounding box around the ball, placed in the centre of the FoV, covers an area of approximately 1600 mm² (40x40 mm). This means, at least initially, attempting to identify a ball having a minimum area of 0.028% of the entire frame (not allowing for fluctuation in ball size due to depth of the frame, lighting or lens distortion). However, if the ball *can* be segmented from the image with sufficient accuracy, the size of the ball becomes a known scale reference when calculating changes in motion.

c. **Shape**

Ideally the shape of the ball should be circular in all frames. However, consideration must be given to the fact that this is only the case when using a combination of very high frames rates, high ISO values, short focal ratio lenses and low ball speeds. At lower shutter speeds, the ball travels a significant distance across the frame whilst the shutter is open. This distorts the image of the ball along its path. Additional elliptical distortion occurs as the angle subtended between the lens, the ball and the light source increases. Finding the precise centre location of the ball from these distorted images requires a number of additional processing considerations. As the shape of the ball is no longer spherical, decisions must be made as to how to calculate the location of the image of the ball (for example, the centroid of the object, or the mean of the visible axis). Once the centre of the ball is known, its perimeter can be calculated; this is of importance when evaluating the point of contact with the net or doubles line.

d. **Speed and spin**

There are several estimates for the maximum ball speed achieved in a table tennis match, with a consensus approaching 110 km/h (Turberville, 2004). At this highest speed the ball is moving 30ms⁻¹; a frame rate of at least 750 Hz is required for the change in ball position between frames to be less than 40 mm (the diameter of the ball). However, research carried out by Iino et al. (Iino & Kojima, 2009) describes an advanced player as able to hit the ball at a mean of 18.7 ms⁻¹ and an intermediate player at 16.7 ms⁻¹. This results in a frame rate of approximately 420 Hz necessary to resolve ball movement less than that of the ball diameter. A high speed smash shot takes approximately 0.09s to travel the length of the table. A shutter speed of anything less than 1/420th second will record ball travel distance as motion blur (ball movement during open shutter) greater than the diameter of the ball itself. The blur causes the ball image edges to be less distinct, creating difficulties in thresholding, and the centre of the ball image to be difficult to derive. The mean hue of the ball image also

changes as it blends with the background. It is at the higher ball speeds where an officiating support solution can add greater value to the game, as rule infringement detection by the human eye becomes increasingly difficult. There are also variations in the reported spin rates of the ball during a 'chop' serve. Using five elite players from the Singapore national team, Lee & Xie (2004) measured spin rates of up to 64.1 revolutions per second (rps) for a 38mm ball and 59.6 rps with a 40mm ball, with the larger ball losing spin more rapidly due to increased surface area i.e. with an increased drag (Fullen, 2010). Using a high speed camera operating at 1000 Hz, Yoshida et al. (2014) discovered speeds by world class players of 50-60 rps for men and 40-50 rps for women. With the tactical importance placed on spin, particularly during the serve, the ability to reliably detect this would add significant value to an automated CV based performance feedback solution.

3.2.4. Camera setting considerations

An understanding of the seven key camera settings of focus, shutter speed, frame rate, lens speed, focal length, ISO speed and sensor design is important. The high speed of the ball, in particular, is demanding of most hardware capabilities. When using low-cost devices in a typical match-play environment, ideal camera settings may not be achievable and a considered compromise is therefore justified. Lower specified camera settings will introduce additional complexities in detection and quantification and this is a rational expectation and a viable constraint of this research. Furthermore, each setting is often intrinsically affected by each other setting and cannot, therefore, be considered in isolation. A brief discussion of each parameter and the considerations for the experiments follows.

a. Focus

The ability to focus the camera lenses consistently and repeatedly across all trials is of critical importance to ensure a) the boundary contrast between the ball and any background feature is clearly defined, b) errors are reduced when measuring the diameter of the ball and c) comparisons of detection algorithms are controlled. For these reasons the Siemens Star Chart (JVC Professional Europe Ltd, 2012) focussing chart (as depicted in Figure 3-1), laser printed at a resolution of 1200dpi and a diameter of 25cm, was used to focus the camera on the nominal object plane of interest at the beginning of each recording session.



Figure 3-1 Siemens Star Test Chart

Unless pinhole cameras are used, all lens based digital cameras focus across a depth determined by the lens focal length, aperture, distance from subject and size of sensor (McHugh [1] n.d.). Once the camera was focussed on the centre of the table there were areas within the scene which were inevitably out of focus; these areas appear primarily in both the far distance and near to the lens, with a small tendency also towards the edges of the image. By using the lowest possible aperture setting and ensuring the camera is not placed close to the field of play, the depth of field/focus has been maximised.

Once the FoV and aperture are set, the chart is placed in the centre of the table tennis table under standard lighting conditions. To avoid the continual re-focus adjustments of automatic focussing during subsequent capture, automatic focus was used only with the Siemens chart; focus was then fixed on the camera by switching to manual focus.

b. **Shutter speed**

In recording a high speed table tennis ball, it has been discussed that the shutter speed is fundamental in capturing a series of ‘stills’ of the ball during its trajectory. The less movement there is of the ball across the frame, the brighter the object, the more clearly defined edges and the easier it becomes to segment. However a fast shutter speed requires more light, faster lenses (large aperture stopped to maximise depth of field) or more sensitive camera sensors; faster lenses reduce the depth of field and, additionally, increased lighting can increase saturation and enlarging of the ball within the frame.

c. **Frame rate**

Manipulating the individual image capture frequency for a given video sequence is the most direct means by which to increase the number of location data points of the moving ball; the greater the number of data points the more confidence can be given to the estimated location of the ball *between* data points. This is critical when examining the output of a table tennis video sequence for rule infringement decision support. The theoretical maximum frame rate (Hz) is equal to the shutter speed. The uncertainty of the ball location between frames is not proportional to the frame rate but increases non-proportionally with reducing frame rates. From the earlier discussion of ball speed (Section 3.2.3) it has been shown that the optimum frame rate, purely from distance travelled between frames, should be at least 420Hz (less than one ball diameter of travel between frames) for the professional level of play. However, increasing frame rate requires more light on the scene to get sensible levels of exposure. More broadly, selecting appropriate frame rates directly affects settings for shutter speeds, lens selection, CV processing overheads and light. In these experiments which are designed for low-cost implementation with club level proficiency, frame rates in the region of 100-200 Hz were considered acceptable for evaluation.

d. **Lens speed**

The increased illumination at the focal plane offered by a lens with a wider aperture allows more light to enter the frame for any given shutter speed. More light entering the image will enable a reduction in gain (and hence the reduction in noise) on the camera's sensor, with the added benefit of potentially reducing shutter speeds and the need for a sensor with increased sensitivity. A faster lens speed yields an improved contrast in low light levels. This will also benefit the requirement for a reduction in shutter speed.

e. **Focal length**

As focal length reduces, the wider the FoV becomes, allowing the camera to be placed closer to the play area/table. The closer the camera is to the table the larger the apparent ball size and the more accurate any measurements become. However, as lens focal length decreases, so does lens curvature thus distorting the ball as it moves further away from the centre. A balance between focal length and apparent image size is justified. A lens with a long focal length (a zoom or telephoto lens) used at maximum setting, with the camera placed further away from the table, will produce a

flat FoV, reducing image curvature and compressing focus. However, it is recommended that only optical zoom is applied, not digital zoom. This will preserve the image resolution.

f. ISO speed

Increasing a camera's sensitivity to light reduces the need for low shutter speeds, increased lighting and high-cost small focal ratio lenses. With an increase in ISO speed there is an increase in signal-to-noise (SNR) ratio. Noise is difficult for CV software to discern from fine texture (McHugh [2] n.d.) and this is a consideration when thresholding and filtering the image to reduce unwanted image artefacts. During the experiments, to increase SNR in the background of the image sequences in a controlled way, green netting was placed behind the playing area.

g. Sensor

The sensor size for CV based detection has a direct impact on image quality. With large sensor size comes the potential for larger photosites (the light sensitive spots on the sensor), which are able to capture more information from a scene, with an improved signal to noise ratio, better dynamic range and improved low-light performance. This point is worth considering when applying smartphone technology or webcam devices, which generally have smaller sensors. Additionally, smaller sensors capture a narrower FoV (if using the same lenses) compared to their larger counterparts. This improves focus quality by restricting the light cone to the centre of the lens. Conversely, a smaller sensor would need to be moved further away from the scene, thereby reducing the pixel size of the ball. Furthermore, the sensor resolution will ultimately determine the usable limits of the image resolution. Although a higher resolution is more desirable to capture greater definition of the ball (particularly around its circumference), this has implications in image processing bandwidth and light sensitivity of the individual photosites. Again, a considered balance of sensor size, image quality and processing capabilities is necessary.

3.3. Apparatus

For initial baseline experiments, video data gathering was performed in a managed environment within Loughborough University's Sports Technology Institute (Figure 3-2). Throughout the recording sessions (except during motion detection validation scenarios), club level table tennis equipment was used, including: 40mm white and orange table tennis balls, selection of rackets and two tables

of standard construction, dimensions and design meeting approved guidelines (ITTF, 2013a). At the start of each recording session the chosen table is placed on a hard standing area and its height from the floor to the playing surface is confirmed to be 76 cm. The net is connected to the table and its height is confirmed to be 15.25 cm above the table. The table is positioned to incorporate differing backgrounds, from a simple matt black to a complicated background with textures, colours, contrasting edges and lighting variations.

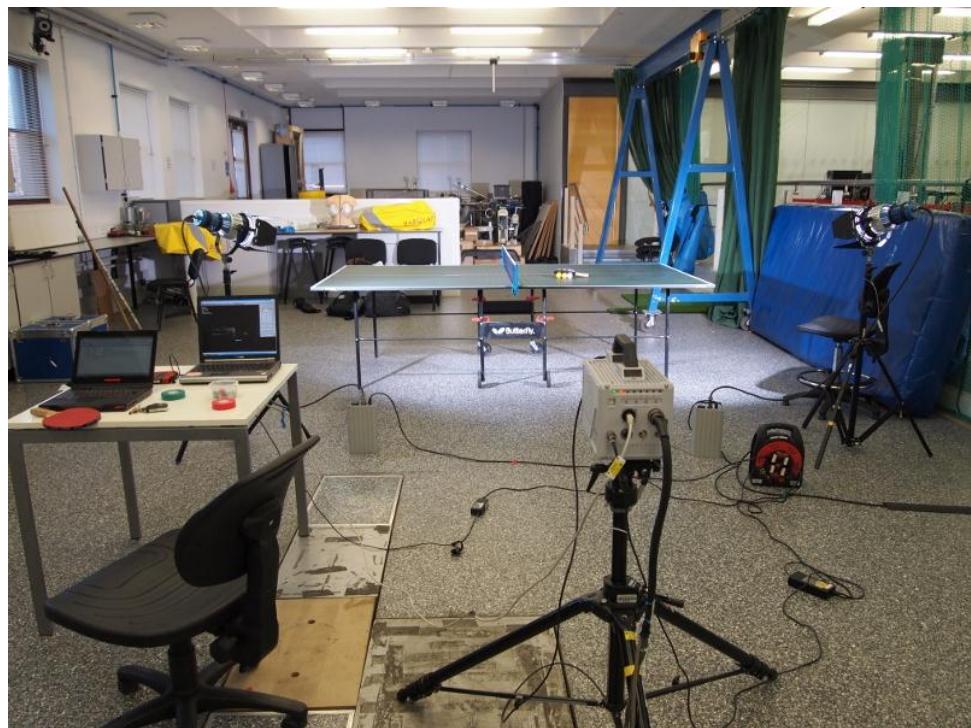


Figure 3-2 Lab with Loughborough University's Sports Technology Institute

A combination of serving machines and athletes, with a range of ability, were used. The serving machines used was a Butterfly 3000 (Butterfly North America, 2013), an industry standard in coaching and practice within table tennis. The Butterfly 3000 is able to direct the ball through two rotating friction discs; the direction of the head can be adjusted and the speed and spin can also be defined using the analogue control pad. Due to the practicalities of setting up each session over several days use of the Butterfly 3000 has the benefits of a) consistency and repeatability in shot selection and b) replicating high performance player characteristics in terms of the level of speed and spin Imparted on the ball. The athletes were primarily used for detecting issues of racket occlusions, multiple moving objects, rallies and the service.

Throughout the experiments, a number of cameras covering a range of imaging specifications, from a Philips ToUcam II webcam (Koninklijke Philips N.V., 2014) through to a Photron Fastcam Ultima APX (Photron USA, 2014) were used. The camera sensor dimensions and lens specifications used throughout this research are presented in Table 3-2.

Table 3-2 Camera sensor dimensions

Camera	Sensor	Sensor	Focal Length	f Ratio	Autofocus
	width (mm)	height(mm)	(mm)		only
Olympus E-PL1™	17.3	13	14-42	f3.5-5.6	No
Photron FastCam™ Ultima APX	12.8	10.2	24-70	f2.8	No
Canon HG10	10.67	8.0	6.1-61	f1.8-3.0	No
Sony HDR-AS15	6.17	4.55	35	f2.8	Yes
Casio Exilim™ ZR-300	6.17	4.55	4.24-53.0	f3.0-5.9	No
Philips ToUcam™ II	4.60	3.97	6	f2.0	Yes
Apple iPhone™ 5	4.54	3.42	4.1	f2.2	Yes
Nokia Lumia™ 900	4.54	3.42	28	f2.2	Yes

The Photron Fastcam Ultima APX and the Olympus E-PL1 (Olympus, n.d.-a) have interchangeable, adjustable and measureable zoom lenses; the Fastcam has a Sigma 24-70mm lens (Sigma Imaging UK, n.d.) and the Olympus a Zuiko Digital ED 14-42mm lens (Olympus, n.d.-b). The Canon HG10 (Canon, n.d.) is used both as a rolling recorder and also for basic event video capture; however, the adjustable zoom was not measureable or consistently repeatable. It is therefore referenced only for its ability to record at 60 Hz in High Definition. The Philips ToUcam II, Sony HDR AS-15 (Sony Corporation, n.d.), iPhone 5 (Apple Inc., n.d.) and Nokia Lumia 900 (Microsoft, n.d.) all have fixed lenses and auto focus only. Those cameras with measurable zoom lenses were set with effective 35mm focal lengths of between 28mm and 32mm. All cameras used within a given series were placed into position on sturdy tripods to enable a stable image containing the full table width. External lighting rigs were occasionally used, as required, to reduce fluorescent flicker and to allow for flexibility in positioning the light intensity zones. They also assisted in the control of lighting conditions during reduced frame exposures.

3.4. Lab design and layout

During the pilot studies, a number of different camera angles and locations relative to the table were evaluated. The primary requirement in finding the optimum location of the camera is to minimise occlusions of the ball; the secondary requirement is being able to detect changes in motion of the ball at key match play events, using a single video recording device. Figure 3-3 depicts several early options for camera placement.



Figure 3-3 Candidate camera positioning

Each of these candidate views has benefits at various times during match-play, but none provide suitable conditions within which the two main requirements could always be met. For example Figure 3-3a appears to present a clear view of the ball for much of the area of the table. While this is suitable for line fault detection during a doubles match on the far side of the table, it has little benefit during the majority of other events, such as the service, let detection or half of all returns i.e. those occluded by the player(s) on the near side of the table. Camera position in Figure 3-3b is the least usable location. Here, the ball is occluded by the player(s) nearest the camera for much of the time and, as the general direction of the ball is along the axis of the lens, the significance in change in

motion is reduced. Figure 3-3c and Figure 3-3d was proposed as it shows much of the ball from one player's perspective, but due to their asymmetric orientation each event description would vary depending upon which side of the table it would occur. Furthermore, the determination of changes in ball motion using only 2D data from these angles is particularly-prone to error/uncertainty.

The optimum location is one in which the 2D projected data can more readily be extracted accurately from the image while capturing the ball for the majority of events (an event is defined here as a significant change in table tennis ball motion which affects game statistics, performance analysis data, or officiating decisions). A review of the positioning of the match officials provides a good indication of the ideal place in which to position a camera for the majority of match play to be observed. With careful consideration, the optimum location of the camera with respect to the table for the majority of the recordings is as shown in Figure 3-4. The camera is placed in line with the net; the distance from the table (d) is between 2.5m and 3m, depending on the camera lens and FoV capability. The height of the camera (h) is relative to the table height and is 30cm above the plane of the table. At this location the camera does not interfere with match play, provides an option for the umpire to be seated directly opposite and is relatively easily positioned.

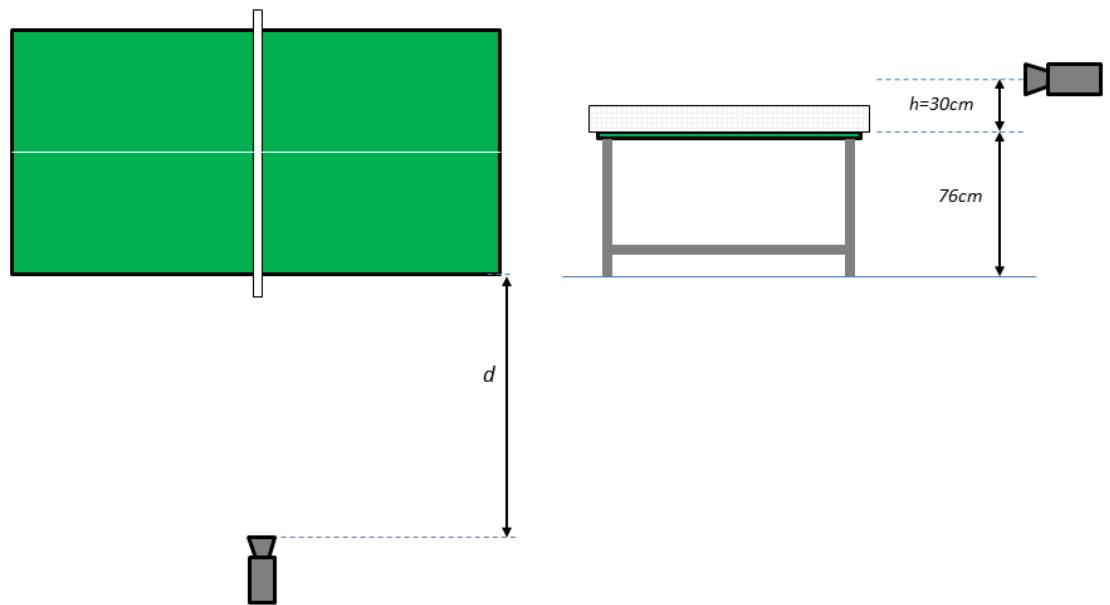


Figure 3-4 Optimum camera position

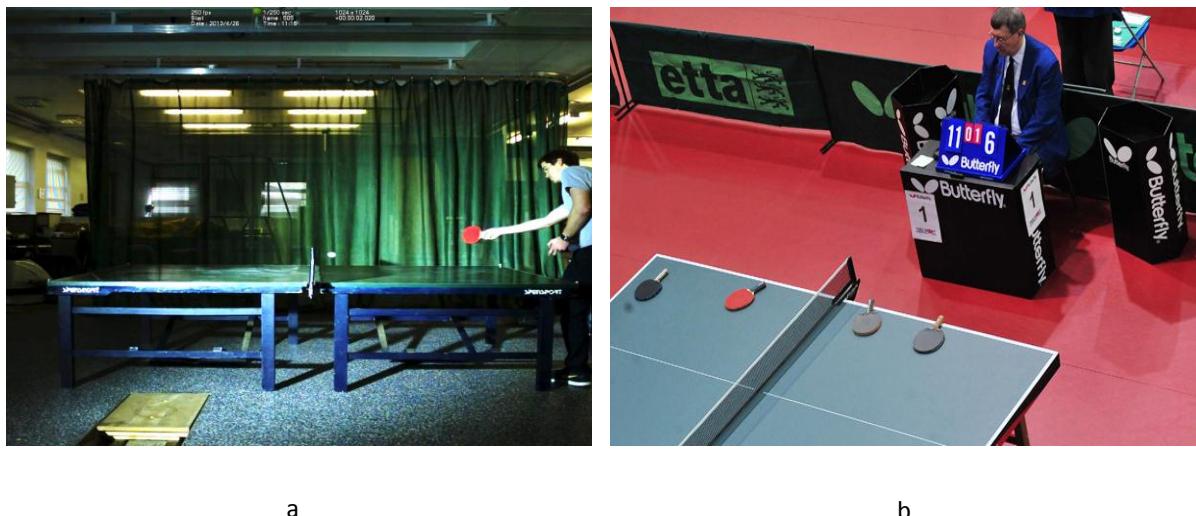


Figure 3-5 Optimum FoV (a) and umpire position (b) (Table Tennis England, 2010)

According to the Handbook for Match Officials (ITTF 2014a, p.7) rule 4.2.4 defines the umpire's position as 'The umpire should be about 2-3 metres from the side of the table, in line with the net, preferably on a slightly raised chair, although this is not essential for singles.'. It is perhaps no coincidence that this camera position approximates the location of the umpire during a table tennis match (Figure 3-5a); the location of the both the camera and umpire are chosen for precisely the same reasons. The majority of experiments involve capture of video sequences using a number of devices recording the scene in parallel. To ensure the effect of variations in camera location on FoV is minimised, the positions of the cameras (in particular their lenses) are as close to coincident as practically possible as shown in Figure 3-6.



Figure 3-6 Cameras Co-located

3.5. Detecting changes in ball motion/trajectory

The central hypothesis of this research is that the detection of sudden changes in ball motion provides an indication of a specific match-play event. Further, it states that data from CV processing of monocular image sequences, when recorded appropriately, contains sufficient information to enable the automatic categorisation of events based on the motion of the ball and its deviations. A high degree of sensitivity to changes in the ball's motion is clearly required to measure events such as a slight graze on a net or a clip on the edge of the table. The issue here is that at the limit of resolution of motion change there is no detectable ball deviation. An additional challenge of using CV processing is that factors such as resolution and ball centre calculations could also indirectly create apparent object movement, due to differences between image processing techniques, their parameterisation and, indirectly, fluctuations in environmental conditions such as lighting and background contrast. Any such processing artefacts need to be eliminated from the results to avoid false-positive event detection scenarios. Data smoothing will reduce the small variations in apparent motion caused by image processing, but, in doing so will ignore the smallest of changes in actual motion. Therefore, the approach presented here is to incorporate a secondary non-CV method of data collection. This parallel reference system does not need to provide detailed location data, but it must enable sequence to be correctly categorised. Data from the two systems can then be compared and the performance of the CV-based system verified.

Several solutions were considered for this secondary control system. These included designing a mapped background grid onto which the image of the ball would be projected, or a specifically designed, predetermined trajectory track. One problem with the grid solution is it still relies on accurate CV based image processing while the engineered track was contrived and would not readily represent various scenarios of true match play. After careful consideration the secondary motion detection system was chosen to be Vicon's motion capture solution (Vicon Motion Systems Ltd., n.d.), using 3D infrared technology. This has the benefits of it being a pre-validated industry standard⁴ solution and to the fact that it can be run in parallel with the optical CV recordings. Using Vicon, twelve infra-red cameras were positioned around and above the playing area to capture the ball's motion frame by frame (see Figure 3-7). Frame data from Vicon is synchronised with the Photron Fastcam

⁴ Although Vicon is an industry standard in its approved use, modifying a table tennis ball to become, in effect, a marker is non-standard use. For the work done here the *absolute* accuracy of Vicon when used in this way may also need to be qualified. It should be noted, however, that the key deliverable is a *relative* measurement from CV to Vicon and for this purpose, modifying the Vicon process is satisfactory.

Ultima APX recorder, enabling two real-time, synchronous data feeds to be created, one containing the image data, the other containing independent location data in 3D.

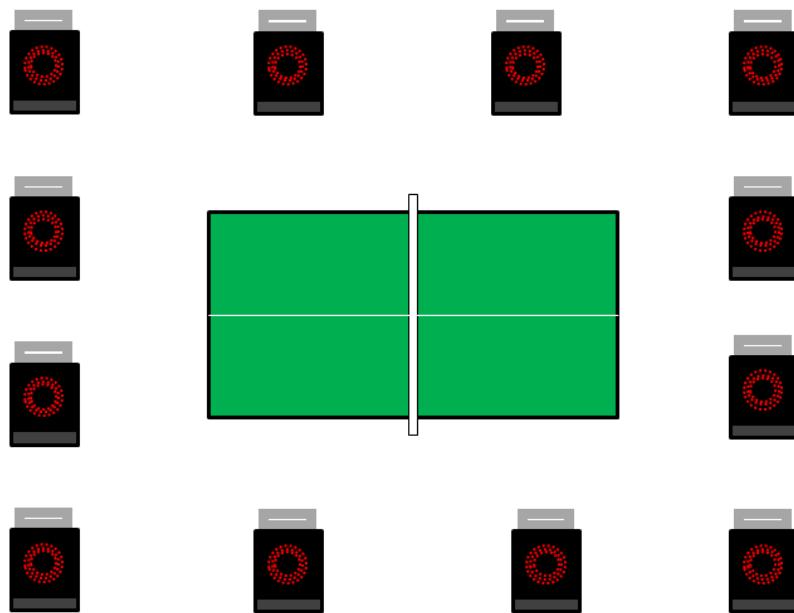


Figure 3-7 Vicon configuration

Attempts were made to use unsynchronised cameras in pilot studies. However manually aligning each video frame by frame is time consuming and prone to human error. Electronic synchronisation was therefore essential as the primary solution to ensure the two sets of data from each video source were temporarily aligned. For electronic synchronisation the two systems, Vicon and Photron Fastcam Ultima APX, are synchronised through a signal generator lock to ensure coincidence of the signals and the output of the signal is confirmed with an oscilloscope reading. Additional verification was made through manual frame alignment spot checks.

In experiments to validate CV based processing for motion change detection, a necessary modification to the ball's surface was required such that it was recognised as an infra-red marker while being still clearly visible to optical video recorders. This choice of material to cover the ball was based on offering the greatest reflectivity, whilst delivering least impact on the ball's dynamics and playability. A modification using Scotchlite™ reflective tape (3M Industrial Adhesives and Tapes Division, 2003), carefully wrapped around the outside of the ball, was made to minimise changes in mass or aerodynamic properties.



Figure 3-8 Modified table tennis balls as Vicon markers showing two types of material

After modification, there were three factors to understand:

- a. The diameter of the test ball increased by approximately 1% compared to the original. The uncoated test ball had a diameter of 39.5mm; once the ball was coated the diameter increased to 40 mm (all measurements made with electronic callipers). As this is within the official ball diameter measurement, the increase was negligible for possible scale factor considerations.
- b. The modified table tennis ball/marker is approximately four times the size of a standard Vicon marker. The marker diameter is a Vicon parameter which is defined prior to recording.
- c. It is accepted that the dynamics of the ball must change due to increased weight and increased aerodynamic resistance.

Considering (a) and (b), detecting a 40mm marker is mitigated through the initial Vicon configuration. Standard 10mm diameter Vicon markers placed at each corner of the table generate datum for Vicon to measure against. Vicon determines the location of a marker as the centre of the marker sphere; the plane therefore is calculated at 5mm above the table surface. This plane is then adjusted for the location of the centre of the table tennis ball, which when placed directly on the table, is a further 15mm above this plane). The issue of impacted dynamics performance of the ball (c) due to

extra mass and airflow restrictions is mitigated by only considering the *relative* motion of the ball. The ball's location is measured relative to the location of the ball detected in a given initial frame, both by Vicon and by the CV algorithm. This is not an experiment to evaluating the performance of an athlete, or the quality of a racket or table, or indeed of the ball itself (ball modification would not then be a possibility). It is noted that player feedback using the ball was positive, with the athlete quickly adapting during general play, servicing and event re-creation. The ball was successfully used with the servicing machine without adverse effects on the ball's delivery.

3.6. Schedule

Each experiment is designed and grouped into a series of experiments. For referencing purposes, each experiment is numbered as *SeriesID-CameraID-SeqID* with a *SeriesID* prefix of 100 for the first series, 200 for the second series, and so on. There are eight series in total. A summary of the schedule is presented in Table 3-3.

Table 3-3 Summary of schedule

Series	Primary experiments designed for specific events and scenarios	Number of recordings
100	Pilot 1; scope assessment	18
200	Pilot 2; location, serving machine (for speed and enhanced curved trajectories) and webcam review. Evaluation of CV comparison software.	18
300	Table bounce, net, ball location and motion change detection; apply full range of recording hardware alongside Vicon.	33
400	Ball speed, acceleration and trajectory analysis	19
500	Detection of net, let, bounce, detection and preliminary analysis of spin	24
600	Increased image complexity for let and bounce	31
700	Full match-play, increased image complexity and doubles line analysis	40
800	Service, return, table bounce, doubles line, let and calibration	82

Prior to each subsequent series, video sequences are subjected to CV processing to assess and analyse for additional requirements. An initial confirmation of the suitability of the video is made and

the relevant data about the video quality (i.e. FPS and resolution) are stored. Where appropriate, the video sequences are edited in length to isolate a particular motion, event or shot. The quality, colour, contrast, brightness and other core image attributes are unaltered. Additionally, access has been given to recorded footage used in a preliminary project within Loughborough University, where the hardware and environment parameters were not available; this is identifiable by the prefix 9999.

In addition to the usable video sequences above, each session was recorded by a rolling video camera for transcribing purposes, plus a number of static photographs were taken of each environment, hardware setting and general configuration. A summary of the key outputs of each series is provided below.

a. Pilot study 1 (Series 100)

The initial pilot study to record table tennis footage is to establish three aspects: (1) overall feasibility of recording and CV processing, (2) a reduction in scope of hardware, configuration, positioning and general environment options, and (3) develop the method of documentation to ensure repeatability and comparison analysis. The pilot studies themselves, however, quickly produced data outputs with features adequate for use throughout the research.

b. Pilot study 2 (Series 200)

A second pilot study, introduces very low-cost technology with the Philips ToUcam webcam, to experiment with different camera locations and to use the high speed and spin capabilities of the serving machine. This series is extending indications from Series 100 and testing early development of the TRASE software. However, its primary aim is to increase baseline data, with broader environmental factors and variations in the ball's trajectory. The results of the two pilot sessions determine the final requirements of the TRASE software

c. Ball motion change detection (Series 300)

This session expanded on the library of match play events recordings using the full range of hardware. The proposed position of the camera was introduced and used for all future recordings. The pilot studies discovered a variation in ball image location depending on the CV algorithm chosen. Series 300 investigates this finding to establish, if this is a consequence of camera configurations, lighting variations, or indeed as a direct results of the CV processing. As a variance in ball location

can be incorrectly identified as a motion deviation, the accuracy of a CV algorithm when detecting the ball must be understood. This session introduced Vicon as an unobtrusive, external comparator running in parallel to CV processing and the standard configuration for Vicon was established during this series. Highly reflective, 10mm, markers were placed at each of the four corners of the table, at the centre of the table below the net, at two ends of a fixed length measure placed on the table and finally placed at the exit of the serving machine.



Figure 3-9 Location of markers around the table

A photograph of the table taken with an LED flash highlights the locations of the markers at the corners of the table and the centre point just below the net (Figure 3-9). Also visible is a fixed length measure placed along the centre line with a marker placed at each end. This remained on the table during all Vicon recordings to provide a standard scale within the images as required. In this picture the ball is the largest reflective marker, to centre left. The service machine is in the centre foreground, also treated with the reflective material so that the exit point of the service could be referenced.

d. Ball Location, speed and acceleration (series 400)

Supplementary event sequences were added to the library and further parallel Vicon data was recorded. The use of a Foresight™ launch monitor (Foresight Sports, 2013) was incorporated in parallel to both Vicon and CV detection processing (see Figure 3-10a). Although designed for measuring golf ball dynamics during and immediately after golf swings, early tests indicated that it may also be triggered by a table tennis ball if the ball's trajectory passes across the line of sight of the device appro-

priately. For this, the table tennis ball was placed on a raised platform (representing a golfer's tee) and a firm shot is made from this stationary position across the front of the launch sensor (see Figure 3-10b). This experiment provided additional supportive data for velocity and acceleration measurements.

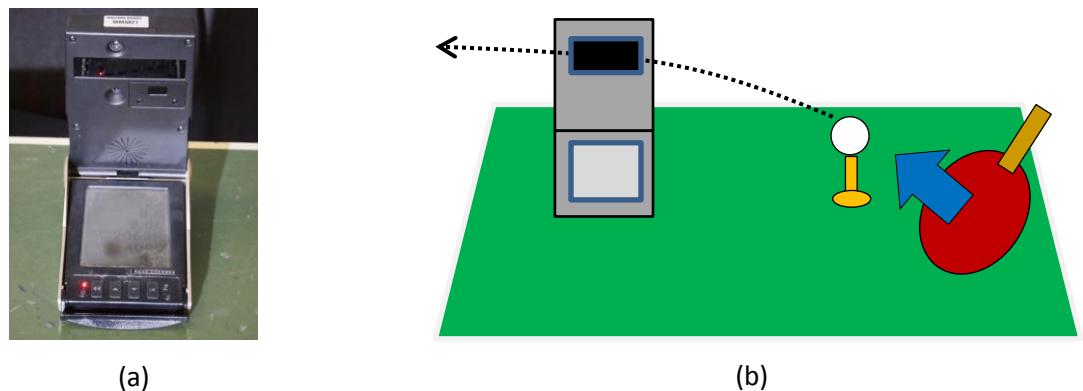


Figure 3-10 Using Foresight™ Launch Monitor to measure speed

e. Events and spin (series 500)

Further recordings of the bounce off the table, net and let calls, net collision, doubles line events and the effects of spin were the focus for experiment series 500. Black contrasting annotations were made on an orange ball, depicting lines of axis separated by a rotation angle of 30° (Figure 3-11).

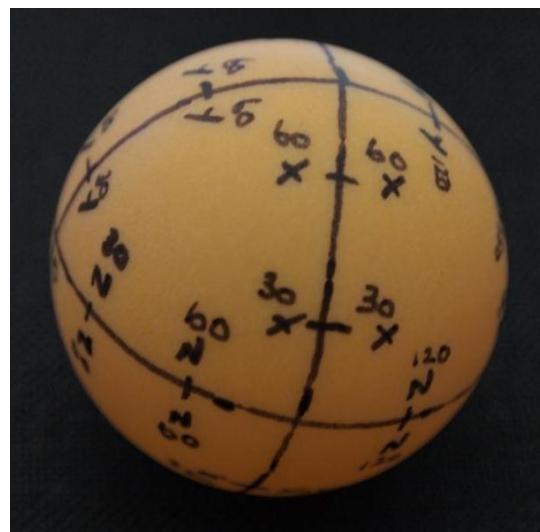


Figure 3-11 Ball annotated with spin markers

The ball was served using the Butterfly serving machine with varying amounts of spin and speed. A range of shutter speeds and frame rates were used in recording the events.

f. Increased image complexity and chalk measurements (series 600)

As a further validation of the Vicon data, introduced for this session is an additional measure of distance using chalk on the surface of the table to mark where the ball lands, which is then measured with tape from the tip of the serving machine from where the ball is released. Furthermore, this series addressed depth of field data requirements using diagonal cross-table ball shots, lets, nets and services. Finally, the Olympus E-PL1, was introduced to generate comparison data using its larger camera sensor whilst running at a lower frame rate (30 Hz).

g. Full match-play, speed and doubles line (series 700)

With the improved understanding made in software analysis requirements from previous experiments, isolated events have been defined and re-recorded. Along with this were recordings of full match-play sequences, from service to point win, with the cameras placed in the optimum location covering the entire FoV of the play area. The expected outcome of this series was to successfully detect and extract the ball at this range, with the complexities of moving players at either side of the table, a complex background and inconsistent lighting. This is designed to provide the most realistic environment in testing the hypothesis of this research. Additional data was collected for further understanding of doubles line detection.

h. Service and calibration (series 800)

The final session completes the data collection phase by re-examining the service and also to generate a matrix of ball sizes across key areas throughout the playing area, to understand the effect of lens aberration and depth of field further. The service recordings were designed to create outputs for evaluation of likely ball occlusions, speed measurements and to also provide source data to detect the legality of the service. In creating a ball diameter map across the table, a uniquely designed rig (Figure 3-12a) was built to enable the ball to be placed in known 3D locations around the table. It is known that the ball's location relative to the camera has a possible effect on the ball's image diameter within the frame. A review of spherical aberration (Cyganek & Siebert, 2009) indicates curvature of the lens and distance of the ball from the camera will impact the size of image the ball. Experiments in this research measured this effect directly, to understand the potential effect on motion

and scale of reference calculations. Whenever monocular devices are used for CV analysis, an appreciation of this effect must be included in any assessment of speed, acceleration and distance measurements across the entire frame. The rig positions the ball at any one of four heights above the table, by alignment of four holes (one at each level) above the intended position (Figure 3-12b). By placing balls onto the rig, recording an image and processing this image in the same way as a video sequence, the size of the ball from fixed locations around and above the table can be calculated.

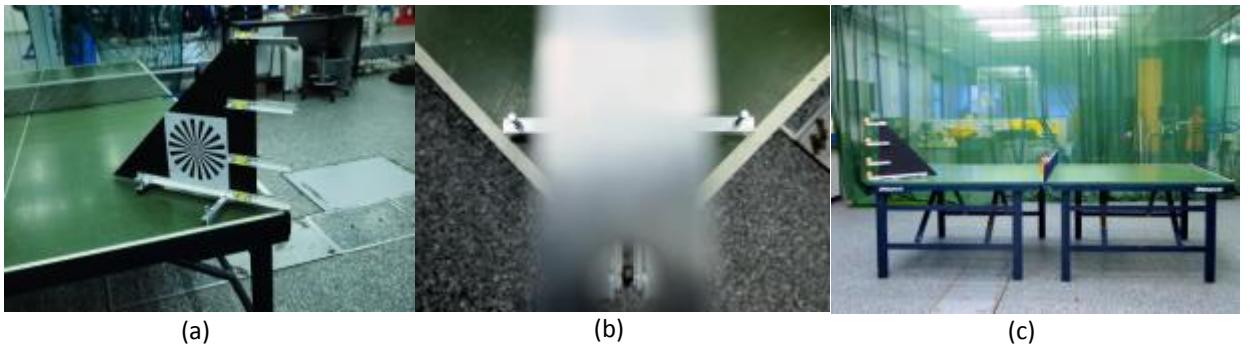


Figure 3-12 Calibration rig in situ

Each of the four different heights in the rig is equal to integer multiples of the height of the net, i.e. 0x15 cm, 1x15 cm, 2x15 cm and 3x15 cm (Figure 3-12c) to provide useful data as the ball moves through important locations during play. The locations chosen were four heights above all four corners, four heights above each end of the length of the net, and four heights at three centres along the lengths of the doubles line, giving thirty-six locations on and above the table in total in a zone representing the volume within which *the majority*⁵ of table tennis events of interest occur (see Figure 3-13).

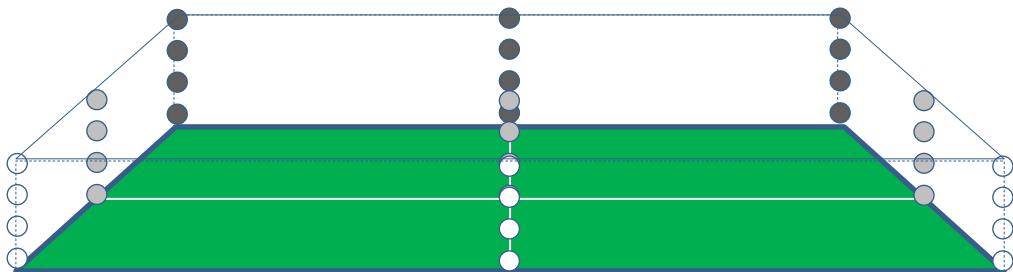


Figure 3-13 Calibration rig measurement zone

⁵ All services and most returns occur outside of this zone. However services generally occur very close to the left of right edge of the zone; additionally and it may be assumed that determination of a return is more straightforward to detect than that, say, of a doubles line or net call. Ultimate design of a solution may require separate, non-linked dedicated cameras installed for each player.

3.7. Summary

When establishing the applicability of CV algorithms to event detection in table tennis, a database of source videos is required for which the recorder settings and environmental conditions are known and controllable. The use of online resources is not ideal, as these videos have been down sampled for on-line accessibility, the camera placement has not been specified or optimised and often the details of cameras configurations and settings are not provided. The video database requires a combination of simple (for proof of concept) single event based videos and complex (for full analysis) videos containing complex backgrounds and full match play. The players themselves should be of sufficiently high level for the intended application where a serving machine for serving is useful when repeating the same shot at high speed and spin rate.

When assessing low-cost monocular video constraints, a range of recording hardware is necessary. Down-sampling of high quality source videos post-recording may not give a true indication of the limits of CV processing. Consideration has been given to the use of strobe lighting and the recreation of events using computer graphics. However, this will not necessarily prove the application when using low-cost hardware within a real match play environment.

The design of the experiments must not impact the game. The players must be free to move and the apparatus should conform to ITTF guidelines. An exception to this has been made during these experiments for the confirmation of event detection from a non-CV based implementation. This research has found the modification of a ball to be detected by optical and infra-red wavelengths allows parallel recordings using both CV and infra-red marker based detections.

The colour, size, shape and speed of the ball are initial considerations when designing an environment within which a typical table tennis match will be played. When managing the hardware, there are seven factors to consider: (1) focus, (2) shutter speed, (3) frame rate, (4) lens speed, (5) focal length, (6) ISO speed and (7) sensor size. Each of these factors cannot be considered in isolation as each has a direct impact on each other and the ultimate usability of the image for ball detection. The speed of the camera's shutter must be as close to 1/420th second for a professional match, but can be considerably less (between 1/100th and 1/200th second) for amateur event detection. Anything less than this and the ball will be too blurred and key event data (such as a let) could be missed. Similarly, the highest possible frame rate should be used, to ensure the greatest number of data points are collected as possible. Frame selection, where alternative frames in a sequence are

used for example, is considered at the CV analysis stage if processing power is limited. However, with a frame rate of less than 100 Hz, the reduced number of data points increases the potential for non-identification.

The scene should be as evenly illuminated as possible. At shutter speeds below 1/250th second the lighting provided by the installed room lights is sufficient to capture the ball clearly. Care should be taken with certain lighting installations which oscillate with the frequency of the utility power supply, such as gas based lighting (fluorescent tubes etc.) as these can cause fluctuations in exposure. This leads to false movements of objects between the frames due to the brightening and dimming of lighting altering the detected sizes of objects. If a low-cost solution is implemented in halls and arenas with fluorescent lighting, then it is a proposal of this research to incorporate additional stable lighting sources such as LEDs, if required.

Lens speed may not be critical and a slow lens may be compensated for by good lighting and a fast shutter speed. ISO (or gain) should be set as low as possible, ideally 400 ISO to reduce potential noise in the image. Consideration of the lens focal length is important to ensure the entire playing area is visible, avoiding extreme wide angle lens which create significant lens distortion, while not having too short a focal length that the camera has to be placed far away from the scene. A 35mm photographic film equivalent range of 28-32mm was suitable to place the camera close enough to the table (2-3m) without significant visual distortion. Placing the camera at this location, in line with the net and having the lens positioned 15cm higher than the net results in the optimum recording aspect with minimisation of ball occlusions. A horizontally level FoV can be achieved using a simple bubble level placed on the camera. However, with a monocular device, it is not possible to avoid all occlusions during typical match play and this must be appreciated by any final single source CV based solution.

4. Computer vision processing

4.1. Introduction

Following the substantial amount of image data generated in Chapter 3¹, a method of processing the data using Computer Vision (CV) algorithms is required to detect the table tennis ball and its motion. Additionally, in order to identify the most suitable CV algorithm for event classification, player performance review and automatic scoring in a typical table tennis match, a comparison of the CV algorithms is also necessary. Combining these two requirements into a software program provides the environment to accurately (and repeatedly) process the videos with consistency, automatically generating outputs for use as comparison data. Such software needs to allow for varying combinations of pre and post processing modules and their parameters and to additionally allow non-CV data generated by external systems to be compared. This chapter describes the necessary design requirements and architecture, with a proposed solution for a table tennis CV comparison workbench for use in this research.

4.2. Software design

The ability to process video sequences in a standard, repeatable environment is important for robust experimental analysis. The care taken in designing experiments for generating quality source data must be carried forward to the design and implementation stages for user interface based software tool. Assumptions based on data processing outputs must ensure that the data processing itself has not introduced variations and so must be consistent in its use. As such, the following high level requirements were identified as necessary for the generic design of monocular video sequence processing and comparison software:

- a) All pre and post processing functionality is modular. This enables consistency and integrity through re-use and provides for different combinations of functionality. Options for grey-tone processing, background subtraction and thresholding techniques are included. Current literature indicates 2-pass thresholding is successful and therefore is included.

¹ Each sequence generated an average of 491 frames each, and with 276 sequences this created over 130,000 images to analyse. Although not all sequences are processed by each selected algorithm, this is still a substantial amount of data to process.

- b) All CV based ball detection algorithms are made modular to enable consistency re-use and extensibility. Based on current literature, the software must be designed, at a minimum, to evaluate common edge detection algorithms. Additionally, in furthering current knowledge and improving comparison data, it is proposed that foreground extraction algorithms should also been included as modules.
- c) Provide options to save and load configurations for reducing human error in configuration setting during repetitive processing.
- d) Provision for raw data to be exported for analysis in external software, as required. The output data incudes processed images, extracted objects of interest (OOIs) and motion data.
- e) Externally generated, non-CV based ball position data to be imported, aligned and compared to equivalent CV generated data.

No off-the-shelf solution with the above functionality was found. Therefore, as part of this research a novel software tool called the Table tennis Recording Analysis Software Environment (TRASE) has been developed in Matlab R2012a (The MathWorks Inc., n.d.-b) with Computer Vision Toolbox version 5 (The MathWorks Inc., n.d.-a). Matlab was chosen for its well documented, supported and community evaluated library of CV functions. Other software development environments have also been considered, including OpenCV for C++ (Itseez, 2015) and SimpleCV for Python (Sight Machine Inc, n.d.). However, in the case of the former, it is not an ideal prototyping environment whereas the latter was rejected due to reduced capability and low popularity within the CV community. Yet these tools should not be ignored for future research. Indeed, if processing power, direct access to the GPU and overall speed of decision making are relevant criterion then it would be a suggestion to investigate the suitability of OpenCV with C++ for this purpose. In the case of this research, where the crucial investigation outputs are to review a breadth of CV algorithms against a selection of key criteria, speed of execution is not important. However, a measurement of relative processing performance is built into TRASE for further research capabilities.

When measuring the relative merits of CV algorithms the critical CV outputs and metadata recording processing outcomes, it is suggested that datasets are generated automatically by the comparison software. With TRASE, the raw output database includes:

- a. The recording of primary ball metrics, such as position within the image, along with ball size and shape properties.

- b. Ball dynamics data recording velocity, acceleration and bearing to assist with event classification and player performance analysis.
- c. Recording the automatic detection of abrupt change in motion across frames in a given video sequence. In parallel, generating automatic records of the calculate differentials between CV motion detection and non-CV event detection.
- d. The creation of the event state sequence, automated for umpiring assistance and scoring functionality.
- e. Processing performance (mean time taken to detect the ball for all frames in a given video sequence).

The TRASE graphical user interface is divided into a number of panels from left to right, broadly reflecting the flow of CV processing for a given video sequence, from standard input configuration parameters, through CV detection algorithm, filtering and final outputs. For an overview of the user interface of TRASE, see diagram Figure 4-1.

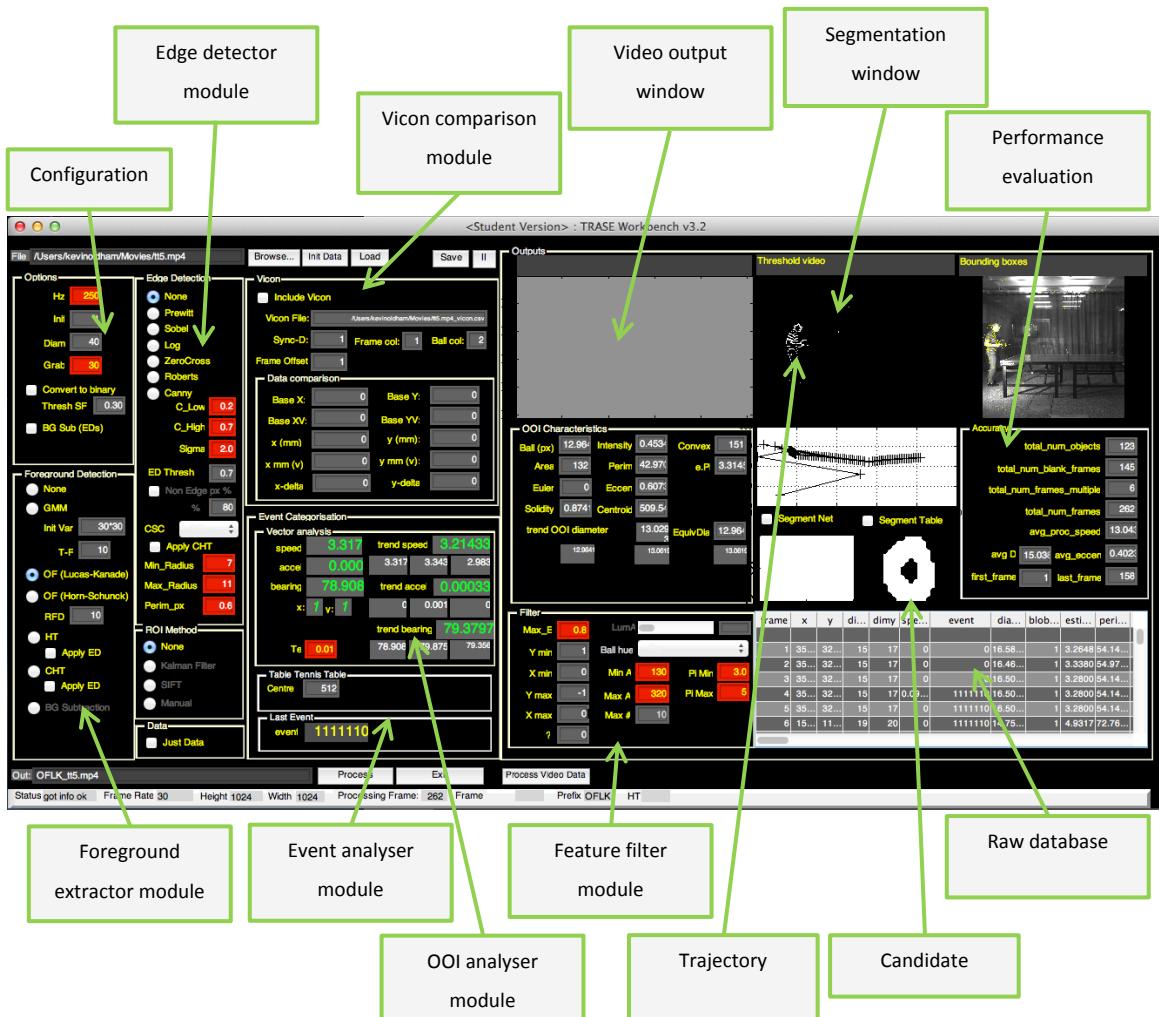


Figure 4-1 TRASE User Interface

Due to its modular architecture, TRASE is designed to allow any video sequence to be processed by any algorithm. Modules may be appended to either the pre-processing, processing or post processing stages. Support for video sequence formats include files compressed using the standard MPEG-4 codec (The Moving Picture Experts Group, 2014), in MOV or MP4 containers or the Microsoft AVI format². This allows for the majority of native video source formats.

The software delivers CV analysis data through two primary stages: 1) comparator analysis and 2) event detection. The first stage aims to select the most suitable CV based algorithm from a defined selection by comparing them against each other, measuring key indicators and identifying the optimum result. The second stage implements this selected algorithm to the event classification and

² Depending on locally installed codecs.

measurement processing. In the first stage of direct CV algorithm comparison, there are six standard modules used in processing for each video sequence. These are referred to as: (a) pre-processing, (b) apply detection algorithm, (c) ball identification, (d) identification metrics, (e) database update and (f) full sequence comparator analysis. This comparison process is represented by Figure 4-2.

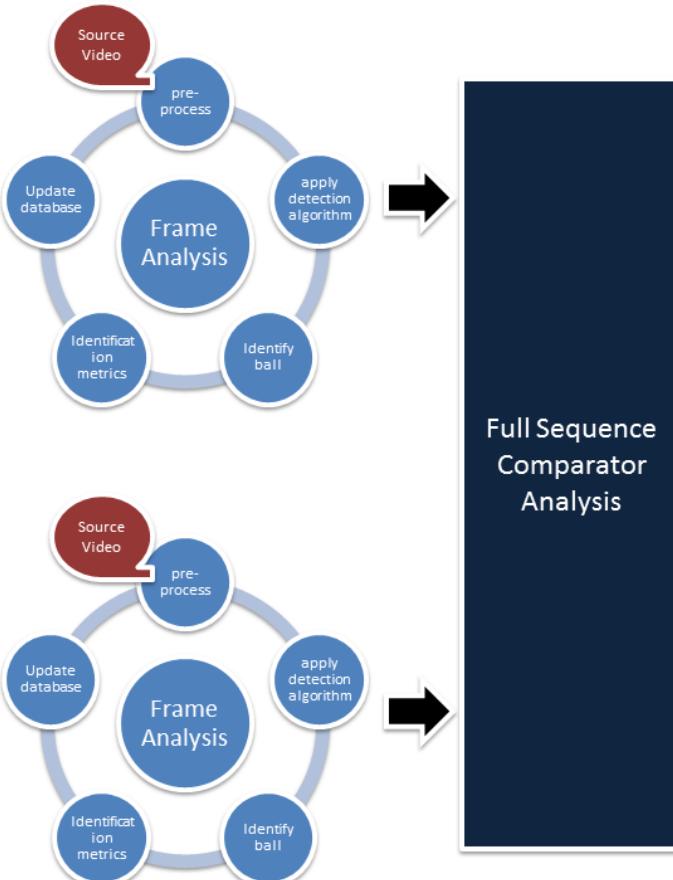


Figure 4-2 Comparison Processing modules

The same source video is processed by at least two algorithms. Each frame in the sequence is processed according the CV detection algorithm being investigated. Resulting outputs for the success (or otherwise) of detection are recorded, along with metadata describing the shape, size, pixel location, and processing speed in the database. This raw data can ultimately be used for event detection and performance measurements in match play. Snapshots of the frame and segmented ball are also saved together with an output video sequencing the segmentation. A complete set of data for each sequence analysis is then stored in a CSV file where direct, frame by frame, comparisons are made. These results are stored in the database for detailed review and further analysis. From this data, a justification for the selection of an optimum CV algorithm for use in table tennis can be made.

Once the optimum CV algorithm has been identified, it is then re-used during the automated event categorisation and measurement investigation. TRASE re-uses coded modules where required. The software process for event analysis is represented in Figure 4-3.

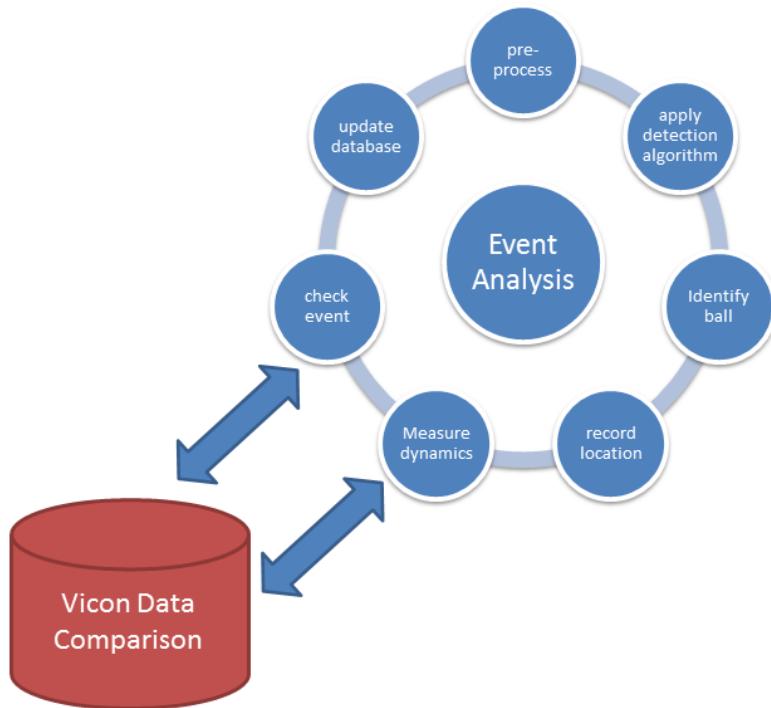


Figure 4-3 Event Analysis

4.3. Workflow

To ensure parity and efficiency in processing each video sequence, a standard approach to the workflow is followed. This includes activities from initial source video review (assessing for usability and applicability) through to final data checks. In total there are six key workflow stages summarised as follows:

1. Video source review

The initial stage involves visually checking the video to ensure it is usable and editing the sequence, if required, to isolate a specific event. Any editing or re-formatting must not alter the original quality (maintaining frame rate, resolution and colour space). The source videos can be either .avi, .mov, or.mp4, depending on the recording equipment's native format. The source format is checked using

Prism Video File Converter v1.88 (NCH Software, 2013). The frame aspect ratio is maintained at 1:1 and to ensure parity in processing speed evaluation. Without frame loss, the playback frame rate for all videos is 30 Hz (thereby creating a slow motion effect if watched directly).

The eight recording sessions, plus suitable videos created during the pilot, generated 276 usable videos, each of which could potentially be processed. At the end of each session, the videos are analysed for suitability of use, catalogued and events manually classified for easy identification and retrieval. The videos were retrospectively grouped into algorithm comparison, motion change validation, dynamics and event analysis for use by TRASE. Any individual video may be active in more than one group. The event group is subcategorised into separate events for ease and consistency of retrieval. All video metadata, such as number of frames, resolution and duration, is confirmed post-recording using iMediaHUD (The Iffmpeg Team, 2013).

2. Pre-processing

Prior to the frames being processed for ball detection, each video sequence is passed through a number of optional pre-processing stages in TRASE for the calibration of the binary threshold mask, background subtraction key frame (for edge detectors only) and frame conversion from RGB to HSV (The MathWorks, n.d.) and to grey-tone. Additionally, replacing standard red, green and blue data for each pixel with one of hue, saturation and value (intensity) is more convenient in the manipulation of the image data. This is due to the fact that the RGB values are all correlated to the lighting around an object and equal lighting around neighbouring objects, making distinction and extraction of objects more involved.

3. Ball detection algorithm

Once the video sequence has been pre-processed, the algorithm under investigation is selected and TRASE steps through each frame to detect the ball. Modules developed for this research allow for either one of three foreground extraction algorithms or one of seven edge detection algorithms, but other algorithms could be applied through the modular capabilities. Lead-in frames selected during pre-processing, ideally not containing players or the ball are used for the removal of static background image data, as required. It is an expectation of this research that no initial manual intervention is required to identify the first instance of the ball. This is a key evaluation parameter of the CV detection process.

4. Post Processing: CHT and the feature filter

During the pilot experiments, it would be the case that many moving candidate objects are detected from each frame. Therefore, a module is required to allow individual frame output to be filtered for similar shapes, eliminating as many candidate balls as possible. If CV algorithms are to be successfully used in table tennis using a single video source, then a method of reducing the number of false positives is necessary; for this research the process is either a Circular Hough Transform (CHT) (Duda & Hart, 1973) or a bespoke feature filter algorithm³. The advantages of a CHT are that it is not greatly affected by image noise or gaps along the perimeter (H. Ballard & Brown, 1982). However, by simply examining the perimeter of the segmented region other table tennis ball characteristics are disregarded which could offer a greater detection success. TRASE combines these characteristics in the bespoke feature filter for direct comparison with the CHT. Extending the work in evaluating table tennis balls from video sequences by Wong (P. Wong, 2009), the following filter features (Oldham, Chung, Edirisinghe, & Halkon, 2015) have been designed and incorporated into TRASE in refining the ball detection process, reducing the number of false positives.

- a. **Shape.** The key features in determining the shape of the ball are described in Table 4-1(this feature set can apply to any sport with a ball as the primary object of interest).

Table 4-1 Shape features

Feature	Definition
Diameter (D)	Calculated using the bounding box method (Weisstein [1] n.d.) as the mean of height and width of the smallest box around the region, or the centroid calculation (the centre of mass);
Intensity (I)	The colour intensity of the region. The intensity is used to filter reflections and shadows.
Perimeter (P)	The total distance around the boundary of the region, measured in pixels.
Eccentricity (E)	The ratio of the distance between the foci of the ellipse and its major axis length. A value between 0 and 1.
Area (A)	The number of pixels in the segmented region.
Solidity (S)	The proportion of pixels in the convex hull which are also in the region
Equivalent diameter (D_e)	Given the area (A) of the region, this is the diameter of a circle with the same area.
Convex Area (A_c)	The number of pixels bounded by the convex hull.
Perimeter to diameter ratio	The ratio of perimeter (P) to diameter (D); this should be approaching π .

³ Again, the modular approach of TRASE allows other techniques to be incorporated.

Given measurements of ball width (diameter) and perimeter, the other features can be estimated and the estimated values compared with the measured values to provide verification of the shape of the segmented object. The length of the ball's perimeter is determined, and measured in relation to its area. An estimation of the equivalent diameter can be made and tested against the results using the following equation (4-1):

$$\text{ball width} \cong \frac{P}{\pi} \quad (4-1)$$

The area contained by the segmented ball candidate is calculated and related to the length of the perimeter to provide further confidence on the ball's overall shape. The area is calculated as:

$$A \cong \left(\frac{\text{ball width}}{2} \right)^2 \times \pi \quad (4-2)$$

Eccentricity (e) should be close to zero for a perfect circle (Weisstein [1] n.d.). The shape of the ball may not be circular and early pilot studies are detecting a value consistently $e > 0.5$. The eccentricity is greatest along the direction of travel. This is especially the case when using slow shutter speeds. However, the eccentricity remains relatively constant throughout all images and can be expected to be so along the axis of travel. For detailed analysis, it is recorded that eccentricity decreases slightly as the ball's speed decreases after being hit by the racket. Within the design of TRASE, this parameter is tuneable.

The ratio of perimeter to area is calculated to reduce false positives from non-curved objects, such as a star. The area can be derived by a pixel count of the segmented object. A verification of this area can be made by applying equation (4-3).

$$\text{ball area} \cong \left(\frac{\text{ball width}}{2} \right)^2 \times \pi \quad (4-3)$$

Furthermore, an approximation of π is derived by dividing the perimeter of the segmented region by its mean width (bounding box method). This again gives a strong indication and validation of its circularity when used in conjunction of the eccentricity calculation. As an initial guide,

the candidate object array is filtered by examining the ratio of a candidate's area to its bounding box perimeter length. When the ratio between the area of the object and the area of the bounding box is above 0.4, it is classified as a feature filter *pass*; any other candidate object is discarded. Detailed experiments conducted within the context of this research reveal that for high quality images, with adequate lighting and fast shutter speeds, the ideal ratio of area to bounding box perimeter length is $\pi/4$; if the ball and bounding box is a perfect circle and square respectively the ratio would be 0.785. However, for many shapes simply measuring a bounding box perimeter and comparing to the object's area will not fully describe the shape contained within. Take, for example, any mono-spaced typeface: if one were to draw a bounding box around each letter, using this method of eccentricity, one may deduce that an asterisk (*) was indeed a circle. This is obviously untrue – what defines an asterisk from a circle in this case is its perimeter. For an object that has the shape of an asterisk, the eccentricity would be similar to a circle but the eccentricity index would be greater than that of a circle. Considering this, a more successful approach to finding a circle from a candidate is to test for square of the perimeter to area ratio as in equation (4-4):

$$\frac{\pi}{4} \cong \frac{P^2}{A} \quad (4-4)$$

The positive impact of this filter is considerable and is therefore placed first in the filtering order. False positives are immediately reduced for minimal resource effort, and some reflection errors are also decreased. However not all false positives can be removed by this filter alone. Parameter ranges for shape filtering is tuneable in TRASE.

- b. **Colour.** The ball's colour is either white or orange (see Section a). However the precise colour values in the video can vary considerably, depending on lighting, camera sensor, and internal colour processing techniques and thus cannot be matched directly. However, pilot studies in this research have shown the ball to be of uniform hue and it is this solidity within the detected region that provides an additional indicator of the correct ball candidate. This is particularly effective in dealing with ball shadows and reflections where the region shape and size are a good fit, however the hue, saturation or intensity are out of range.
- c. **Region of interest (RoI).** Given a ball's location in a frame, and knowing its general speed and the frame rate of the camera, the distance that the ball will have travelled in the time between

frames can be extrapolated. This information is then used to create an area of investigation, in the next frame, of where the ball is expected to be.

5. Outputs

Apart from the Comma Separated Value (CSV) files storing dynamics, features and Vicon differential data, which are created and updated every frame, there are a number of other outputs generated by TRASE. The first of these is an image and location output of any false positive detection (one that is discounted by the filter or CHT process), which provides additional data during the development lifecycle of TRASE throughout this research. Additionally, for all ball detections, two images are stored: the fully processed frame and the ball candidate as defined by the bounding box. A configuration data file is created at the end of processing, recording any configuration options and summary ball detection accuracy results showing, for example, when the ball was first detected, in how many frames the ball was not detected and how many potential ball candidates were found pre and post filtering.

Each region-segmented frame is compiled to produce a single, corresponding video file (.avi format) for human observational analysis and comparison with the original video sequence. This provides a visual record of the areas detected by the algorithms and aids further improvement. There is the option to save a single frame (both original and process) through the configuration options within TRASE and it is stored as a JPEG image. Finally, a screenshot of the TRASE window is made to provide evidence of all settings, configuration options, and accuracy settings and file naming etc. All of these data output files are stored in a unique directory for each algorithm processed video sequence.

Each filtered candidate has its pixel-based location within the image and all feature properties relating to its size, shape and colour recorded. The pixel location is also converted to a *relative* location from the previous ball detection, which is then converted to millimetres based on the size of the ball as a scale of reference. The distance travelled from the previous location is calculated based on both the Euclidean distance (Wolfram [1], n.d.) and the Manhattan distance (Wolfram [2], n.d.); knowing the frame rate (FPS) and distance, the speed can be determined. Given two sets of points, the bearing is determined (with bearing of 0° where $x = 0$ and $y > 0$) and recorded with the location data, using the method described. Once the speed for two sequential frames is known, the acceleration is determined as the derivative of speed with respect to time. Once there are three sequential frames worth of data, the process of data smoothing using the Butterworth filter (Butterworth, 1930) starts

and continues for the remainder of the video sequence. All dynamic (smoothed and non-smoothed) data are saved and then reviewed outside of TRASE for further analysis.

4.4. TRASE Findings

During the pilot stage of experiment design, a number of algorithms, image artefacts and processing techniques were evaluated to understand their impact, feasibility and potential benefits. These were specifically:

1. The effect of binary thresholding
2. Binary thresholding optimisation
3. Background subtraction
4. Tracking
5. Ball centre calculation
6. Variations in shape characteristics
7. Ball diameter calculation
8. Variations in CV algorithm's positional data
9. Reflections, lighting and shadows
10. Filter ordering

This retrospective appraisal impacted the initial requirements and design of TRASE, where it was discovered tracking and background subtraction were largely irrelevant. Furthermore, thresholding and feature filtering required substantial review to improve efficiency. The following sections detail these findings.

1. The effect of binary thresholding

Applying a binary threshold to the image has been suggested as a key initial stage for CV based ball detection in table tennis by Modi et al. (2005), Wong et al. (2010) and also for other sports, such as tennis ball detection by Yan et al. (2005). With such a common suggestion, thresholding has been made available for pre-processing of images for all CV detection algorithms in TRASE. Observations during piloting demonstrated that, with either the white or yellow hue ball, a visible contrast is ap-

parent with the majority of non-contrived backgrounds⁴. Using this to our advantage, appropriate binary thresholding alone is able to isolate the ball from the background, often with good results. An example from the high quality pilot sequence 100-P-03 (camera setting CS12, see Appendix C) is shown in Figure 4-4, using a threshold of $T = 0.9$ ⁵. When applying the threshold the ball is clearly visible, with only a small number of other segmented objects requiring filtering. Using this processing it may be argued that the ball's approximate location could be found and, for the general requirements of detecting the ball and determining its position within the image, could be considered a successful processing stage.

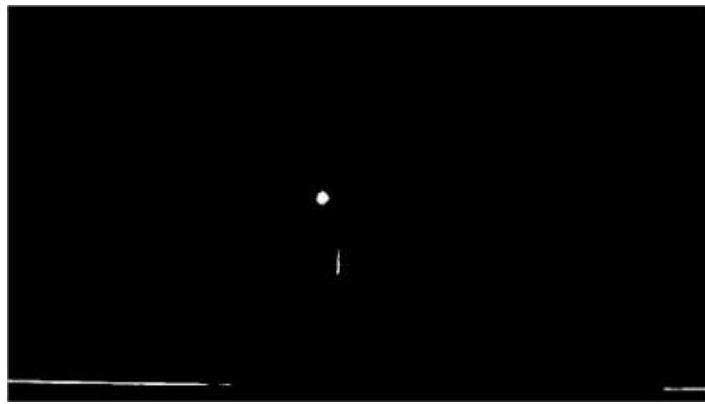


Figure 4-4 The effect of binary thresholding (100-P-03 frame 50 T=0.9)

However on closer review the shape of the OOI is irregularly distorted along its circumference. This side effect is well understood and the importance of accurate thresholding is emphasised in work by Feng et al. (Xu, Li, Feng, Xu, & Chen, 2013). This distortion effect is caused by variations in intensity across the curved edge of the OOI, becoming susceptible to loss during thresholding. This irregularity can vary from frame to frame (see consecutive frame sequences a, b and c from 700-P-15-SEQ5 using camera setting CS13, in Figure 4-5) causing problems when determining the true trajectory of the ball and hence its dynamics, as values for centroid, eccentricity and radius become difficult to determine.

⁴ It would be natural to expect the yellow to have a reduced contrast. However, experiments have shown that the yellow hue 'burns' towards white in the majority of video sequences. This is most likely due to a slight over exposure from the necessary lighting for match play. This makes the contrast variation between yellow and white minimal.

⁵ The background is green netting, with variations in contrast. The ball is the only moving object in the sequence.

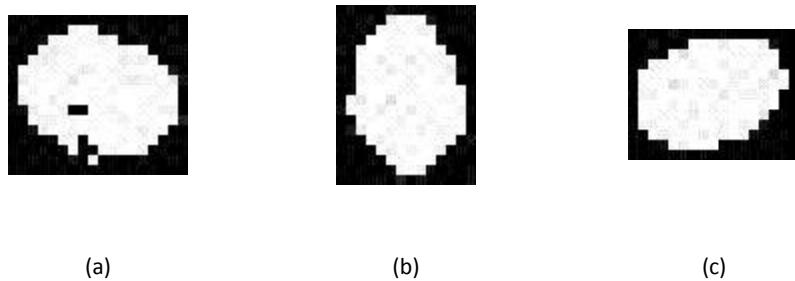


Figure 4-5 Shape variation when thresholding the ball in consecutive frames

Furthermore, in the case of a medium quality video produced by a Canon HG10 (Canon, n.d.) using camera setting CS01, with a complex background and smaller ball, isolating the ball required the threshold value to be reduced to $T=0.5$; in doing so this introduced many unwanted regions (see Figure 4-6).

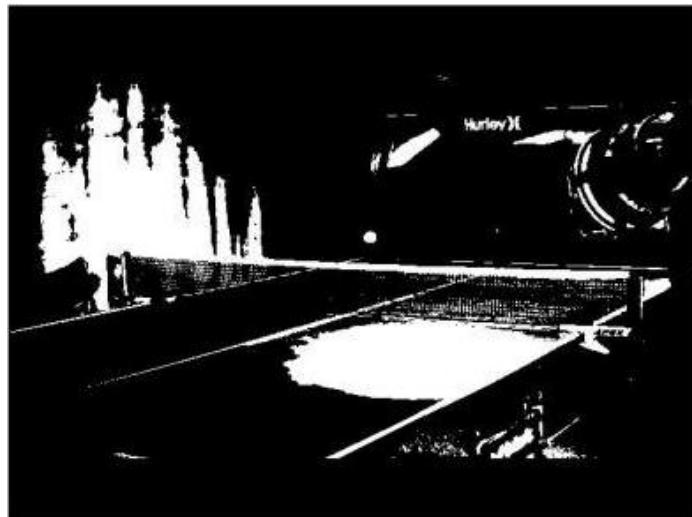


Figure 4-6 Binary 100-C-14 frame 20 $T=0.5$

In this outcome, determining an accurate segmentation of the ball (and hence its shape and location) using binary thresholding alone is not possible as the ball crosses each of these unwanted regions. A tracking solution could be incorporated here, but estimating the tracking of the ball through these regions will still miss key events such as a let or bounce. For comparison, in the lowest quality video available with reduced shutter speeds and a background having a constant but marginal contrast, the ball is detected successfully, using a threshold of $T=0.7$. There are a number of unwanted regions segmented in the image; however isolation of the ball is less likely to be affected due to the improved camera angle (see Figure 4-7).

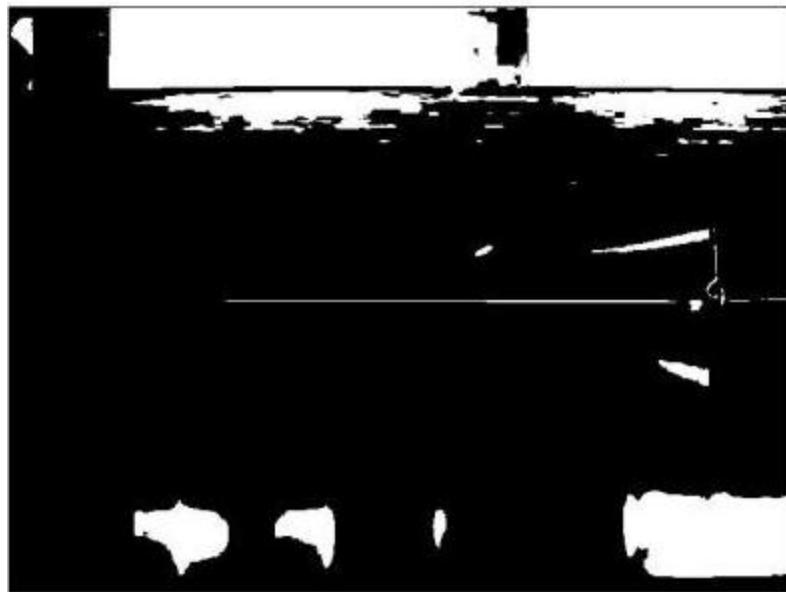


Figure 4-7 Binary 9999-U-01 frame 50 T=0.7

In summary, binary thresholding is capable of detecting the ball, but careful consideration must be given to ensure initially there is a high contrast between the ball and the background. Minimising lighting reflections across the table (as can be seen here in Figure 4-6) is also necessary for simple thresholding. Practical considerations should be given to the former by aligning the table against a dark backdrop; the latter by positioning the camera and/or lighting to avoid reflections. However, even with these practical camera and environment design considerations, the value of the binary threshold varies between different video sequences and must be accurately determined to reduce unwanted regions and to minimise the effect of altering the ball's shape. A consistent and repeatable method for determining the threshold value without using full automation is necessary.

2. Binary thresholding optimisation

Due to the curved surface and shading of the ball at its edge, calculating a threshold value which is able to consistently isolate the *complete* ball from the background for any given video sequence is not achievable. It is important to calculate a threshold value which removes as much of the background data as possible, isolating the ball without losing the darker detail around its circumference. For ball location detection a partial ball isolation, with a uniform processing around its circumference, is to be considered sufficient. However, segmenting the *entire* ball becomes critical when determining the shape and size of the ball if it is to be used as a scale of reference in speed and distance measurements. Consideration has been given to automatic calculation of the threshold value.

However, during these research experiments this has been found to be an unpredictable (and therefore non-repeatable) variable in the design of these experiments and additionally, is not always reliable in isolating the ball with maximum area. Therefore, a partially manual process has been developed in which an estimated threshold value can be determined and then entered into the TRASE GUI for acceptance or finer tuning.

A particular characteristic of table tennis is that, under most conditions, the ball is often one of the brightest objects in the frame. The high intensity of the ball when compared to the general background may be used as a guide to establish the optimal threshold value. Using this knowledge, a threshold range has been determined which is inclusive of the majority of the ball yet exclusive for the background. In detail, the three steps in determining a fixed threshold value for a given image sequence is as follows:

- a. Step 1: Manually isolate the ball from a colour frame with a dark background, defining a bounding box just covering the ball. Convert this sub-image to a grey-tone image and remove any immediate pixels bordering the ball having intensity lower than a specific value. Pixel removal can be done automatically with an intensity threshold of 64 found to be sufficient in isolating the ball in all typical match play video sequences generated for this research. All pixels in this new sub-image (the object of interest) are now of just the ball. Determine the mean (μ_{Oi}) and standard deviation (σ_{Oi}) of the object's intensity.
- b. Step 2: Determine the mean background intensity by selecting a colour frame *not* containing the ball, convert to grey-tone and calculate the mean (μ_{Bi}) and standard deviation (σ_{Bi}) of the background intensity.
- c. Step 3: An estimated binary threshold value range T_{min} to T_{max} is then determined, as in equations (4-5) and (4-6):

$$T_{min} = (\mu_{Bi} + (2 \times \sigma_{Bi})) \quad (4-5)$$

$$T_{max} = (\mu_{Oi} - (2 \times \sigma_{Oi})) \quad (4-6)$$

This provides a range of values within which the optimum threshold will exist. At the upper extreme of this range the threshold removes the darker edges of the ball (a problem caused by directional lighting and the curved surface). Therefore an initial threshold value is chosen close to T_{min} and a

process of threshold iteration is applied until a) the ball is identified and b) the circumference of the ball is smooth in appearance (limited by the resolution). Xu et al. refers to this process as the ‘Traditional iterative threshold’ (Xu et al., 2013). This process is repeated for each video and the threshold configuration stored to allow for standard re-processing of the sequences. An example output from sequence 100-P-06 (camera setting CS08) is provided in Figure 4-8.

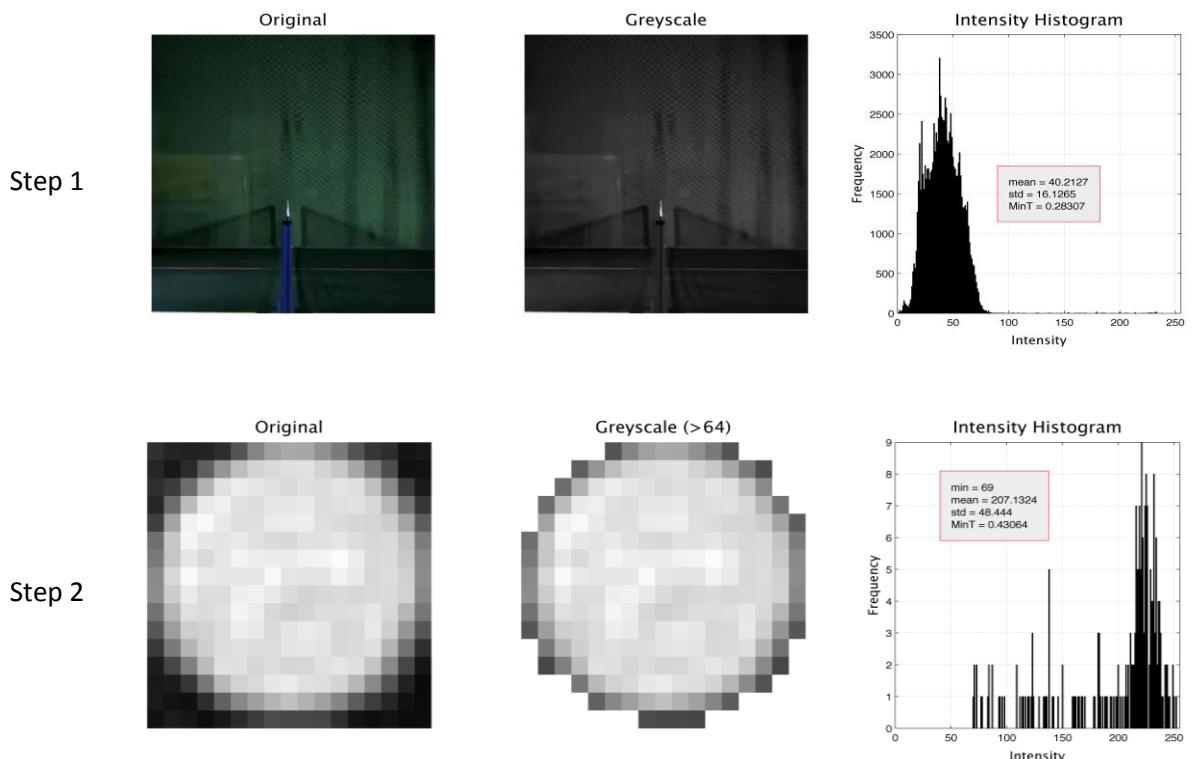


Figure 4-8 Determining the binary threshold

An analysis of the effect of improving contrast has also been made in an attempt to improve the success rate and optimize the delineation of the thresholding. Furthermore this analysis has been made across the RGB, L*a*b* and HSV colour spaces. Sample outputs can be seen in Figure 4-9.

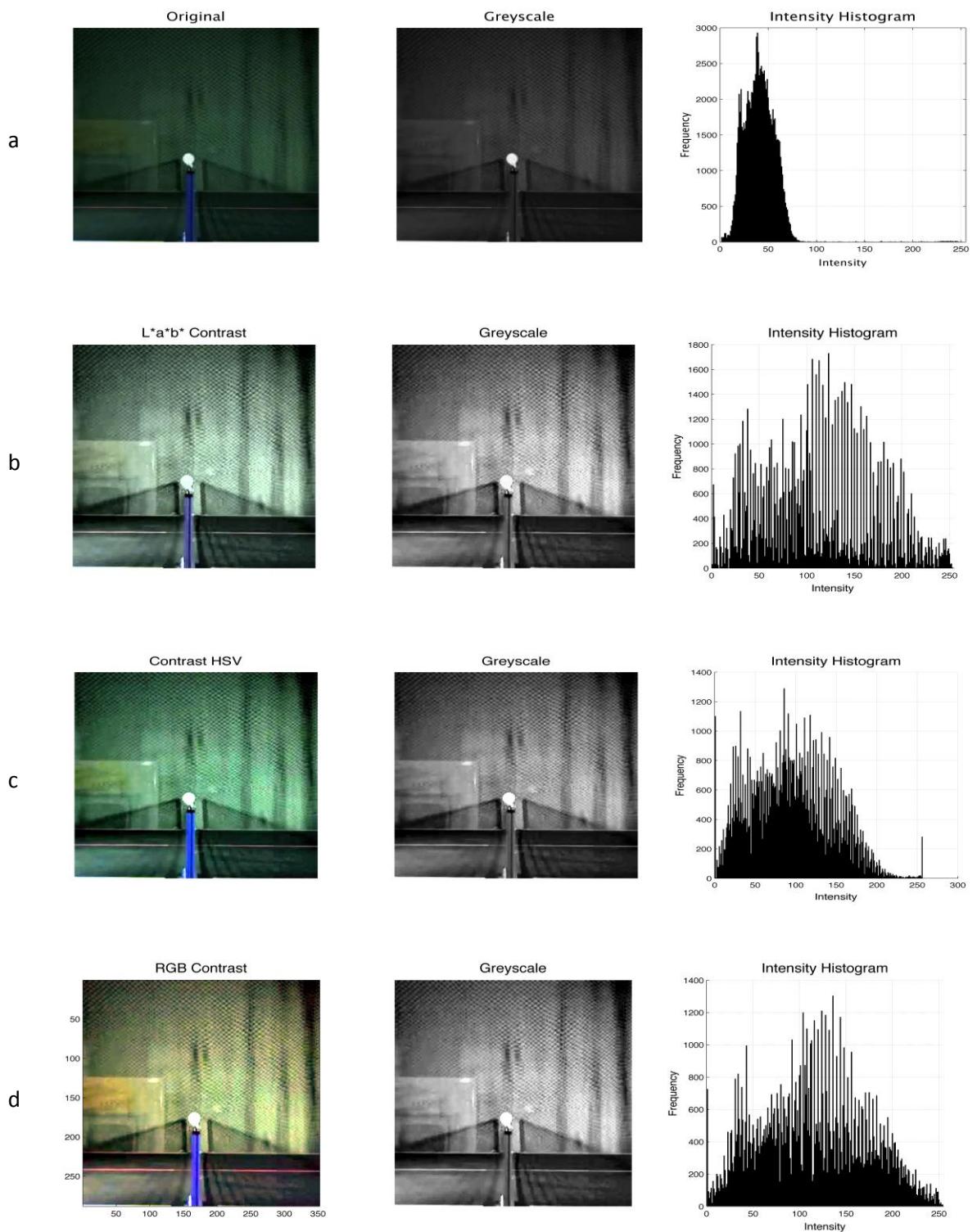


Figure 4-9 The effect of contrast on binary threshold

The results show that contrast adjustments intended to improve thresholding of the ball by widening the intensity difference between the ball and the background instead stretch and spread the overall contrast of the image (see Figure 4-9 b, c and d). In table tennis the bright ball already has a distinc-

tive intensity profile shifted to the right of the intensity graph. Increasing the contrast of the frame moves all other pixels towards this profile. Therefore, it is proposed that no contrast adjustments are made (see Figure 4-9a). This has the additional benefit of reducing the processing overheads. The method of determining the threshold value above is intended as a guide only. It provides a value (T_{min}) from which an initial binary threshold can be tested and is tunable by TRASE for each video sequence. Once chosen this value is *not* modified for each application of the algorithm to the video sequence, ensuring a consistent approach for comparison.

6. Background subtraction

During piloting, all edge detection algorithms were first processed through a frame differencing algorithm to attempt to remove static regions in the images. However, it was discovered that there was only a marginal benefit to segmentation and no improvement on rate of detection, particularly with complex and often moving objects in the background. Therefore, the suggestion is that any potential benefits of background subtraction do not outweigh the added processing overheads when real-time processing becomes critical for live umpiring and scoring. However, it has been observed that there is an application for background subtraction during the let detection event. Results have shown that a slight movement of the net during net collision, or let, is sufficient to be detected by background subtraction using a difference frame with a stationary net. See Figure 4-10 (b), with a zoomed view of the ball hitting the net in Figure 4-10 (a).

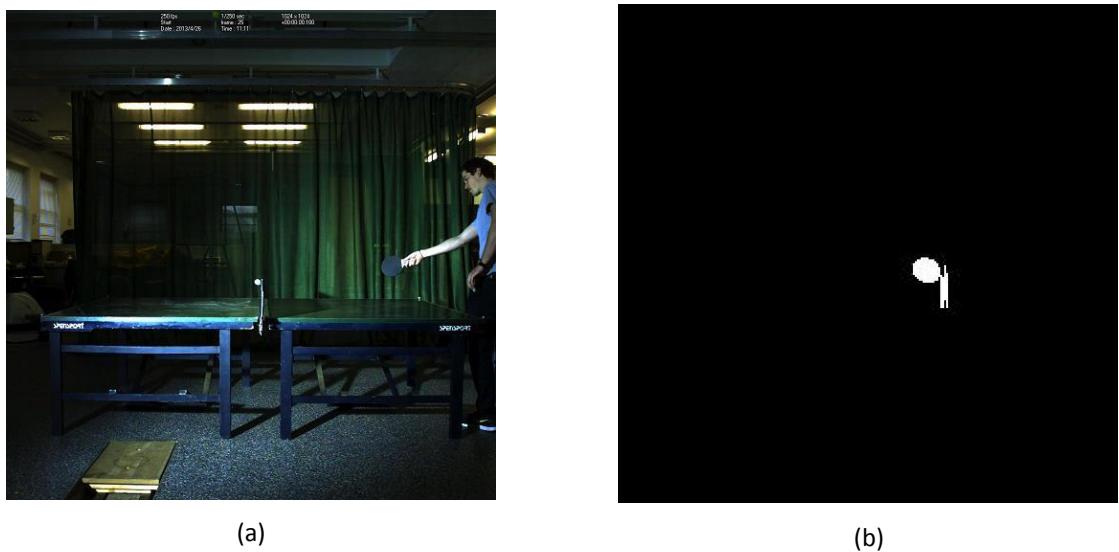


Figure 4-10 Collision (let) induced segmented regions

From this result, an additional process has been designed in TRASE to place a region of interest around the net, called an event activity hotspot (EVH). For the event detection of a let or net collision, immediately post detection of a deviation of the ball's trajectory, a review of additional general rectangular shape segmentations within the EVH uses background subtraction. Observation of these segmented areas provides substantial evidence of the ball making contact with the net.

7. Tracking

Tracking traditionally offers two potential benefits. The first is that during table tennis ball occlusions, the position of the ball in the next frame may be estimated, allowing for constant awareness of the location of the ball. The second is that if the expected location of the ball is known, then it is not necessary to process the whole frame; the algorithm for object detection can focus on a more precise region. A fundamental evaluation of the Kalman Filter (Kalman 1960) has been made using the high quality footage from the Photron Fastcam cameras. Video sequence 100-P-03 has been used for this process due to its relative low noise, reduced background interference and the presence of only a few moving objects in the scene. It was assumed that, if successful using this simple video sequence, the concepts described here may be scaled up for use in a wider field of view (FoV), with other moving objects, players, lighting and background noise.

As a comparator, the video sequences are also processed without the Kalman Filter, using binary threshold only. Figure 4-11 illustrates two resulting frames from the video sequence. Detection of the ball using binary threshold has been highlighted in green and predicted location of the ball using Kalman Filtering is highlighted in red.



Figure 4-11 100-P-03 Processed using Kalman Filter

The images show a relative determined location of the Kalman Filter when compared to the binary threshold algorithm alone. On these and other images there is a substantial overlap between the green and red circles, indicating that the Kalman Filter is able to locate the ball with a degree of accuracy. The Kalman Filter is able to provide a value indicating the expected location of the ball and hence velocity and acceleration variations are possible. For each step in the Kalman Filter applied to the video sequence 100-P-03 above, the location data has been plotted with horizontal positioning (red) and vertical positioning (green) both over time (x-axis) as illustrated in Figure 4-12.

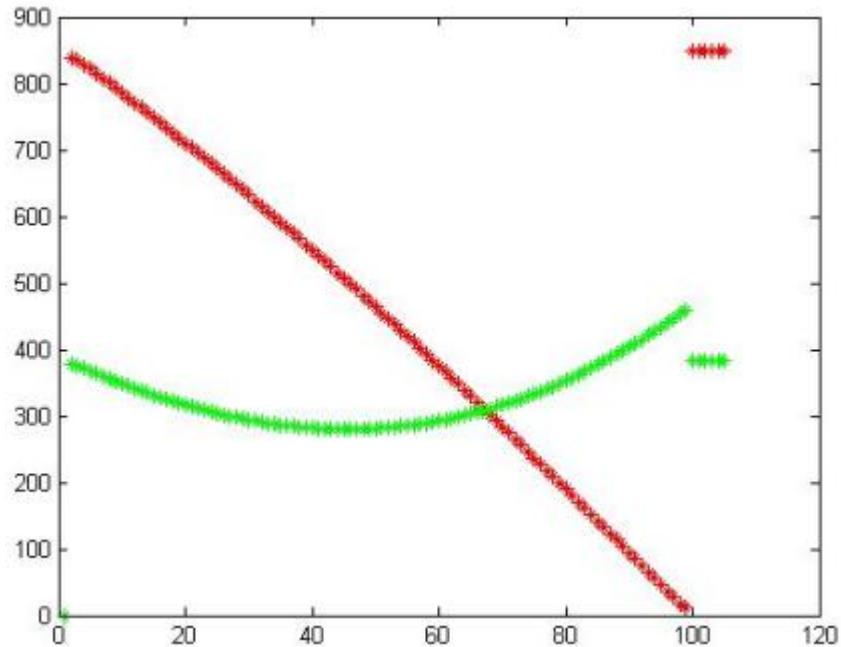


Figure 4-12 Ball location using Kalman Filter

Even without performing direct calculus on the resulting data there is the potential for detecting the slowing of the ball along the red path and the peak height of the ball through the green path. This shows potential for using the Kalman filter in a non-complex image and during free flight. However, when applied to video sequences containing more than the candidate ball, or where the ball takes an unexpected path, such as during a bounce, service or net collision, this becomes problematic. A Kalman Filter expects a consistent, parabolic flight of the ball based on past positions. It will perform well only when the object is moving with salient motion (one direction). It goes wrong soon after a change of direction as the past positions will indicate that the ball is moving in one direction but different to the direction after the ball has been hit. The mathematics of the Kalman Filter can be used to explain this. As a consequence, whenever there is a sudden deviation in the ball's flight due to a

collision, the Kalman Filter needs to be re-configured to re-locate the ball and reconfigure the filter's parameters. This takes several frames to calculate and each frame is lost data. Successful ball detection then requires an increasingly sophisticated segmentation process to avoid false positives. As the ball abruptly changes direction several times per second during a typical table tennis match play, this is a significant challenge within a low-cost environment. Furthermore, whenever there is a sudden change in direction, location data derived from the Kalman Filter is missing and it is the analysis of this change in motion which is necessary for an event based classification solution. Results from experiments using Kalman Filter detailing missed frames are shown in Table 4-2 Kalman Filter missing frame data

Table 4-2 Kalman Filter missing frame data

Video	Event	Event Frame	Frame Lost	Frame regained	#Frames missing
100-P-06-LET	Let	42	36	47	11
300-P-03	Table bounce	324	314	331	17
300-P-04	Table bounce	276	264	285	19
300-P-08	Table bounce	120	113	136	23
500-P-08	Net collision (top spin)	746	738	755	17
500-P-08	Table bounce (top spin)	771	766	780	14
500-P-09	Net collision (top spin)	521	514	530	16

A simpler solution proposed here is not to use a tracker but to consider defining an EVH set within the image where sudden changes in direction are likely to happen. With the observations made of background subtraction during a net collision, for table tennis all EVHs can be summarised as a) the immediate area surrounding the ball, the diameter of which is a function of the ball diameter, frame rate and speed, b) the table surface (and importantly its edges) in the lower third of the image, c) the location of the net defined by the horizontal plane and, finally, d) the area in which the ball is likely to be returned by a players, to be considered at the extreme left and right zones of the image. Using the EVH location definitions, predictions and expectations are made for the region of the ball in the subsequent frame and the cause of any event as it occurs.

8. Ball centre calculation

To mitigate ball distortion errors due to slow shutter speeds or fast play, the centre of the ball needs to be found in each frame. Yet it cannot be assumed that the ball shape and size will be the same in every frame. The impact of thresholding and CV segmentation algorithms on ball shape and size has been found experimentally across multiple video sequences and applying different CV algorithms. Typically the shutter speed is not sufficient to capture the ball without at least some motion blur. Therefore, in this case, the shape of the ball should be treated as an ellipsoid, exaggerated for diagrammatic purposes in Figure 4-13.

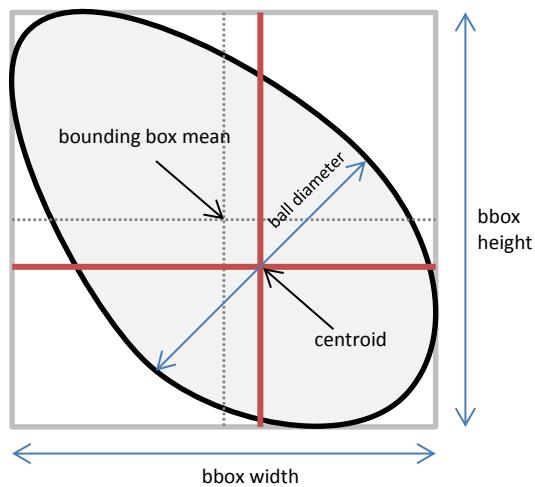


Figure 4-13 Representation of the ball shape when detected at low shutter speeds

The simplest method of finding the ball's centre (by taking the mean of the lengths of the minimum bounding box) is only successful when the ball is uniformly proportional and not sheared or distorted. The rigid ball in table tennis has a fixed, known diameter and using this as the scale of reference, investigations using fourteen of the available video sequences⁶ have shown that when the bounding box method is applied, the difference between ball centres from adjacent frames has an increase of 24% of the diameter of the ball when compared to the same Vicon⁷ data. This is a significant difference and as a first attempt provided an initial benchmark. A second approach is to apply similar methods as those used to detect the centre of a star recorded on a CCD for space craft guidance systems (Arbabmir, Mohammadi, Salahshour, & Somayeh, 2014), this is by calculating the centroid of the image. This process measures the mean image matrix of the segmented object to find its intensity centroid. Applying this to the same fourteen evaluation video sequences, a similar experiment

⁶ Sequences are chosen based on typical speeds from a good player, using consistent services from the service machine. Segmentation was performed using both Canny and Gaussian Mixture Model algorithms.

⁷ This makes the assumption that Vicon derived motion difference between adjacent frames is the ground truth.

comparing distances of centroid location data to Vicon location data across frames reduces the mean difference between the two systems to 8.2%, with the additional benefit of offering sub-pixel accuracy. As the majority of ball segmented sub images are complex shapes, due to over exposure of the image or blurring of the ball during a slow shutter speed, the preferred ball centre calculation in TRASE is the centroid, or centre of mass, algorithm.

9. Variations in shape characteristics

Experimental observations have resulted in evidence that there are a number of ball shape characteristics which are not constant as the ball travels across the frame. Taking video sequence 100-P-06 as an example, Figure 4-14 shows the resulting measurements of the ball's area as it travels from right to left across the FoV, parallel to the FoV of the camera. The ball strikes the top of the net in the centre of the frame and continues to travel towards the opposite side of the table.

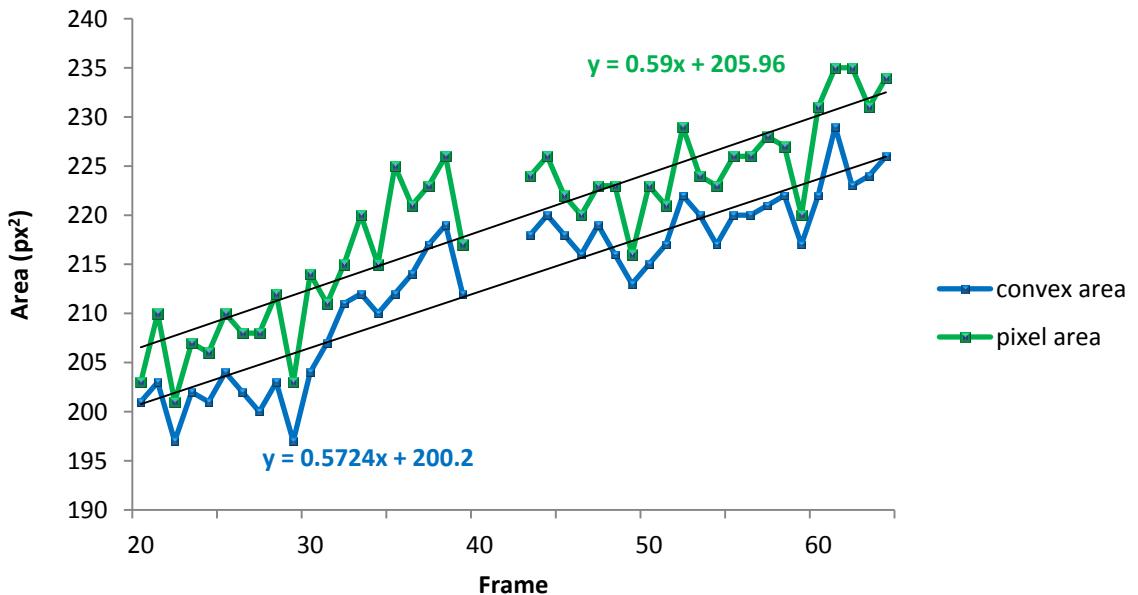


Figure 4-14 Variations in ball area across the frame

The ball enters the sequence at frame 20, leaving at frame 65. The first full ball detection occurs at frame 20 with a convex area of 200.2 px^2 . This appears to steadily increase with a mean convex area increment frame on frame of 0.57% until its convex area is 223 px^2 in the final frame. The area based on pixel count follows a similar gradient of 0.59%. A study of the two diameter methods (ED and BB)

confirmed the ball indeed appeared to be increasing in size, with the BB method diameter increasing by 0.02% per frame and the ED diameter method increasing similarly by 0.02% (Figure 4-15).

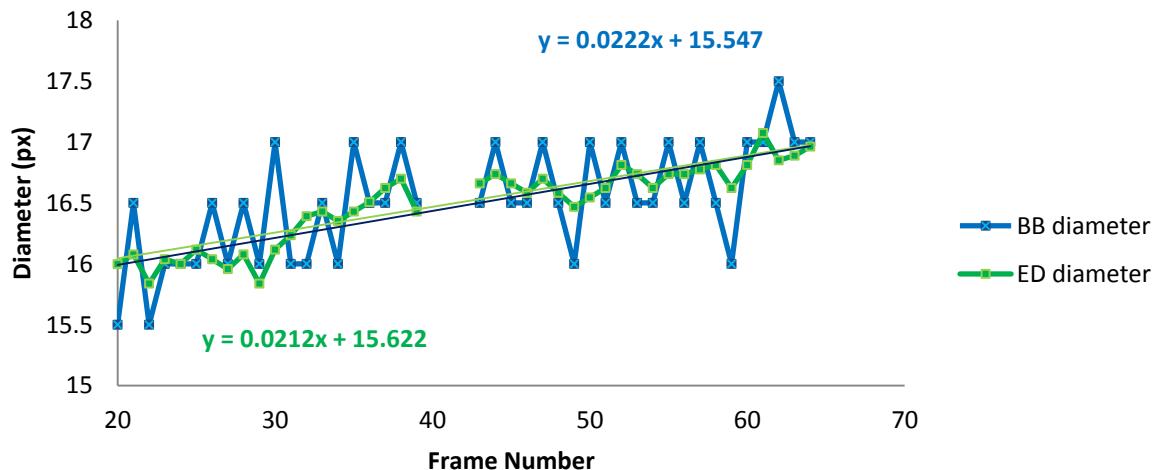


Figure 4-15 Variations in ball diameter across the frame

The first consideration was that this was caused by increased lighting towards one side of the table. However, when investigating solidity and intensity, there is no significant change in the gradient coefficient across the frame, $-5 \times 10^{-6}\%$ and $-4 \times 10^{-4}\%$ respectively (Figure 4-16). To confirm this, recordings of the ball travelling from left to right with the same setup of apparatus produced similar patterns in ball characteristics. Further analysis of two methods of measuring the ball's eccentricity (BB ratio and eccentricity) presented an indication that the ball is tending slightly towards a perfect circle (Figure 4-16).

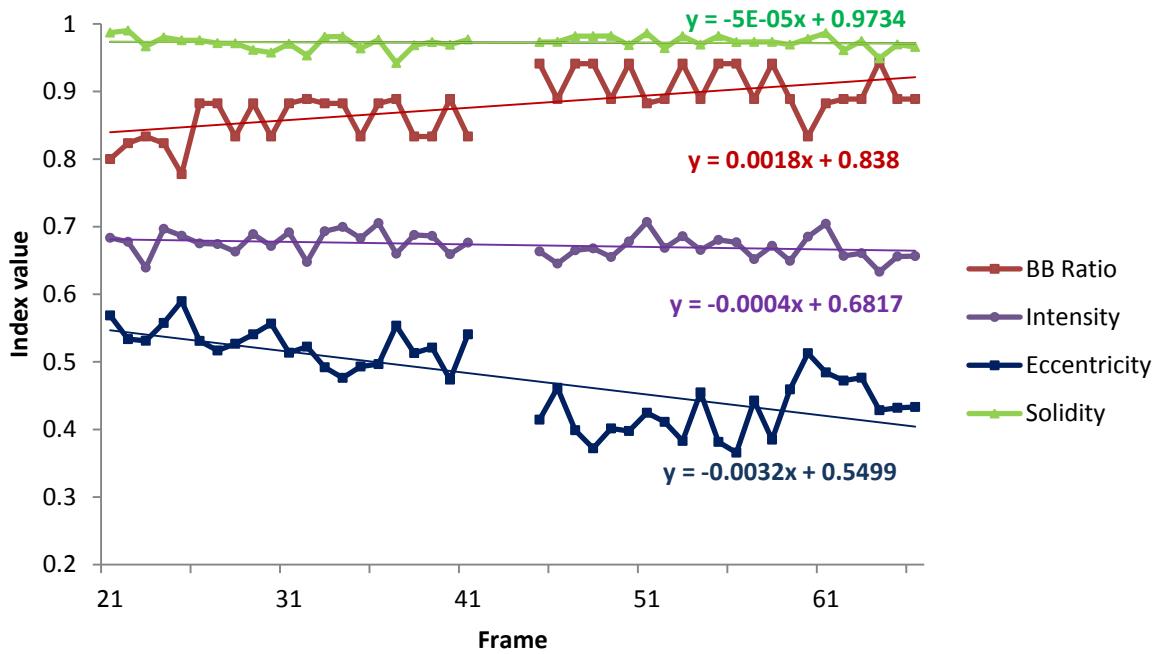


Figure 4-16 Variations in ball characteristic indices

We would expect the ball to decrease in speed as it travels across the table, largely due to air resistance. A slower ball would have an increase in definition around its circumference. But this video sequence also captured an important event, a let, where the ball grazed the top of the net and continued across to the other side of the table. In doing so the ball would have lost more energy than when in the air alone, through heat and sound as it collided with the stretched fabric and decreased in speed⁸. If this is the case, comparing the two halves of each chart, before and after the event would show a step change in the area gradient. This characteristic can clearly be seen in Figure 4-17.

⁸ At this point, the CV detection algorithm being used (Horn-Schunck implementation of Optical Flow) failed to segment the ball and data was missed (the gap in all the charts around frame 41)

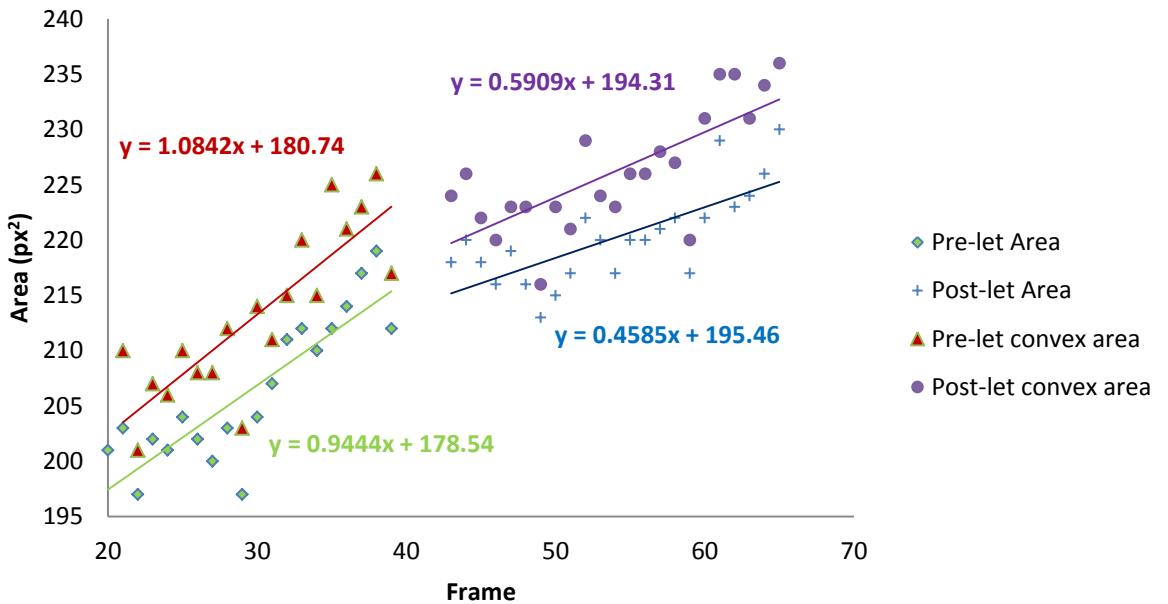


Figure 4-17 Area pre and post let event

For this particular video sequence, the gradient co-efficient of the pre-let convex area measurement is 1.08, changing to 0.59 post-let⁹. The gradient for pixel based area calculation changes from 0.94 pre-let to 0.46 post-let which is consistent with the convex area observation. This clearly demonstrates the step change in gradient is a direct visual representation of the capability of the thresholding algorithm. At fast speeds (prior to the let) the ball appears smaller, as the sensitivity of the threshold is too low for the soft, blurred edges of the ball. This has been confirmed through measurements of the segmented diameter.

As the ball then slows down, the contrast between the ball and the background increases allowing the thresholding to segment more of the ball as shown in Figure 4-18.

⁹ This change of 50% in the gradient is nothing more than coincidence.

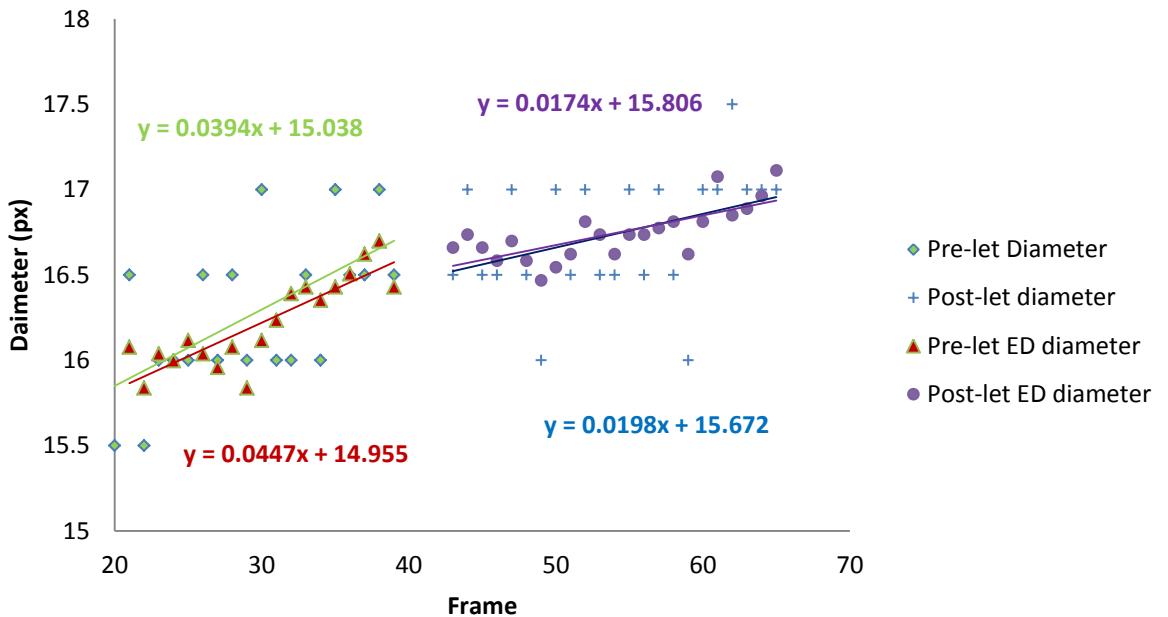


Figure 4-18 Diameter measurements pre and post let event

Not only is this a tool for optimisation of the threshold value, but also provides an indication of the occurrence of an event. The difference in gradients also provides a relative indication of the change in ball speed. This finding requires further research and needs to be considered in motion metrics analysis, if the size of the ball is to be used as the scale of reference.

10. Ball diameter calculation

The width of the segmented moving ball (when an optimal threshold is used) will always appear larger than the width of a segmented stationary ball, due to motion blur. This can be reduced with high shutter speeds, but no matter how fast the shutter there will always be blurring up to the point where ball motion is less than one pixel whilst the shutter is open (at which point any inaccuracy in measuring the ball in the image is limited by the resolution of the camera sensor). However, if the ball is to be used as the scale of reference, as proposed in this research, then a reliable estimate of the diameter must be made. For comparison, presented are two options for calculating the diameter of the ball: a) bounding box (BB) and b) equivalent diameter (ED). The BB method defines the rectangle with a length and a width which can just cover the image of the ball. The dimensions of the rectangle are calculated and the diameter is assumed to be the mean of these two values. The ED calculates the area of the region defined by the ball (i.e. counts the pixels within the region) and applies the following equation (4-7):

$$ED = \sqrt{(4 * \frac{area}{\pi})} \quad (4-7)$$

The BB accuracy is limited by the image resolution and is a useful approximation, working best with high shutter speeds (to ‘freeze’ the image of the ball) and high resolutions. The ED method in comparison is proportional to the area of the ball (The MathWorks Inc., n.d.-c), offering the potential of finer precision. A comparison of the ball diameter ranges using the two methods was made, repeated across seven video sequences covering a variety of size, speed and colour (Table 4-3).

Table 4-3 Comparison of ball diameter calculations

Sequence	BB Range	ED Range
100-P-06	15.5-17.5	16.3-17.1
200-P-03	15-17	16.1-16.9
300-P-03	17-23	18.1-20.4
600-P-08	14.5-18	15.3-16.2
600-P-09	13.5-18	14.9-15.9
700-P-10	15.5-18.5	16.2-17.9
700-P-15	16-19	17.0-18.1

It can be seen that the ED method does appear to provide a more consistent smaller range of ball diameter values, all within the range of the BB results. With this indication ED was chosen in TRASE for the continued research into using the ball dimensions as the scale of reference when determining its speed and acceleration. These results also provide a useful general guide for the size of the image of the ball on the sensor. Additionally it has been found that for between 2m to 3m distance, the ball covers a horizontal diameter of between 13 and 19 pixels when using the Photron Fastcam Ultima APX with its relatively small horizontal resolution of 1024 pixels.

11. Variations in CV algorithm’s positional data

When extracting the ball from the video sequence, the pixel location (x, y) of the centre of the ball varies between algorithms. This is due to the differences in processing to consistently find a boundary of an object when it is moving at high speed, or when slightly out of focus or with variations in lighting. As an example, when reviewing a simple table tennis video (100-P-06) and comparing image co-ordinate values for each algorithm, at first sight (Figure 4-19 and Figure 4-20) there is little difference between the algorithms.

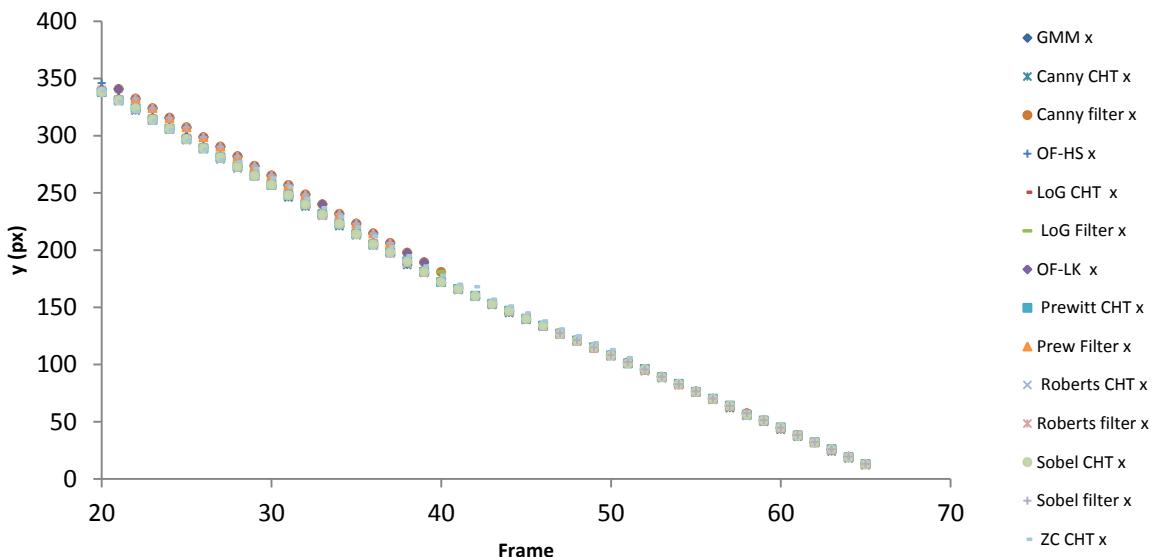


Figure 4-19 Algorithm y-values for 100-P-6

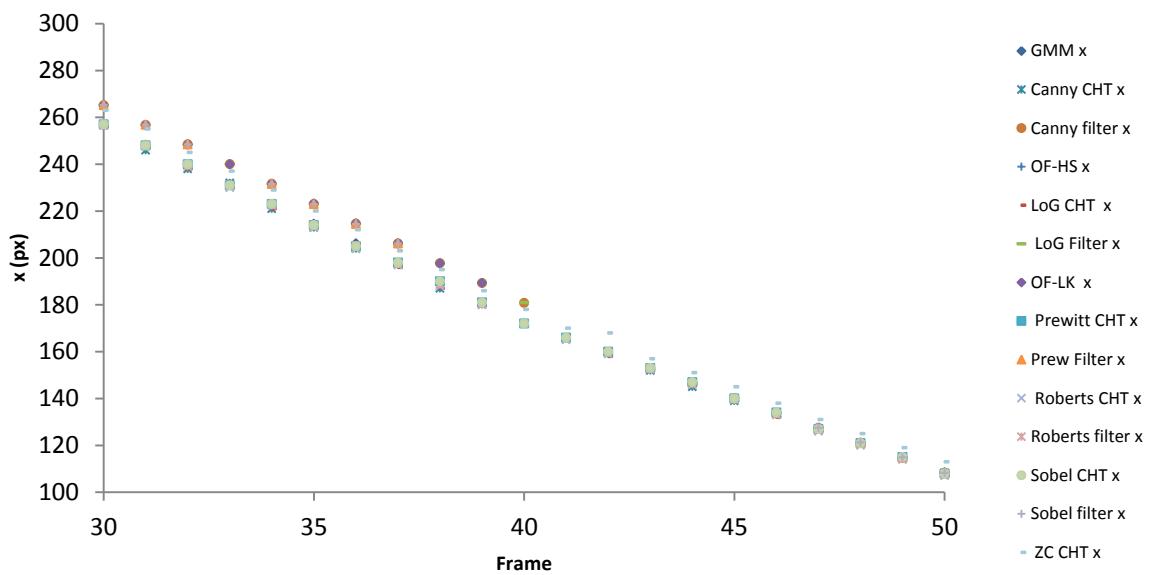


Figure 4-20 Algorithm x-values for 100-P-6

However a closer view of a section of the data (Figure 4-21) begins to show there are significant differences, particularly when the ball is travelling faster (prior to frame 41). These small differences could have a large effect on detecting events.

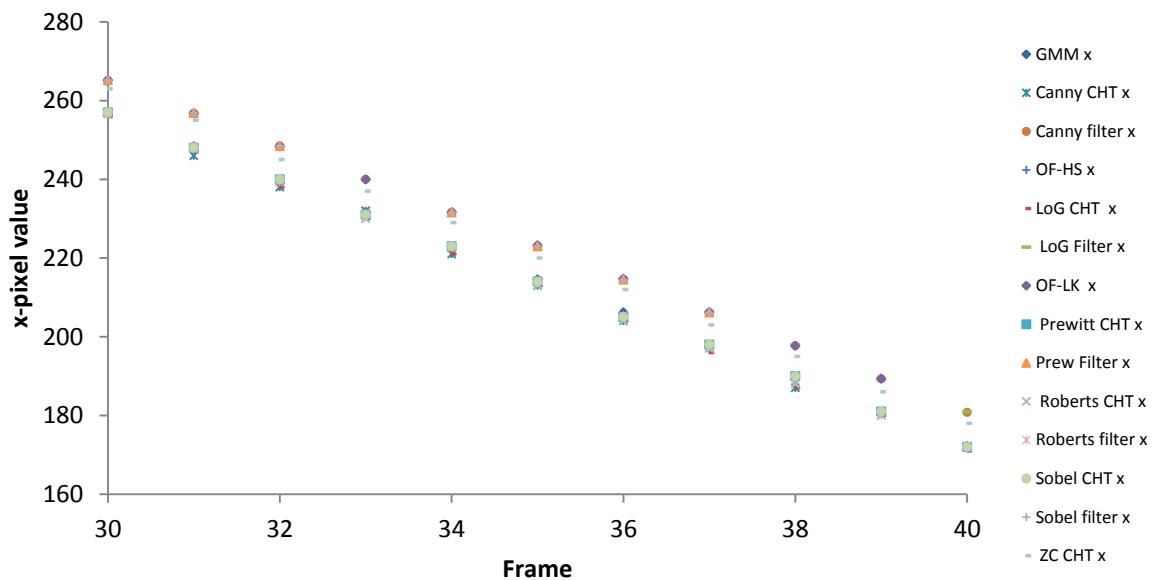


Figure 4-21 100-P-06 x-values between frames 30 and 40

Consider a let call, where the ball makes contact with the top of the net and continues its forward motion to the far side of the table. To satisfy the hypothesis for this research, this event must be detected by a sudden change in either one or both components of the motion vector (horizontal or vertical). With the angle of the camera placed in line with the net (the defined optimum location), it would be natural to expect a deviation in the y (or vertical) component of the vector, with the smaller the area of contact during collision, the lower the angle of deviation. As the angle reduces, the ability for the CV algorithm to successfully detect a deviation also reduces, as the deviation falls outside the tolerance range of the algorithm. Video sequence 100-P-06 contains a let event and Figure 4-19 visibly shows a change in gradient; however the gradient change varies between the algorithms. Reviewing the outputs of 100-P-06 we can measure the minimum and maximum values for x and y for each frame and review the differences. On doing this we see the results presented in Figure 4-22.

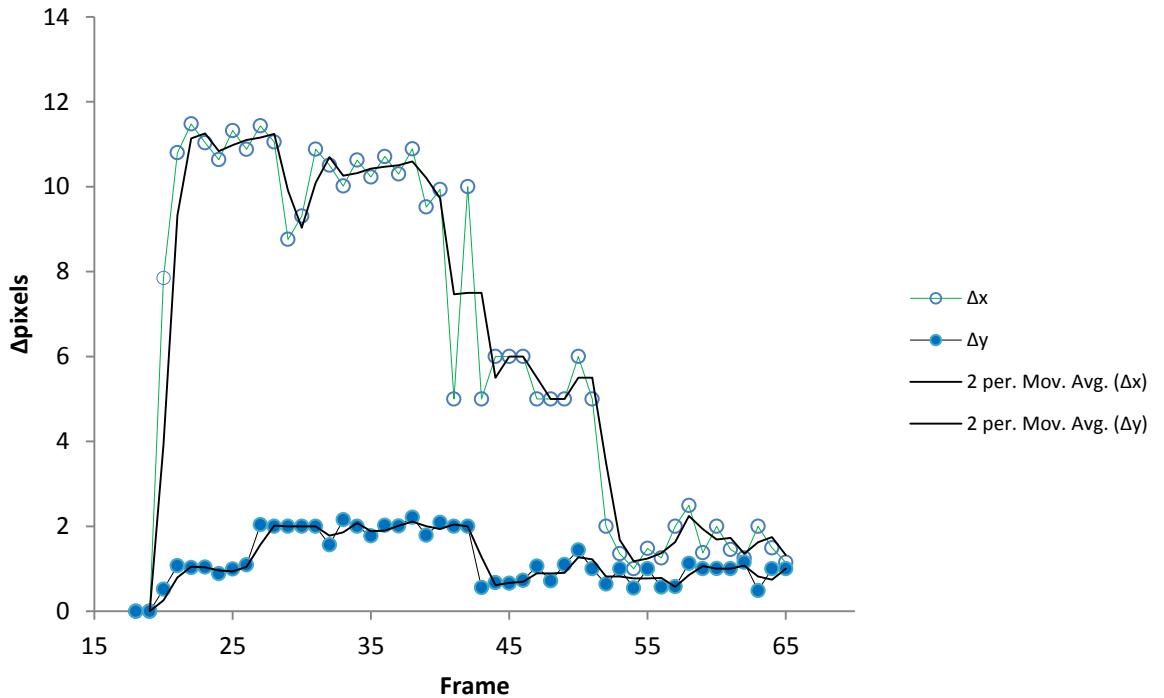


Figure 4-22 Maximum difference in ball location between algorithms (100-P-06)

Again we see a maximum difference across the selected CV algorithms of 11.47 pixels in the x-plane and 2.21 pixels in the y-plane. From the point of striking the net this falls to below 6 pixels. It is this discrepancy between the data values which requires further 2D location validation.

12. Reflections, lighting and shadows

It has been observed from the experiments that as the ball approaches the proximity of the table (usually within 15 cm) a reflection of the ball can appear with similar image characteristics to the actual ball. The shapes, size and relative motion cause the segmentation algorithm to find multiple candidate objects within the frame, obviously detrimental to the overall performance and success rate of the filter. Then, as the ball comes towards the table, the segmented image of the ball and the segmented reflection merge through a wedge sum of circles, through a mathematical rose where $n = 1/2$ (Weisstein [2] n.d.), into a highly elongated oval making automatic ball filtering and detection during this phase of the ball's motion difficult to identify, analyse and predict. The reflections may not be apparent to a player, perhaps due to their position and height in relation to the table. As the ball gets within this reflection range of the table, multiple OOs are detected in each frame causing noise and complications in detecting the point of the bounce, again required for accurate dynamics.

analysis and event analysis. An example of this occurrence (video sequence 600-P-10-OF-LK), prior to implementing the feature filter module, can be shown by the trajectory output of TRASE highlighted in red in Figure 4-23.



Figure 4-23 Issues of reflection as observed in TRASE

In this example the ball is travelling right to left as viewed from the camera. As the ball arcs towards the left hand side of the table a reflection is detected before the bounce, creating what appears to be another ball along the plane of the table underneath the trajectory of the ball. This becomes a problem when identifying changes in trajectory, as the point of contact with the table is important when establishing a bounce event, evaluating the legal area boundary of the game, or when statistically analysing a ball's motion during play. A variety of balls and two tables were tested with different reflective surface properties and the same effect was observed. The cause is, therefore most likely, due to a combination of overhead lighting and the deliberately positioned low angle of the camera. It has also been observed that shadows of the ball can appear on background surfaces due to the positioning of lighting. The shadow exhibits many characteristics of the ball, size, location and motion for example. However, they commonly differ in brightness, intensity and colour.

Figure 4-24 represents the output from another section of the same video (600-P-10-OF-LK) prior to filtering, where the general parabolic motion of the ball is clearly apparent to the human eye and

also where reflections on the table as the ball approaches its surface become false positives (NB: the false detections above the curve are caused by minor lighting and background complexities).

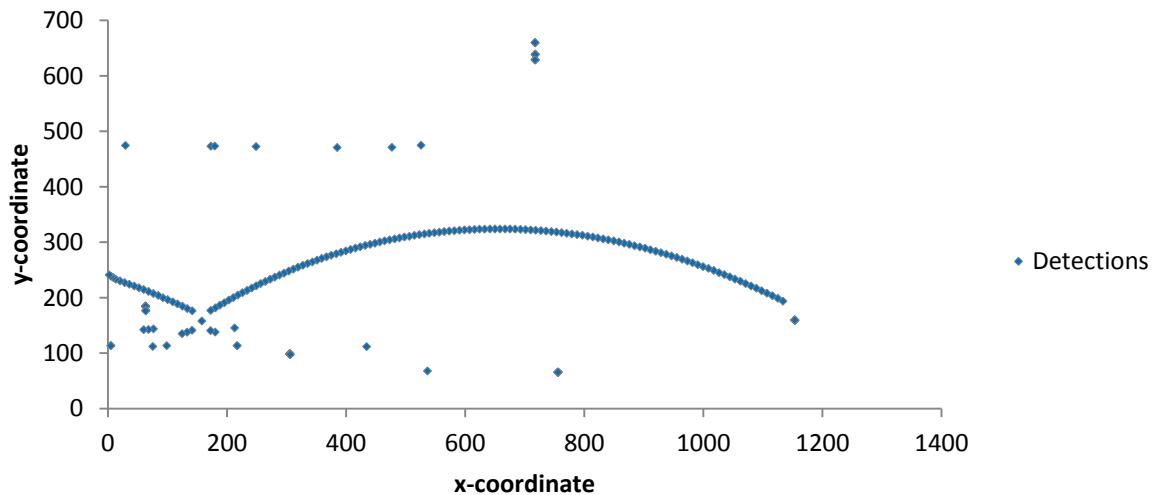


Figure 4-24 Pre-filtering of TRASE output including reflections

The sample data in Table 4-4, taken from sequence 600-P-10-OF-LK, show pairs of records for both the ball and its reflection, prior to the ball making contact with the table. Analysing this particular data clearly demonstrates the difference in shape of the reflected object (actual ball data is in grey) where the height and width of the bounding box around the ball is closer to being equivalent, compared to the reflected object.

Table 4-4 Example reflection data for 600-P-10 (OF-LK-100-600-10)

frame	x	y	dimx	dimy	blobnum	centroid	perimeter
142	240	121	39	9	160	-0.09	0.98
142	242	245	30	28	161	0.21	0.41
143	231	121	41	9	162	-0.09	0.98
143	234	240	30	27	163	0.21	0.43
151	170	117	31	6	181	-0.06	0.98
151	174	204	31	27	182	0.28	0.46
149	186	118	32	7	176	0	0.98
149	189	214	31	28	177	0.26	0.45

As can be seen from this data sample, each frame has multiple candidates for which the Optical Flow algorithm is unable to distinguishing the ball from its reflection. It, therefore, generates a pair of candidate values in each frame with both appearing and moving similarly. From this data, we can see that: a) the x-axis coordinate is very similar for both, b) the y-axis coordinate of the reflected object is lower (further down the image from top left after being transformed), and c) the value for the centroid is always greater in the actual ball.

There are examples of solutions for the removal of reflections from static images using CV, such as the proposal by Tan et al. (2004) who suggests not to use explicit colour based segmentation due to complex colour textures, but instead describes an algorithm applying chromaticity and repetitive intensity differentiation. However, detecting, isolating and avoiding reflections within sports video sequences have not been fully researched. In the particular case of table tennis, the reflections are not ‘mirror-like’ in their appearance. Considerations were made when designing the filter of the objects to discard the lower trajectory, or to detect the trajectory patterns which mirror those of a natural flight. Unfortunately, these proved unreliable and resource intensive. Moving the camera to a higher position created new problems by making the detection of bounces, net collisions and service infringements more complex with monocular hardware. Instead, improvements to the feature filter were made to isolate the ball from the reflection by examining the solidity and colour properties of the reflection or shadow. Concepts, suggested by Tan et al., to examine intensity differentials are also included in the filter.

Although not an initial design consideration for the experiments, there are several video sequences within which reflections and shadows have been observed; therefore these videos have been used as the basis for improving the feature filter. The observations of reflection and shadow are directly responsible for the additions of intensity and solidity parameter configuration in the feature filter. Presented below are the results of three experiments using the improved feature filter. Table 4-5 shows a number of segmented images of a table tennis ball from a series of three video sequences: Experiment 1 is from sequence 600-P-05-OF-LK; Experiment 2 is from sequence 600-P-07-OF-LK, while Experiment 3 applies the filter to sequence 600-P-09-OF-LK. All have identical frame rates (250 FPS) and use the Photron Fastcam Ultima APX camera.

Table 4-5 Observed ball and reflection pairs

Experiment 1		Experiment 2					Experiment 3

In the case of all three experiments, each candidate ball has been detected using the same Optical Flow (OF) algorithm with identical configuration settings. The top row contains the actual ball, the bottom the observed reflection. Each pair of images is from the same frame and has been segmented in an attempt to isolate just the ball through segmentation with the reflected image being a false positive segmentation. In the case of experiment three, two minor reflected images have been detected and given here for comparison of the variety of reflected aberrations detected. Careful observations of the images can give an indication to the direction and motion of the ball in its proximity to the table.

An additional experiment using images from 700-P-12 demonstrates a sequence of four frames captured either side of a table bounce using the improved filter (Table 4-6). The reflection of the ball is clearly visible on two of the frames either side of the bounce before the filter is applied; the reflections are ignored after using the filter, not only reducing the number of candidate balls, but also improving the bounding box dimensions used for ball identification and scale of reference.

Table 4-6 Reflection sequence in 700-P-12

Frame	Pre filter	Post filter	Description
480			The ball is moving right to left, in a downwards motion at an angle approximately 30° to the surface of the table
481			The ball is still moving downwards, immediately prior to contact with the table. Segmentation detects both the ball and the reflected image.

Frame	Pre filter	Post filter	Description
482			The ball has made contact with the table and now moving away from the table. Segmentation shows both the ball and the reflected image.
483			The ball has moved away from the table

A review of the set of ball metrics, derived from CV processing alone, indicate that there is an optimum sequence of stage when filtering objects in the table tennis video sequences. This is discussed in the following section.

Filter ordering

It is important for the feature filter to processes images based on shape characteristics. However, the order in which they are processed is equally important; each filter stage should remove the greatest number of false detections. If processor speed is important, then the fewer filter stages required, the greater the detection performance in return. Once only one candidate remains, the filtering is halted and the next frame can be processed. To establish the order of each filter stage, TRASE was configured to filter only one characteristic at a time. A baseline was determined by selecting a video sequence with a large number of reflections (600-P-10-SEQ5) and allowing TRASE to detect all possible candidate balls using the Gaussian Mixture Model algorithm. The video was then re-processed with one of the filters selected and the difference in the number of ball candidates, as a percentage, was recorded. This was repeated for all nine filter characteristics. This analysis was then further repeated across sixteen video sequences containing a selection of bounces with reflections, logos and lighting artifacts which were testing of the filter, generating a mean percentage difference across a range of different video sequences. This outcome for each video sequence was found to be generally consistent, the conclusion of which is presented in Figure 4-25.

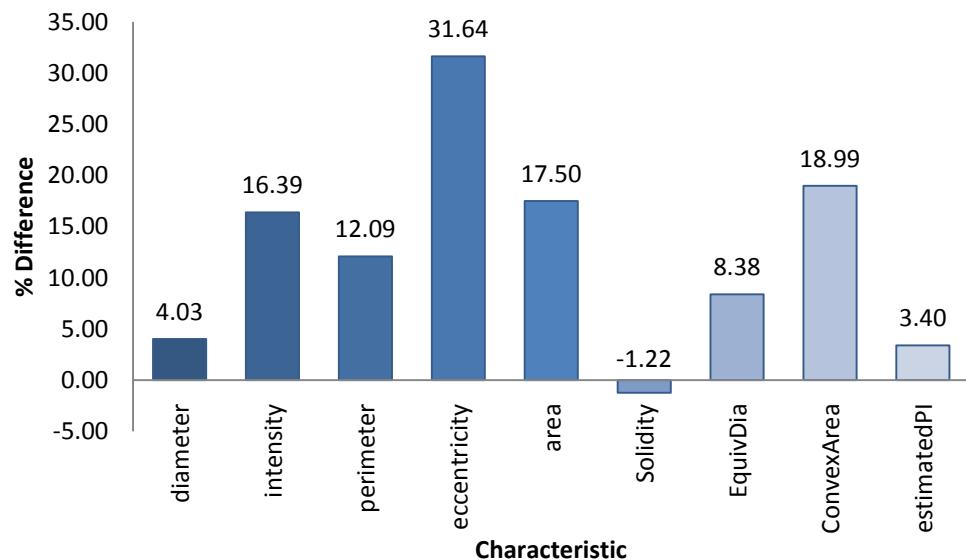


Figure 4-25 Comparison of characteristic filtering

We can see from this that there is a range of success when applying each filter individually. The greater the percentage difference, the more successful the filter has been¹⁰. In order of priority, the following features (Table 4-7) have the greatest effect on isolating the ball from the reflection:

Table 4-7 Feature priorities for ball Isolation

Priority	Feature
1	eccentricity
2	convex area
3	area
4	intensity
5	perimeter
6	equivalent diameter
7	diameter
8	estimated pi
9	solidity

¹⁰ It is possible to have a negative difference where the value being compared is an index, as was the case here for the solidity characteristic.

Of interest is the preference *equivalent diameter* holds when compared to the *diameter* characteristic. Additionally, the convex area measurement was more successful than the basic pixel area measurement. Also of note are the minimal benefits of *estimated Pi* and the improvement filtering on *intensity* when compared to *solidity*. Considering processor implications and duplication of effort, a refined optimum order for use in table tennis ball detection filtering is as follows (Table 4-8):

Table 4-8 Refined feature priorities for ball Isolation

Priority	Feature
1	eccentricity
2	convex area
3	intensity
4	perimeter
5	equivalent diameter
6	estimated pi
7	solidity
8	area
9	diameter

The filter executes each of these modules in turn until only one candidate object remains. The filter then halts and the ball metrics may be measured and output for event analysis.

4.5. Qualitative comparison¹¹

Before providing a quantitative analysis of the capabilities of both edge detectors and foreground extractors, provided here is a qualitative review of selected CV algorithms when applied to a single image from each of three representative videos, 100-P-03, 100-C-14 and 9999-U-01. Video sequence 100-P-03 was chosen for its clarity and distinction of the ball from the background. The 100-C-14 video was selected for the unusual camera angle and lower quality of video. Finally 9999-U-01 was identified to represent a more challenging source. The raw images are presented in Figure 4-26 A1-A3 below, along with their corresponding outputs for Canny (row B), LoG (row C), Prewitt (row D), Roberts (row E), Sobel (row F) and Zerocross (row G).

¹¹ This section should not be read in isolation, but in conjunction with the detailed quantitative analysis Chapter 5.

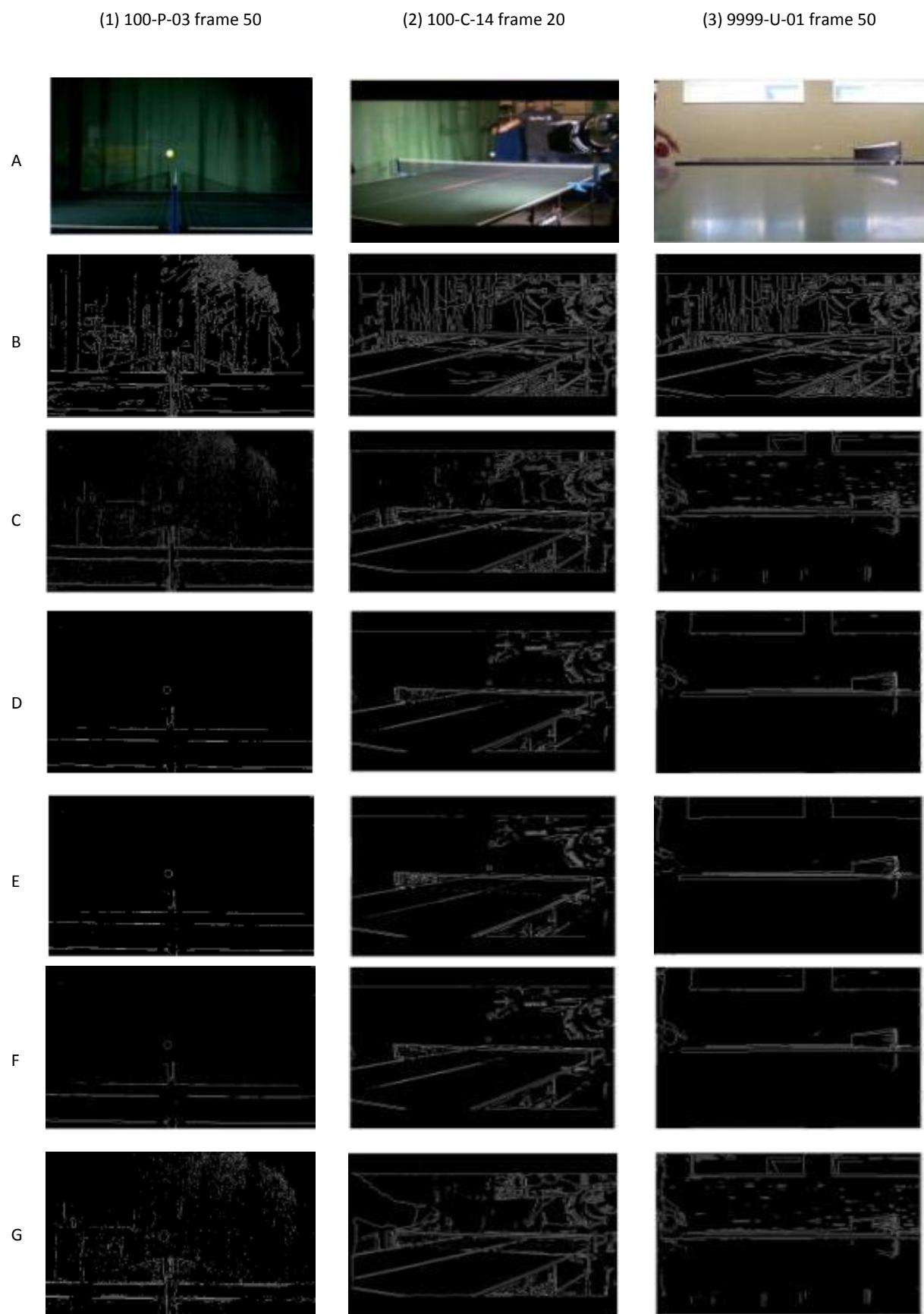


Figure 4-26 Edge detector results

The outputs also provided further confirmation of the functionality and desired outputs of the coded algorithms being employed in TRASE. Below is a commentary of the observations of the algorithms, their potential strengths and problems encountered whilst processing the images.

1. **Canny (B)**

Taking video 100-P-03 frame 50 with a lower threshold of 0.125 and an upper threshold of 0.0313, we get output B1. The ball is visible in the centre of the image. The table and net are also identifiable in the bottom third of the image. There is a great deal of noise however, particularly in the background where the green netting is and where the reflection of the net casts a shadow in the background from the two high powered lights. Various thresholds were applied; this was the best quality image with the ball still visible and the minimum of background noise. In 100-C-14, the ball is clearly identified using Canny Edge as seen in image B2. The perimeter of the net and the table are also clearly defined. The complex background textures are causing some unwanted region detection. The thresholds in this image are from 0.0188 to 0.0469. To the eye, the Canny edge detection algorithm has successfully identified the player and the racket. The ball and table have also been isolated; however there is a great deal of noise and so filtering out just the OOI will require an extra stage of processing and further resource. The net (and its reflection in the table) has been clearly segmented. The threshold value was $0.0125 < T < 0.0313$.

2. **LoG (C)**

Examining frame 50 in trial 100-P-03 the LoG edge detection algorithm successful isolates the ball from the background, defines sharp lines for the table and reduces some of the background noise when compared to Canny edge detection. $T=0.015$. In sequence 100-C-14 at frame 20re too we clearly see the table and net boundaries, with the ball separated in the middle of the frame. It is possible to also detect the region around the player along with their extended arm. Background noise is minimal. The threshold applied to this image is $T=0.0913$. Finding the ball in 9999-U-01 frame 50, even with the human eye, is not easy. The straight edges of the highly contrasting table are clearly visible, with the bow in the net showing its curvature. The racket is also clearly represented in this image, at least as well as Canny but without the extra noise. The threshold used for this sequence is $T=0.0022$.

3. Prewitt, Roberts and Sobel (D, E and F)

When applying Prewitt, Roberts and Sobel, outputs are similar in appearance and the ball is easily been segmented from its surrounding area, with less background noise than LoG. However this is at the expense of a clearly defined table. This indicates the possible need for different algorithms for different aspects of the image: one for the ball, one for the table and one for the player. Examining these three algorithms against 100-C-14 the table and the ball have been distinguished well. These algorithms also delineate the table clearly final image 9999-U-01. In all three algorithms the ball has, perhaps surprisingly, also been successfully outlined against a very similar contrasting wall, albeit the ball shape is distorted due to camera settings. Threshold is $T=0.00713$.

4. Zero Cross (G)

Bearing some similarities to the Canny edge detector, when implementing Zero Cross to 100-P-03 frame 50, the OOI has been extracted (with a broken edge) along with most of the table. However, the background noise will make most frames in this high quality video more difficult to segment the ball. In 100-C-14 frame 20 both players have clearly been segmented. The lines of the net and table edges have been highlighted well, useful for most shot analysis applications. The ball has also been selected, but there is still much noise around these objects for the process to contend with. In Zero-cross 9999-U-01 there are fewer detected regions than Canny, but not as few as Prewitt, Sobel and Roberts. When implementing this algorithm, consideration should be given to extra resource overheads and the increased likelihood of false detections.

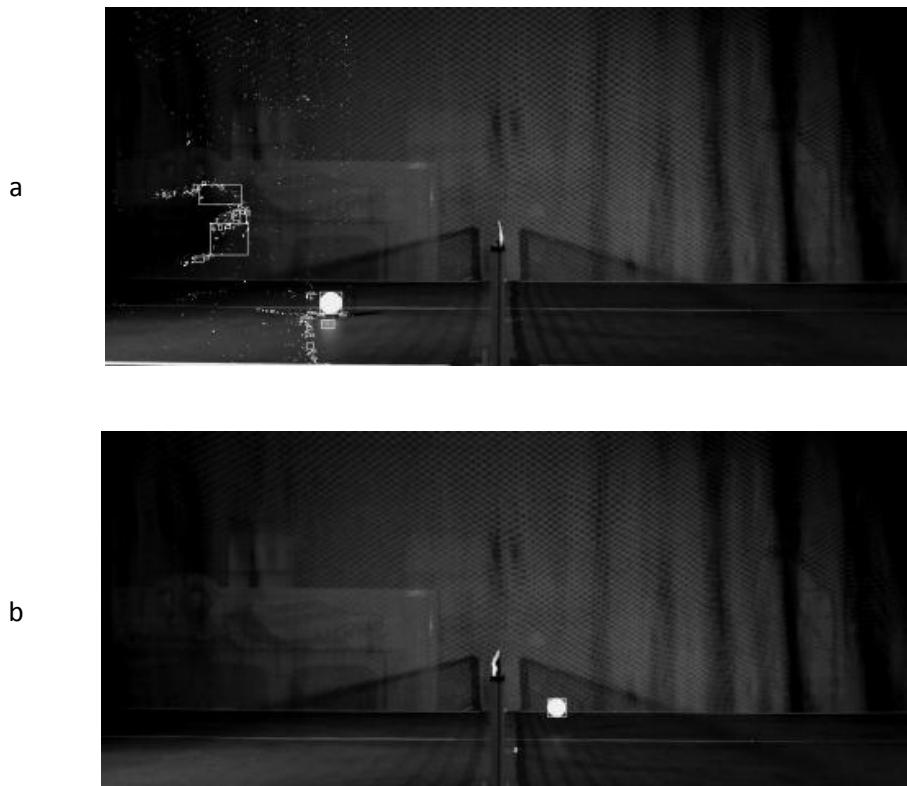
Aside from the described edge detector algorithms, the Gaussian Mixture Model (GMM) and two implementations of optical flow algorithms have been processed against a series of images for discussion below.

5. Gaussian Mixture Model

The video frame was taken direct, no image pre-processing, except for converting to grey-tone. Initial results from applying GMM demonstrate a general success in detecting the moving ball, but with many outlying false positives. In Figure 4-27a a number of false positive detection boundaries can be seen clearly to the left caused in this case by subtle changes in shade from the flicker of fluorescent lighting. Filtering using ball features will generally remove these false positives. Figure 4-27b demonstrates the good success rate achieved when the only moving object in the scene is the ball.

As in the previous image, the border placed around the ball provides location and ball dimensions, a requirement for assessing accuracy and determining rules compliance. In Figure 4-27c the shadow of the ball in the following image demonstrates one of the complications of the GMM. Thresholding the image, or converting it to binary, may avoid these difficulties. Further research in this area will be followed up in subsequent experiments. Figure 4-27d clearly demonstrates some of the issues of the GMM. Of particular problem here are the moving arm, racket and shadows, all clearly identified by the algorithm. This could be reduced with a combination of post-processing region of interest (ROI) and feature filtering.

A point to note about applying the GMM is that it only detects the moving object. So whilst generally good at detecting the ball and its size and boundary, assuming the camera is stationary, it will not detect the position of the fixed table. This is a further indication of the need for a combination of different segmentation algorithms for different objects within the table, if necessary.



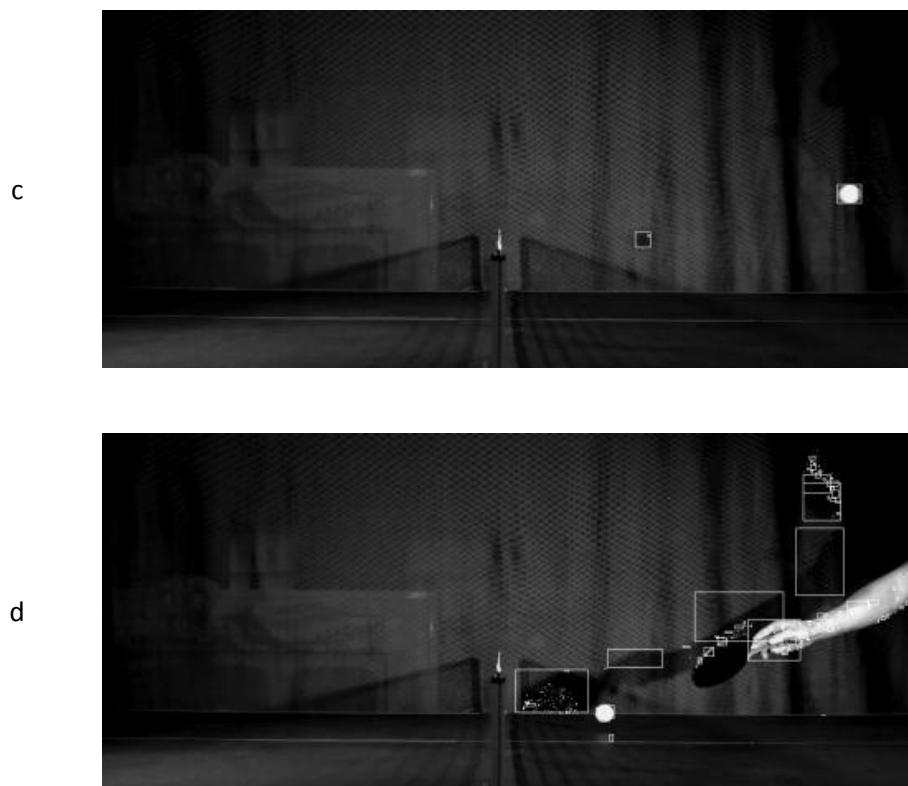


Figure 4-27 GMM applied to 100-P-06

6. Optical Flow

The Optical Flow Horn-Schunck (OF-HS) algorithm has been applied to three video images of varying quality, with all images pre-processed to grey tone. In the high quality image from the Photron Fastcam Ultima APX, the only moving objects are the table tennis ball and its shadow. As can be seen from Figure 4-28a, OF-HS was successful in detecting and isolating the ball from the scene. Its motion is represented by applying the vector rays around the ball's image. Several frames later in the same video (Figure 4-28b), it is clear that OF-HS has also successfully discarded the shadow on the background. However, in the more complex image taken from sequence 100-C-14 (Figure 4-28c) taken from the pilot study, OF-HS has also detected movement around the player's hand and also the logo on his clothing. With filtering, on initial observation, this still suggests a more successful algorithm when compared to GMM.



Figure 4-28 Applying Optical Flow (H-S) to detect the ball

When applied to the lowest quality sequence 9999-U-01 (which was made prior to this research and for which details were not recorded) it can be seen in that the OF-HS algorithm has successfully detected the ball (Figure 4-29a). However, just one frame later (Figure 4-29b), the hand and racket

motion is still detected but the ball goes undetected. Initial areas for investigation are to enhance the contrast within the image or to use a Region of Interest (RoI) algorithm to assist in isolating the ball. However, this will require further analysis to discover the true success rate.

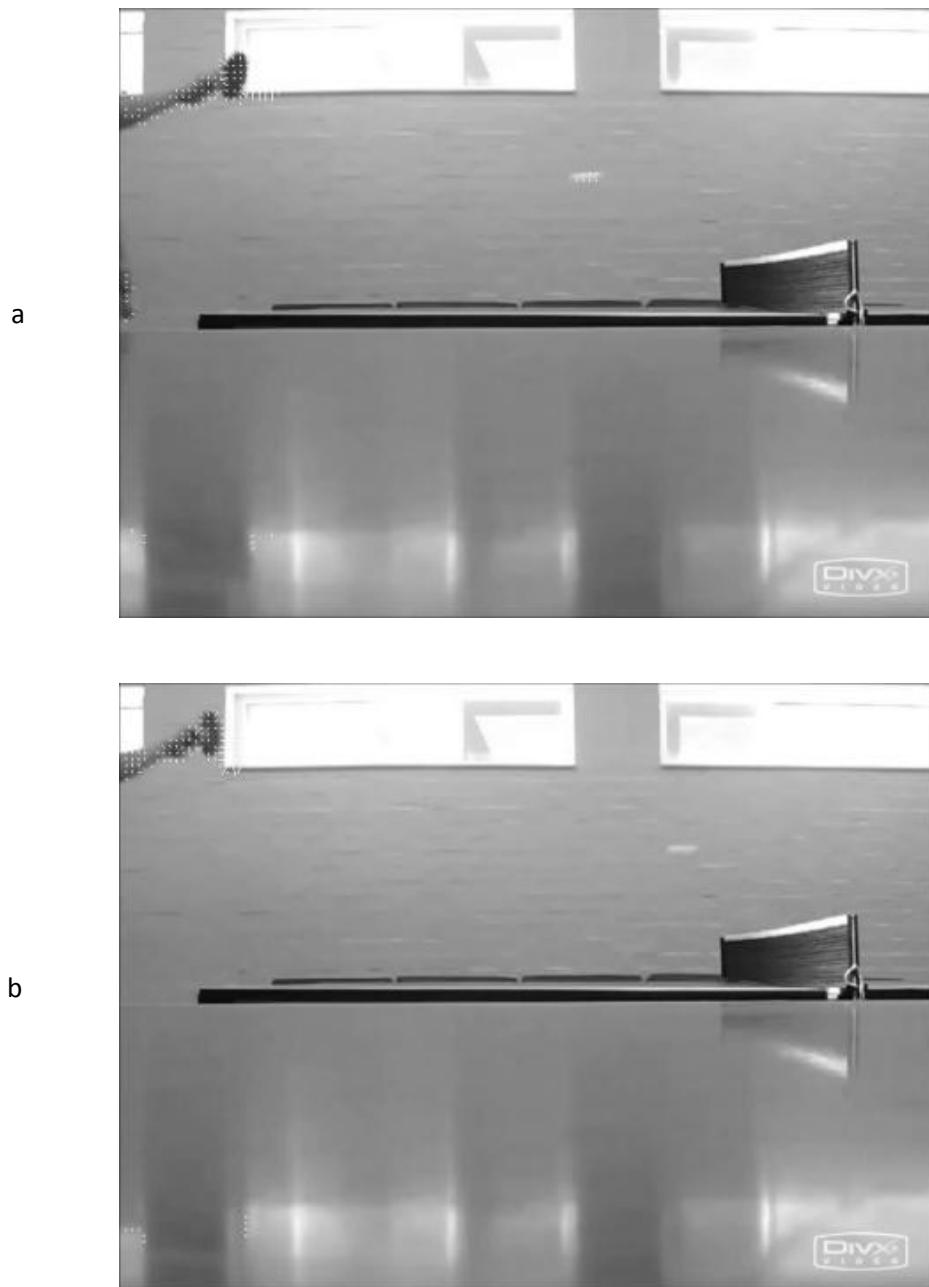


Figure 4-29 Optical Flow applied to 9999-U-01

Optical Flow: Horn-Schunck vs Lucas Kanade

A qualitative analysis of OF-HS compared to the Lucas-Kanade Optical Flow implementation (OF-LK) has been made. Across many videos, Horn-Schunck appeared to show a greater success in identifying the ball with the image when compared to OF-LK. Conversely, this then became increasingly error prone with false positive detections. For example, samples from video sequence 700-P-10 shows a clear difference between the two algorithms. OF-HS is capable of detecting more occurrences of the ball, but at a reduced level of overall accuracy. In key frames analysing the beginning, middle and end sections of a let event, it is becomes apparent that there is a difference between the two segmentation processes Table 4-9.

Table 4-9 Horn-Schunck vs Lucas-Kanade in key frames from 700-P-10

Frame	LK candidate detections	HS candidate detections
1		
2		
3		
4		
5		
20		
23	No detection	
24		
25	No detection	No detection
26	No detection	No detection
27		
80		
81	No detection	
82	No detection	
83	No detection	
84	No detection	
85	No detection	
86	No detection	

Frame	LK candidate detections	HS candidate detections
87	No detection	[]
88	No detection	[]
89	No detection	[]
90	No detection	[]
91	No detection	[]
92	No detection	[]
93	No detection	[]
94	No detection	[]
106	No detection	[42]
107	No detection	[42]
108	No detection	[42]

It may be deduced from this one video sequence that the OF-HS algorithm is better at segmenting objects during low contrast between foreground and background. However, this also leads to a greater potential for false ball detections. These observations are also confirmed from the qualitative analysis of the algorithms.

4.6. Qualitative summary

When the selected edge detectors and foreground extractors are compared to each other using the same (single) video sequences an initial assessment and understanding of the relative success of each, and the benefits each may bring, is presented in Table 4-10. This qualitative analysis provides a useful context within which to interpret comparison results. These learnings then form the basis of a less subjective and more rigorous detailed quantitative analysis.

Table 4-10 Initial qualitative assessment

Process	Qualitative Results			Qualitative Suitability	Qualitative Preference?
	100-P-03	100-C-14	9999-U-01		
Canny	Large number of segmentations detail and noise. Ball detected with broken lines surrounding ball. Complications in identifying table against background.	Good definition, background noise/detail also segmented, clear table identification	Noise level in background high, net and racket clearly identified	Due to high level of background noise detection, not suitable for any segmentation.	No
LoG	Ball clearly segmented with complete lines. Background noise reduced over Canny. Table and net visible, however confused by background shadows.	Table and net clearly identified. Ball segmented successfully. Reduced noise over Canny.	Reduction in noise over Canny. Ball segmented successfully, however difficult to identify without additional knowledge (such as tracking) due to irregular deformity caused through slow shutter speed.	Due to level of background noise segmentation, not suitable for general table tennis applications. Successful in table line detection.	No
Prewitt	Minimal background noise. Ball segmented, however boundary is broken. Table borders are broken. Net unidentifiable.	Ball clearly segmented, as are table and net boundaries. Background noise virtually disappears. Player and racket not identified	Ball segmented, yet broken and difficult to recognise due to shutter speed. Good job of removing background noise. Table and net are clear, as are the player and racket.	For medium quality video, good for ball segmentation and table/net identification. Good potential for general table tennis use.	Limited. Good for table detection and ball isolating after binary thresholding.
Roberts	Clear ball, yet the definition is susceptible to linear breakage. Little background noise. Table difficult to distinguish.	Minimal background interference. Ball clearly segmented. Table boundaries identified. Player and racket not segmented.	No background noise. Ball segmented, yet difficult to recognise with usual factors (shape, size etc.). Table and racket clearly visible.	Due to reduced levels of background noise detection, suitable for ball segmentation, but not player and racket identification.	No
Sobel	Ball perfectly segmented no background interference. However the table definition is lost and not useable.	Ball and table clearly delineated. Racket and player lost. No background noise.	Racket and player instantly recognisable. Ball segmented well, if difficult to identify due to shutter speed. Net and table clear.	Excellent for ball segmentation. Not useable for table, net or player.	Limited. Poor table detection but good for ball isolating after binary thresholding.
Zerocross	High degree of background noise. Ball lineage broken. Table edges identified, but not segmentation incongruent and not clear.	Table and net edges clearly defined. Ball successfully segmented. Player identified, racket not segmented.	Background noise visible, ball segmented clearly, if identification could be made. Table, net, player and racket edges clearly segmented with unbroken lines.	Potential use in low quality, but requires further evaluation.	Yes
Binary thresholding	Ball clearly identified. No background noise. Table and net however are below the threshold and not segmented.	Ball, net and table edges identified. Player and racket more difficult to isolate. Lighting reflections and shadow causing background noise.	Ball successfully separated from the background. Net and table less reliable with broken lineage. Player and racket below threshold.	Useful for positioning ball in all situations. Successful results based on an even light background illumination, with the ball having a substantial differential.	Yes. Applicable in combination with all other approved segmentation methods.

Process	Qualitative Results			Qualitative Suitability	Qualitative Preference?
	100-P-03	100-C-14	9999-U-01		
GMM	Ball segmented successfully. When there are other motions (background movements and shadows), these are susceptible to false positive identifications, creating additional noise and filtering requirements			Unable to detect.	Useful for ball size measurements. Other than this no benefit over Optical Flow (see next), is less reliable and more easily confused by other moving objects or changes in lighting.
Optical Flow	Ball clearly identified. No other moving Objects detected.	Ball clearly identified, along with player's arms and racket.		Arm, racket and ball clearly identified, even against similar background hue and luminance. However, success not constant from frame to frame.	Not useful for static objects, such as the table. May detect the net moving

4.7. Conclusion

The fundamental requirements of a monocular CV based ball detection algorithm must include a) detection of the ball in the majority of frames, b) the ability to segment the ball with minimal shape deformity, c) to determine its centre-point within the image, and d) to segment the ball during key match-play events. Evaluating this requires direct, repeatable comparisons of different CV algorithms, applied at different stages in the detection process, using a wide range of video sequences. To do this efficiently and accurately, it should be performed in dedicated and configurable CV comparison software to allow the investigator to perform a large number of detections and comparisons in a controlled environment. The re-use of software modules reduces the potential for software design errors and allows for ease of maintenance when moving from an experimental prototype situation through to a comparison with non CV data.

Determining the centre of the ball is important in establishing the validity of a serve, or the detection of a doubles line fault. It is during these events that the size and position of the ball can be used as a scale of reference to measure distance. This is also a desired calculation when using such a configuration for player performance assessment. As the representation of the ball in the image is often not circular, due to the relative speed of the ball and camera shutter speeds and lighting configurations, a comparison of two methods was compared: the mean of bounding box and the centroid. The sub pixel accuracy of the centroid calculation has been proven to have a greater success in measuring the diameter.

False positive detections cause complexities in monitoring the correct object of interest, particularly in a low-cost, real-time application. Tracking the ball in table tennis, for example when using the Kalman Filter, is unreliable and computationally expensive. The ball abruptly changes direction many times per second and even in free flight the ball's trajectory is non-parabolic due to induced spin. Therefore a more pragmatic solution is to use a simple Region of Interest (RoI) algorithm, estimating the boundaries of the ball's location based on mean speed and bearing derived from contiguous frames. A significant number of false positive detections occurred during the experiments due to the reflection of the ball as it entered proximity with the table surface. This was due primarily to overhead lighting, the quality of the table surface material and the low angle required by the monocular camera hardware for optimum detection of the ball. An initial feature filter based on size and shape was refined and enhanced to detect reflections. A sophisticated sequence of filtering enabled the software to distinguish between the ball and its reflection, with the most successful filters being measurements of eccentricity, convex area and intensity.

Binary thresholding is a common method in previous research for this subject in simplifying the source image data. However, the threshold value has a significant impact, not only on the success of ball detection, but also in determining the object of interest size and shape. Measurements of the diameter and area of the segmented object have found it to increase in size, as it travels across the frame. Elimination of potential causes of lighting and camera configuration has enabled the discovery of an improvement in thresholding as the ball slows down. This was particularly noticeable after an abrupt reduction in speed due to a let or table bounce. A beneficial side effect of binary thresholding has been discovered by observing the sudden appearance of a new region around the area of the net as the ball causes the net to flex and oscillate. Combining this observation with a detection of change in motion provides substantial evidence in the let event occurring. Investigations have been made to establish whether adjusting the contrast of the frame improves or simplifies ball thresholding. However, this has not proved successful and so the proposal is to not make any image adjustments to contrast.

5. CV algorithm quantitative comparison: A case study

5.1. Introduction

When developing any Computer Vision (CV) system, the choice of algorithm is critical and can affect the performance of the output (McGuinness, 2009). It is clear from current literature that there are several CV segmentation algorithms with the potential to detect a table tennis ball within a given video sequence. One algorithm suggestion (P. Wong, 2009) is to apply a binary threshold, followed by feature detection and classification using an artificial neural network. An alternative approach (Z. Z. Zhang et al., 2010), is to implement a stacked process of binary thresholding, adjacent frame differencing and feature extraction. A further suggestion (Teachabarikiti, Chalidabhongse, & Thammano, 2010), conversely, is to implement a mean shift algorithm. Common to many suggestions is an image point transform process (Morris 2004), such as reducing to a binary image, connectivity process (finding edges), followed by a filtering stage and optional final classification. All purport to successfully identify the ball and, in some cases, evaluate a specific type of match-play event (K. C. P. Wong & Dooley, 2010). The use of a thresholding algorithm with filtering may be suitable when using individual still images (P. K. C. Wong, 2008) retrieved from the International Table Tennis Federation (ITTF) Photo Gallery (ITTF, n.d.), but there is no evidence of its relative performance compared to other algorithms or across a variety of image sequences. In these examples, the recording hardware configuration and environments are not documented. There is no justification for the algorithm selected and no comparison with other algorithms presented. Therefore, a method to compare and index algorithms across a range of video sources is required.

General comparison frameworks for CV algorithms have been suggested by Zhang et al. (2008) as the issue of increased use of CV algorithms without substantial methods for assessing and comparing their accuracy is recognised. They apply sixteen image processing characteristics to the framework, measuring the properties of inter-region uniformity, inter-region disparity and semantic cues. There have also been comparisons of general segmentation techniques (Y. J. Zhang & Gerbrands, 1994) and prior to that, of thresholding techniques by Weszka (1978). An additional proposal for performance measures is made by Erdem et al. (2004) based on spatial differences of colour and motion, combined with temporal differences between the colour histograms within an image. Both comparison frameworks are useful when a) evaluating a new CV algorithm, giving it a solid comparator baseline against other algorithms and b) providing a relative comparison of two generalised algorithm imple-

mentations. Specifically, they offer good measures for the success of an algorithm when applied to a single image. In their research to provide an index comparing two or more segmentation algorithms, Pantofaru (Pantofaru, 2005) also recognised the general issue of having a selection of algorithms without a justification method for the most appropriate technique, proposing a quantitative evaluation based on segmentation correctness. A more recent low-level, generalist comparison of segmentation techniques has been made by McGuiness (McGuinness, 2009). Here McGuiness developed the *K-space video segmentation tool*, providing a standard interface for automatic comparison. However, none of these evaluation methods consider measurements of object detection during specific sports match-play events. As such, the proposal presented here differs from other CV algorithm comparison frameworks by measuring the efficacy of an algorithm at key events during a sports video sequence. Without a suitable method available, this requires the introduction of a novel measurement system, the Efficacy Metric Set (EMS), to define the set of events being measured. These key events are directly related to the general rules and moments of significance in ball sports, covering areas such as:

- a. service (or starting procedure)
- b. bounce detection
- c. ball deflection
- d. detection of the ball when approaching occlusions caused by a net, goal, or player

Focusing only on the ability to detect changes in the ball's motion at these match-critical moments, increases the pertinence of the comparison data, whilst hardware and processing requirements can be minimised.

In describing the proposed novel comparison method, this chapter begins with a qualitative appraisal of different algorithms to provide the problem context, application and broader considerations. After a description of the proposed comparison EMS, a case study is presented for the application of the EMS to table tennis. Building on the findings of Chapter 4, TRASE has been developed and enhanced for this rigorous CV algorithm comparison using the 276 videos generated in Chapter 3. The case study results are provided, demonstrating a distinction between algorithms for specific events and a justification for performing CV comparisons. The chapter concludes with a suggestion for ensuring that the algorithm comparison is made prior to sports video CV implementations. Furthermore, it is suggested that CV implementations in ball sports should not consider one algorithm for all

events, but integrate a switching mechanism between algorithms for optimum ball detection throughout the game, using the comparison data as a tool for selection.

The selection of algorithms presented here for comparison is not, and can never be conclusive or exhaustive; progress in algorithms design is continually being made. The algorithm selection does, however, provide a foundation from which to establish a method of comparison and offer, at least, a result which has been systematically evaluated and ready for comparison when future candidates are introduced. Results for this comparison¹, using a wide range of video sequences and events, show that the Horn-Schunck variant of the Optical Flow (OF) segmentation technique (Horn & Schunck 1981) has the greatest success for the *majority* of events. This not only provides appropriateness for use, but also highlights deficiencies in an algorithm. By understanding the EMS comparison results, algorithm weaknesses can be mitigated by switching to another algorithm whose properties act as a counter to the deficiencies. In this case a Canny edge detector² can be combined with the LK OF algorithm to provide a justified CV algorithm optimisation procedure for event analysis, umpire support and performance analysis.

5.2. A novel CV comparison methodology: The Efficacy Metric Set (EMS)

It has been established that not every CV based ball detection algorithm will produce the same results in ability of detection and measurement of ball metrics (Oldham et al., 2015). It is often the case that a CV algorithm is implemented to solve a problem because it *appears* to successfully segment the object in question. This may be satisfactory when the image matrix conforms to standards, such as in medical imaging. In other areas, such as sport, visual surveillance or autonomous transport, no two events are likely to be the same. However, for applications requiring integrity when analysing complex scenes, due diligence with comparative assessments should be made. Presented here is a proposal for a novel set of CV measurements, the EMS, to calculate an unbiased, Relative Efficacy Index (REI) of CV algorithms in the application of ball sports. These metrics are designed to test an algorithm for its ability to identify a ball at the *limits* of detection. Core to this proposal is the suggestion, for ball sports, a CV algorithm's efficacy must be measured by understanding

¹ Any evaluation for an optimum solution is only ever as good as the number of techniques being evaluated. It is recognised and understood that not every possible algorithm can be evaluated within TRASE. However this framework can be used to 'plug-in' new or different algorithms as required or requested.

² An edge detector is particularly useful when detecting the initial table and net location. However it is also proving successful for an improved accuracy with let detections.

the *reasons* for detecting the ball. The most common reasons for implementing CV in sports are for coaching, umpire decision support and automatic scoring or event annotation. Therefore, efficacy must be measured by processing and measuring frame sequences at the moment when these implementations require data. For sport in general, this occurs when:

- a. There is a partial occlusion, or transit, of the ball when close to a trap (net, goal, bag or basket), the racket, bat or player (**Partial occlusion**).
- b. There is a total occlusion or transit of the ball as its speed and direction are influenced by a trap, bat or player (**Total occlusion**).
- c. The game is being initiated as the ball is being served, hit or kicked from a stationary position (**Movement initiation**).
- d. The ball is at the point of a change in direction, such as a bounce, kick, deflection or return (**Direction change**).
- e. The ball hits or crosses a line or boundary of the match-play area (**Boundary detection**).

It has been proposed (Y. J. Zhang & Gerbrands, 1994) that there is a derived metric, the object of interest (OOI) location, which should be considered as a high priority when using CV for determining speed, distance and accuracy for coaching and performance analysis. Commercial ball detection solutions with multi-camera devices, such as Hawk-Eye (Hawk-Eye Innovations Ltd, 2013), are able to detect, track and determine the location of a ball in a variety of sports. These high-end solutions have the benefit of camera installations and configurations within carefully controlled stadia and auditoria facilities. Not all sports have such a significant investment. Taking the example of table tennis, the game is regularly played in multi-use halls with general-purpose lighting, on a table that is not fixed within its surroundings. Recordings of the game may use low-cost hardware and as such, consideration is given here for monocular sequences only. This introduces a set of complications if the algorithm is to be applied to single video sequences, recorded in a variety of locations with different recording hardware and with different resolutions. For this reason, the EMS is tested by applying it against a representative selection of playing and recording environments with consideration for the following:

- a. The ball is set against a range of contrasting and noisy backgrounds.

- b. The ball is travelling at high speeds, relative to a low camera shutter speed, resulting in ball shape distortion.
- c. Directional light sources are of sufficient brightness to distort the shape of the ball through self-shadowing.
- d. The ball moves across the boundary of the frame, for example when entering or leaving the aperture's field of view (FoV); this is a good indicator of partial occlusion and early detection efficacy, with consideration for lens aberration and poor peripheral lighting.

These four environmental factors directly affect the ability to determine the size and circularity of the ball's image with accuracy. Already discussed is the thresholding variability when calculating the centre point and diameter of the ball, which is important for distance, scale and speed calculations, particularly in a monocular video sequence (Section 4.4). This becomes more relevant in low-cost, single source, consumer-grade hardware configurations. Therefore, the ability to maintain as much of the ball's original shape in the image is also a valid measure of efficacy. Combining these requirements, the proposed thirteen metrics for the detection of a ball within monocular video sequences is provided in Table 5-1.

Table 5-1 Summary of CV algorithm efficacy metrics for sports video sequences

ID	Efficacy Metric	Group
EMS1	Initial frame detection (ΔIFD)	Detection
EMS2	Final frame detection (ΔFFD)	Detection
EMS3	Multiple false candidate detection (MFD)	Detection
EMS4	Total candidate detection (ΔTCD)	Detection
EMS5	False frame detection (ΔFFD)	Detection
EMS6	Non-detection (%ND)	Detection
EMS7	Undetected collision and detection frames (%UCDF)	Event
EMS8	Undetected ball trap frames (ΔUBTF)	Event
EMS9	Undetected bounce frames (ΔUBF)	Event
EMS10	Undetected during hit (ΔNDH)	Event
EMS11	Mean ball eccentricity (MBE)	Ball Metric
EMS12	Mean ball diameter (ΔMBD)	Ball Metric

ID	Efficacy Metric	Group
EMS13	Mean Processing Speed (PS)	Processing

In summary, the EMS is divided into four distinct groups: (1) detection, (2) event, (3) ball metric and (4) processing speed. Groups 1, 3 and 4 apply directly to any sport. Groups 2 and 3 also coincide with the critical events analysis required for automatic annotation, coaching and umpire support. The fourth group, processing speed, is included when evaluating the algorithm for real-time processing in a low-cost environment, as effort required by the computer to detect the ball quickly is also pertinent. These four groups form the proposal for a novel framework for which any combination of algorithms can be evaluated for efficacy of detecting a ball in sports video sequences. Where required, a supervised (manual observation) assessment is made prior to EMS evaluation to provide ground truth data from which the difference between the CV output and the human observed output may be calculated. When a number of candidate objects detected in a single frame exceed a pre-defined cut-off, then the efficacy metric is marked as a FAIL (see Appendix B for the definition of a FAIL during analysis). Detailed definitions for each metric are presented below.

ΔIFD: The difference in number of frames between supervised and unsupervised initial identification, as the ball *enters* the FoV. This provides a general indication of the algorithm's ability to detect the ball at the edges of the frame, during occlusions, at the extremes of lens distortion or in less brightly lit environments. ΔIFD provides a good comparison between edge detectors and motion (foreground) detectors. Exclude false positives.

ΔAFFD: The difference in number of frames between the supervised and unsupervised final identification as the ball *exits* the FoV. This provides a general indication of the algorithm's success to detect the ball at the edges of the frame, during occlusions, at the extremes of lens distortion or in less brightly lit environments. ΔAFFD is a similar measure to ΔIFD with additional assessment of tracking or Region of Interest (RoI) algorithms. Exclude all false positives.

MFD: The number of multiple (> 1) false candidates per frame. If this is a key factor of success for a given implementation, then it is suggested that this metric is weighted.

ΔTCD: The difference between numbers of supervised and unsupervised detections, including false positives. This metric provides an indication of the success of the algorithm to detect *potential* ball-like object within the frame and aims to mitigate the potential for over-aggressive or underperforming threshold and filtering parameters. The ΔTCD has additional significance when detecting more than one ball in the frame. Where more candidates than actual balls are detected, further filtering configuration is required to reduce false positives.

%FFD: The percentage of frames with *only* false positives. Poorly designed feature filters, or objects moving within the region of interest, are most often the cause. A low value indicates algorithm success when rejecting non-ball objects.

%ND: The percentage of frames containing a full ball without it being detected. This includes frames with only false positives. To obtain optimum results, the videos are edited to remove leading and trailing frames which do not contain the ball. A lower %ND indicates a greater efficacy of the algorithm.

ΔUCDF: The difference between the supervised and unsupervised frame count with the ball being undetected during specific collision and deflection frames. This is relevant to detect illegal deflections with other objects, such as another ball, the net, or a player.

ΔUBTF: The difference between the supervised and unsupervised frame count of undetected frames whilst the ball is occluded by a trap, such as the net, goal, bag or basket. Failure by the algorithm to detect the ball at the event is most likely due to occlusions between the camera and the trap. Frames including only false positives are recorded as an undetected frame.

ΔUBF: The difference between supervised and unsupervised frame count of undetected bounce frames (UBF). The number of frames undetected during the event of the ball colliding with a surface within the playing area, such as the table, wall, pitch or floor. Tracking failure, reflection, object deformation, or partial occlusion causes this. A FAIL indicates insufficient detections at the moment of bounce (there were no frames close to the bounce within which the ball could be detected). Loss of the ball during a bounce can adversely affect the processing capacity when using tracking or region of interest algorithms. A low UBF indicates a high ability of the algorithm to detect the ball during a bounce.

ΔUDH: The difference between the supervised and unsupervised frame count of frames without detection during ball hit by a player, bat or racket. This is a further measure of an algorithm's efficacy with occlusions and transits. The ability to detect the ball during this phase is useful for officiator support. A high ΔNDH indicates a low efficacy of the algorithm.

MBE: The mean eccentricity of the segmented ball is calculated as the ratio of the distance between the foci of the ellipse and its major axis length, multiplied by 100. The value range is between 0 and 100. The algorithm with a value closest to 0 indicates the algorithm with the greatest efficacy.

ΔMBD: The difference between the supervised and unsupervised value of the mean of the segmented ball's diameter, as a percentage of the sum of differences recorded. Along with MBE, a measure of the ball's mean diameter is useful for speed and location analysis

PS: A relative estimate of the average amount of time taken (seconds) to detect candidate object(s) in a single frame, as a percentage of the sum of average times recorded. For an accurate comparison this includes frame based pre-processing, missed frames and false detections. All results must be performed on a single computer, with minimal background processing and with the selection of video sequences stored on the same hardware device. Ideally, the performance will be evaluated for a number of events, from which an average is taken. The results should be taken as a relative estimate and not as an absolute measure. It is important to be aware that PS measurements are only a snapshot indicator, as algorithms may be improved through hardware or code efficiency.

Applying the EMS to a range of controlled and defined video sources ensures that accurate, rigorous and consistent measurements are made. The metrics are designed to have a proportionally negative ranking as the value of the metric increases. The summation of all metrics determines a new performance reference index, the Relative Efficacy Index (REI). The most effective algorithm is the one with the lowest REI³. To provide a detection comparator baseline, a supervised frame-by-frame inspection is performed by an experienced operator. Each image sequence is processed against all selected CV algorithms. In the case study presented in the rest of this chapter, TRASE has been used to automatically calculate and record the comparator metrics.

³ Although not included in this case study, it may be possible, under specific circumstance, to extend the EMS method by adding weightings to each metric, depending on the priority of individual events.

5.3. TRASE algorithm processing

With no documented evidence of an optimum table tennis CV based ball detection algorithm, an evaluation of previously suggested algorithms is made alongside a number of established and highly regarded CV processing algorithms. A substantial number of video sequences, as described in Chapter 3, are used as source data for this case study. The proposed REI is calculated and the results used to justify an optimum algorithm.

The literature review presented in Chapter 2 outlined proposals for detecting a ball in table tennis video sequences. These are grouped on pre-processing to create binary images and post-processing the results using a filter. Wong (K. C. P. Wong & Dooley, 2010) details a two-pass⁴ pre-process thresholding algorithm to successfully segment the ball. Additionally, Wang suggests using the intensity values of homogenous regions (Wang, 1998) to isolate regions with similar characteristics. These types of processing have been included for this study. Modularised algorithms to compare a number of edge detection algorithms alongside Gaussian Mixture Model (GMM) (Stauffer & Grimson, 1999) and two OF algorithms, namely Lucas Kanade (Lucas & Kanade, 1981) and Horn-Schunck (Horn & Schunck 1981), are also included. The edge detectors are Prewitt (Prewitt, 1970), Sobel Operator (Ballard & Brown, 1982), Roberts Cross (Roberts, 1965), Laplacian of Gaussian and Zero Cross (Marr & Hildreth, 1980) and Canny (J Canny, 1986). Prewitt, Sobel and Roberts have been selected, in particular, due to their similarity in design; the aim being to establish whether a distinction between these is achievable. All of the edge detectors are applied to binary conversions of the source video.

Critically for this comparison, with the difficulties in obtaining ball size metrics, thresholding the image is not necessary in the two OF algorithms and the statistical based intensity segmentation algorithm of GMM. Post-processing to identify the ball from the segmented regions is made using either the Circular Hough Transform (CHT) (Duda & Hart, 1973) or the TRASE bespoke feature filter. For intensity and other feature filter post-processing, candidate regions are overlaid with their equivalent grey-tone representations. Figure 5-1 represents an overview of the configuration of TRASE for this particular CV algorithm comparison and evaluation.

⁴ Although details are not specified it is assumed this is for low and high intensity values

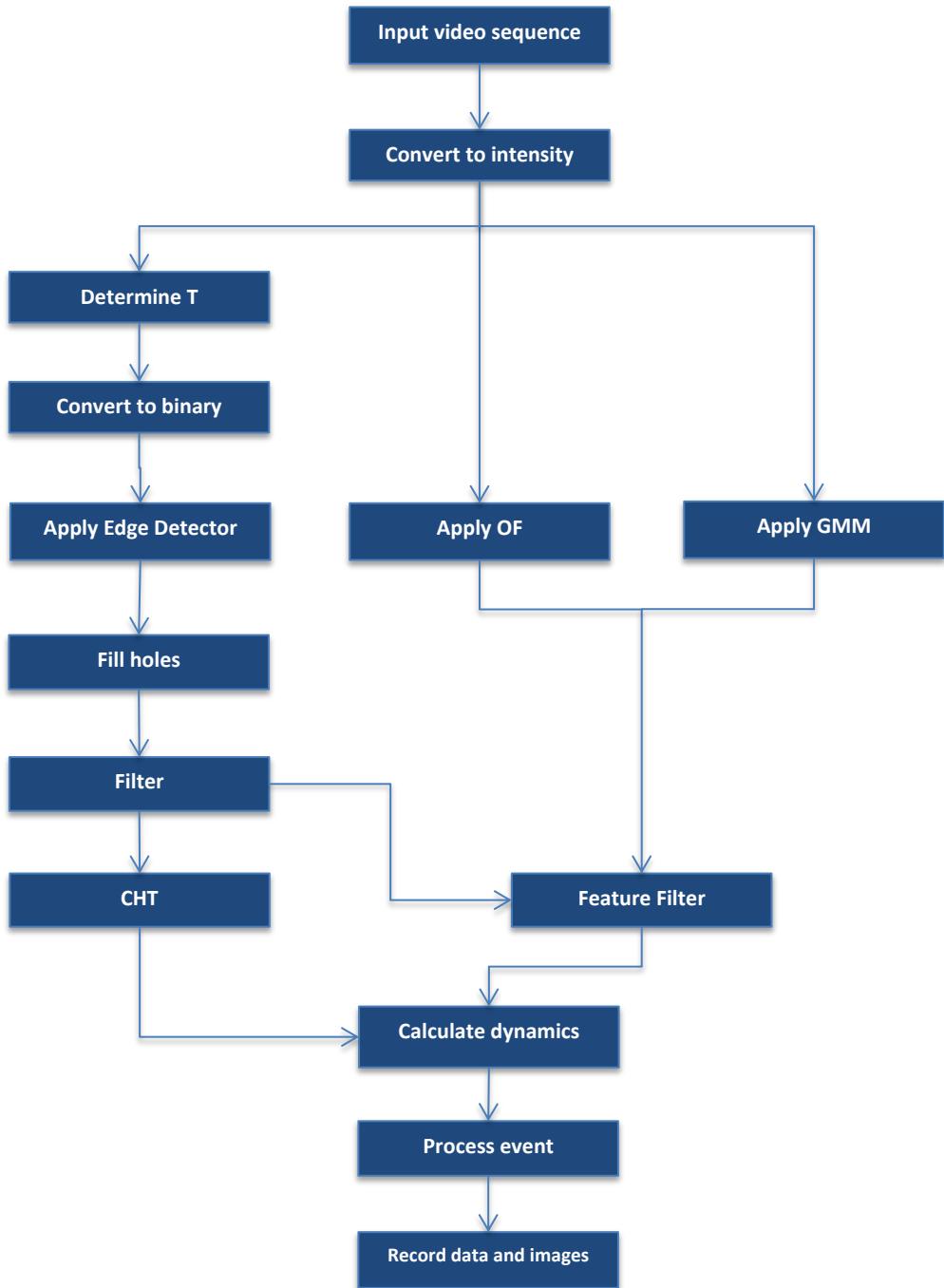


Figure 5-1 Overview of the TRASE processing stages

Multiple candidate objects detected in a single frame causes a number of problems when comparing algorithms. It has been observed that, on occasion with particular detection algorithms, no amount of careful tuning of CHT or feature filter parameters can reduce this effect. Therefore, an efficacy metric is marked as a FAIL when the number of detections reaches a pre-defined threshold. For this

study the threshold is when eight possible ball candidates have been detected in the frame (this is a tuneable parameter in the EMS). Results indicate this only occurs with edge detectors in the more complicated scenes; conversely foreground extraction techniques do not appear to be vulnerable to this problem. Although TRASE has the functionality, no background subtraction has been applied due to the lack of benefit found in Chapter 4. All algorithm parameters are saved within TRASE and re-loaded when the algorithm is processed again. It is also given that, for an effective comparison, these configuration parameters do not change during the comparison process. In total this provides fifteen algorithm combinations for comparison (see Table 5-2).

Table 5-2 Computer vision algorithms selected for comparison

ID	Type	Method	Filter
1	Edge Detection	Prewitt	CHT
2	Edge Detection	Sobel	CHT
3	Edge Detection	Roberts	CHT
4	Edge Detection	Laplacian of Gaussians (LoG)	CHT
5	Edge Detection	Zerocross (ZC)	CHT
6	Edge Detection	Prewitt	FF
7	Edge Detection	Sobel	FF
8	Edge Detection	Roberts	FF
9	Edge Detection	Laplacian of Gaussians (LoG)	FF
10	Edge Detection	Zerocross (ZC)	FF
11	Edge Detection	Canny	CHT
12	Edge Detection	Canny	FF
13	Foreground extraction	Gaussian Mixture Model	FF
14	Foreground extraction	Optical Flow (Lucas-Kanade)	FF
15	Foreground extraction	Optical Flow (Horn-Schunck)	FF

It would be a largely irrelevant and an immediately out-dated task (if it were even possible) to include all ball detection CV algorithms, with variations thereof, and present a comparison for any video data set. It is a core axiom of this thesis that an algorithm applied to one recording environment is not necessarily the most effective algorithm for any recording environment. Each source data presents its own unique images, specific challenges and desired outcomes and to ensure a chosen algo-

rithm is effective, a justification through the EMS should be given. Algorithms are created, modified and refined perpetually and as such, the introduction of additional algorithms for comparison should be performed on a regular basis, if an optimal output is required. The optimum algorithm assessment should not be made once only but continually re-assessed and evaluated, with the previously justified selection acting as the benchmark.

All videos are processed on the same computer⁵ using TRASE. As with any multi-tasking operating system, the processing power is never continuously dedicated to a single program. To minimise these effects the computer's processing capacity was, as far as possible, being used for TRASE by not being connected to the internet, not running unnecessary applications and by disabling the anti-virus software. An unbiased baseline for all metrics was generated for each video sequence through supervised, non-automated evaluation, by three human observers. The mean value was recorded and applied as the standard differentiator for each algorithm evaluation. Additionally to eliminate variations in camera hardware, all videos have been recorded using the same camera configuration CS13 (see Appendix C). Ideally, to obtain a more accurate reading each algorithm-sequence combination would be executed multiple times to get an average reading. However, for this case study where the primary evaluation is for event analysis, the comparison focus was on algorithm detection capability.

5.4. EMS analysis

Figure 5-2 through to Figure 5-24 show the EMS results for this particular case study of the comparison of fifteen algorithms. Each metric is followed, where relevant, by its corresponding failure result to demonstrate the importance of at least having successful detection as the primary consideration for any object segmentation algorithm. A discussion of the results follows the charts.

⁵ An Apple MacBook Air, dual-core i5, 8GB RAM, Intel HD Graphics 5000 with Snow Leopard OSX.

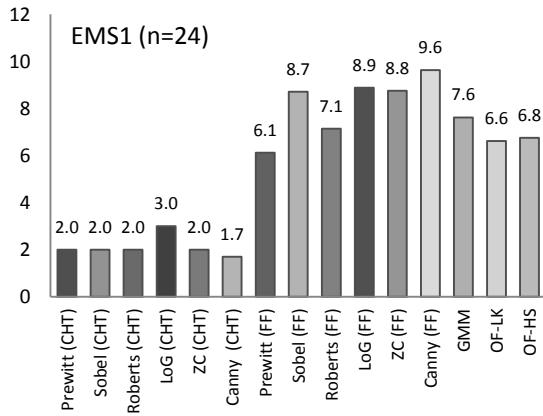


Figure 5-2 Initial frame detection ($\mu\Delta\text{IFD}$)

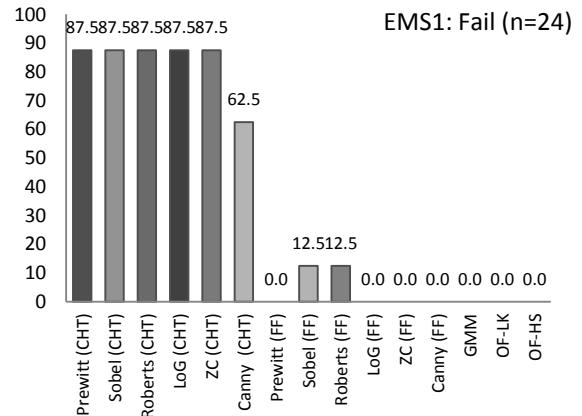


Figure 5-3 % Failure Initial frame detection (% ΔIFD)

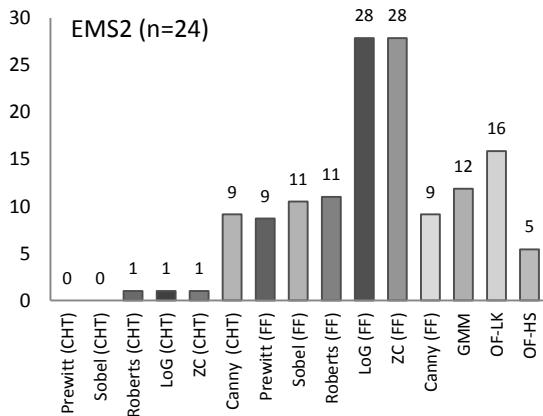


Figure 5-4 Final frame detection ($\mu\Delta\text{FFD}$)

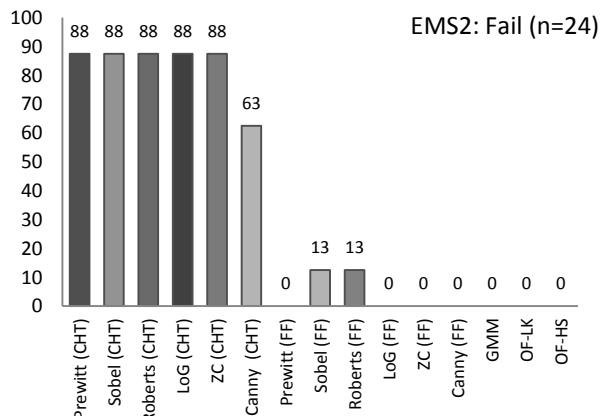


Figure 5-5 $\Delta\text{FFD FAIL}(\%)$

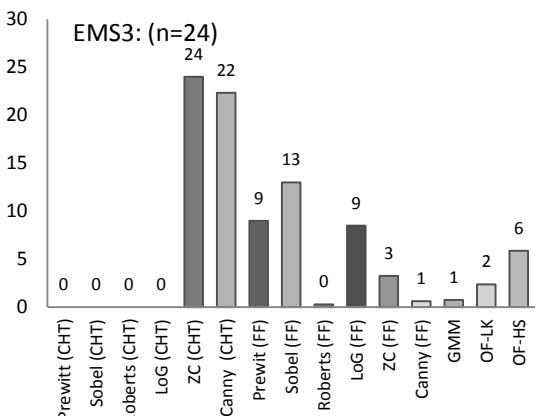


Figure 5-6 μ Multiple false candidate detection (MFD)

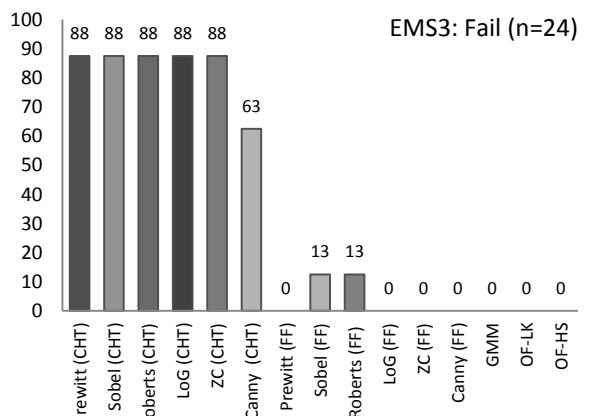


Figure 5-7 MFD FAIL (%)

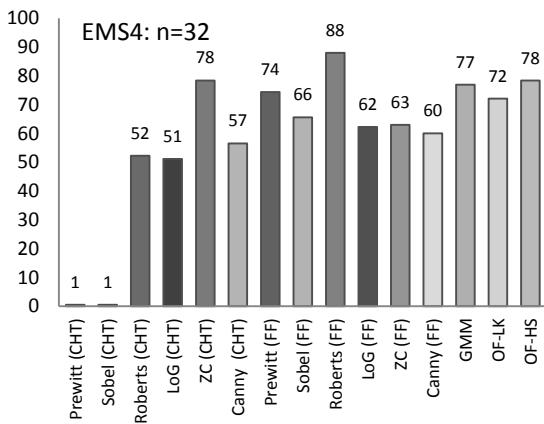
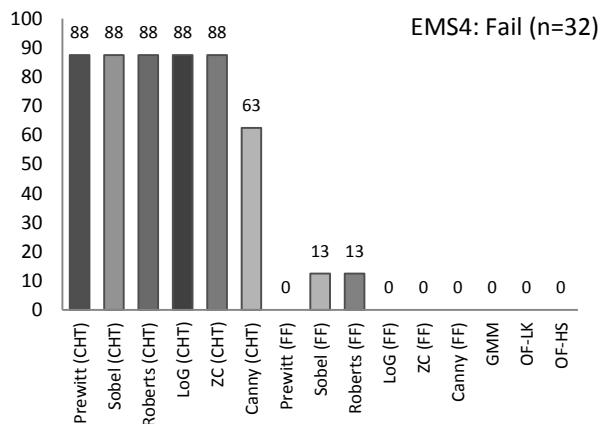
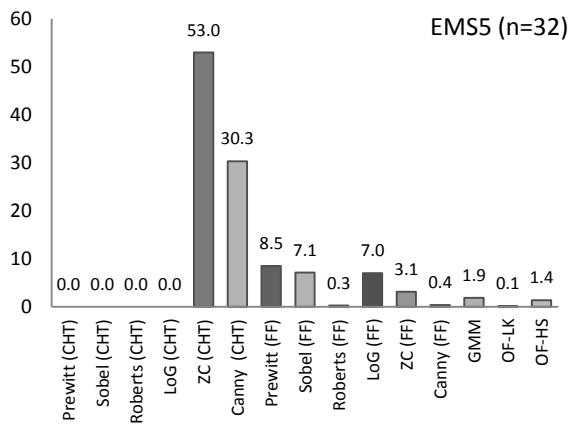

 Figure 5-8 μ Total candidate detection (ΔTCD)

 Figure 5-9 ΔTCD FAIL (%)


Figure 5-10 False frame detection (%FFD)

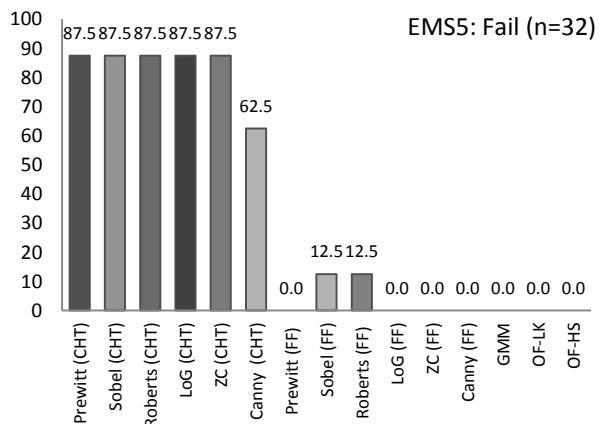


Figure 5-11 FFD FAIL (%)

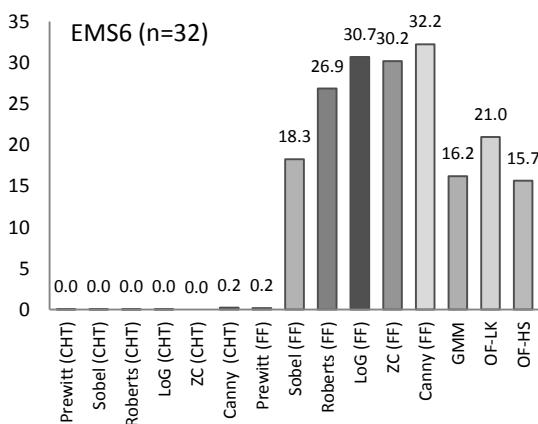


Figure 5-12 Non-detection (%ND)

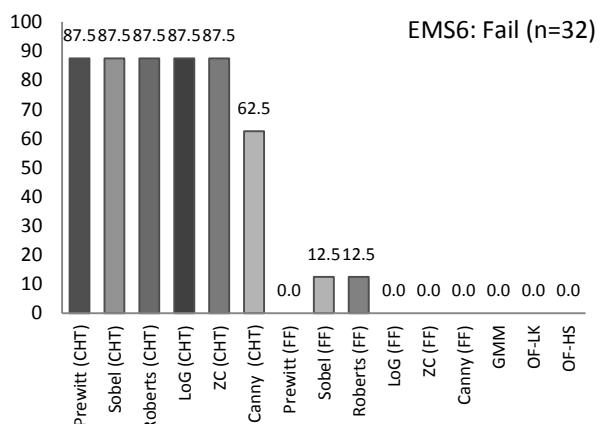


Figure 5-13 ND FAIL (%)

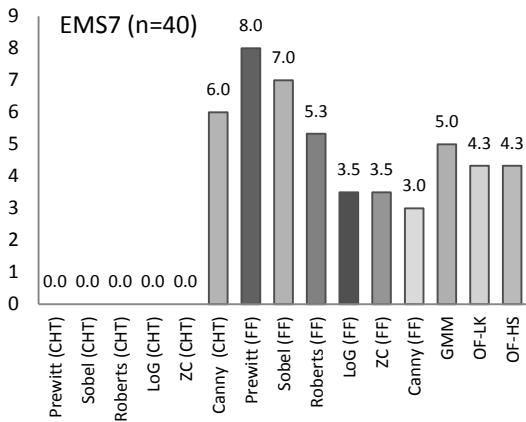


Figure 5-14 μ Undetected collision and detection frames (%UCDF)

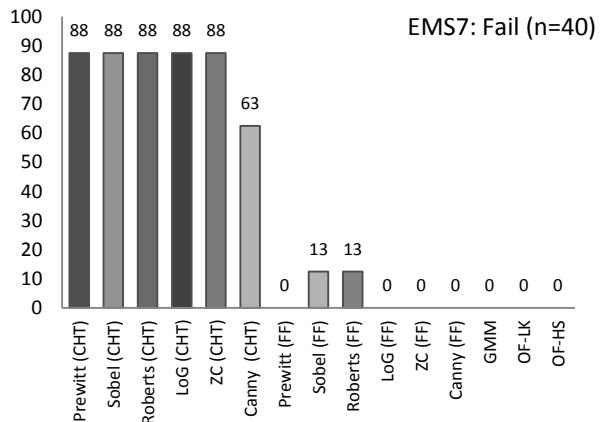


Figure 5-15 FAIL UCDF (%)

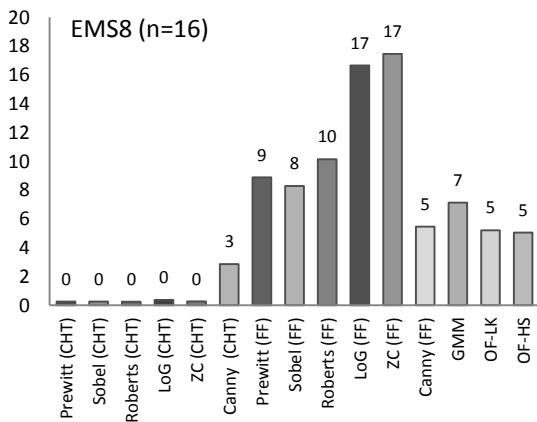


Figure 5-16 μ Undetected ball trap frames (ΔUBTF)

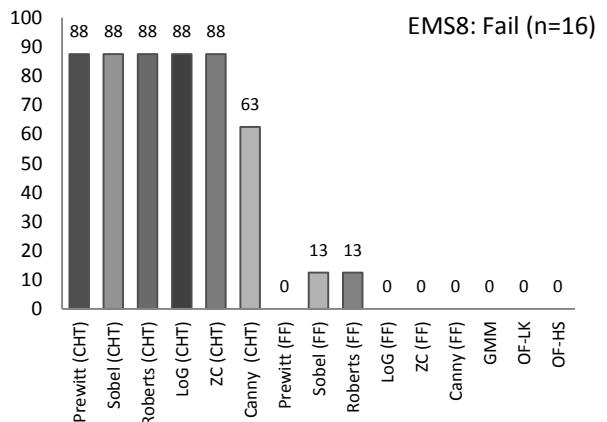


Figure 5-17 FAIL UBTF (%)

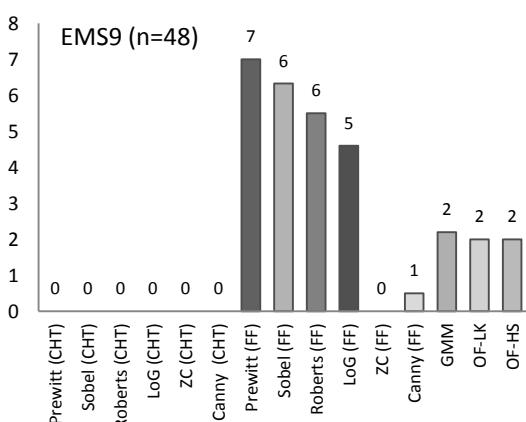


Figure 5-18 μ Undetected bounce frames (ΔUBF)

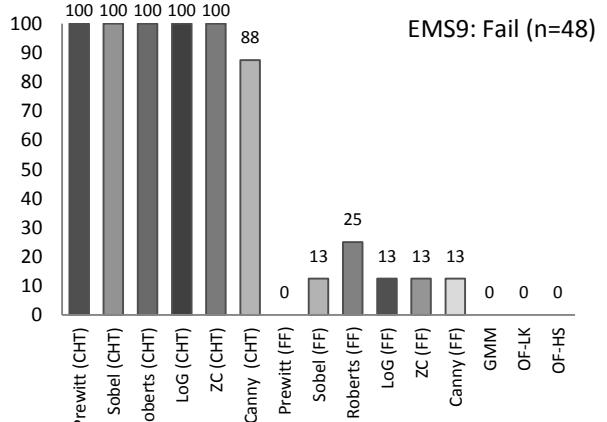


Figure 5-19 FAIL ΔUBF (%)

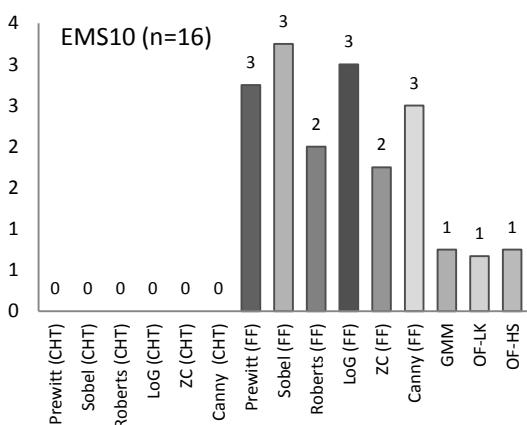


Figure 5-20 μ Undetected during hit (Δ UDH)

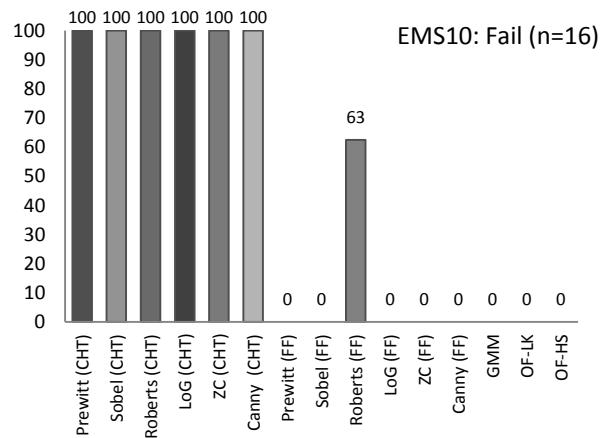


Figure 5-21 FAIL Δ UDH (%)

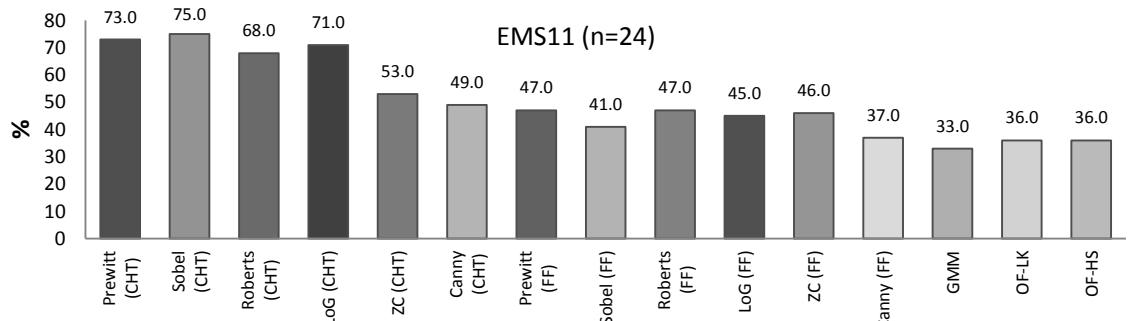


Figure 5-22 μ MBE (%)

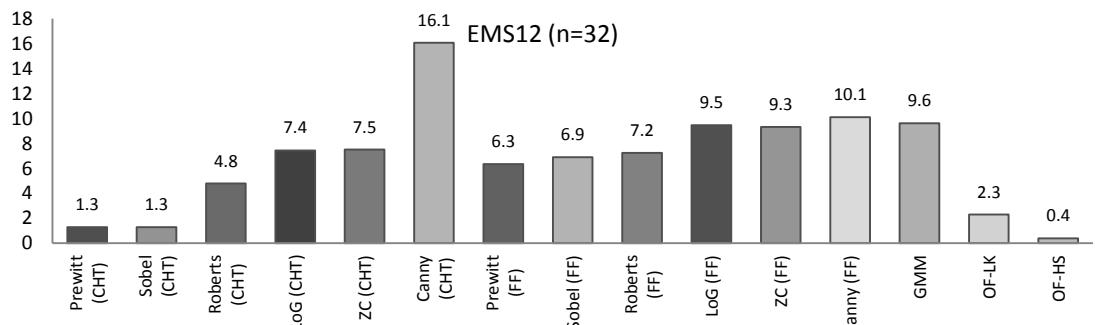


Figure 5-23 Relative $\mu\Delta$ ball diameter (Δ MBD)

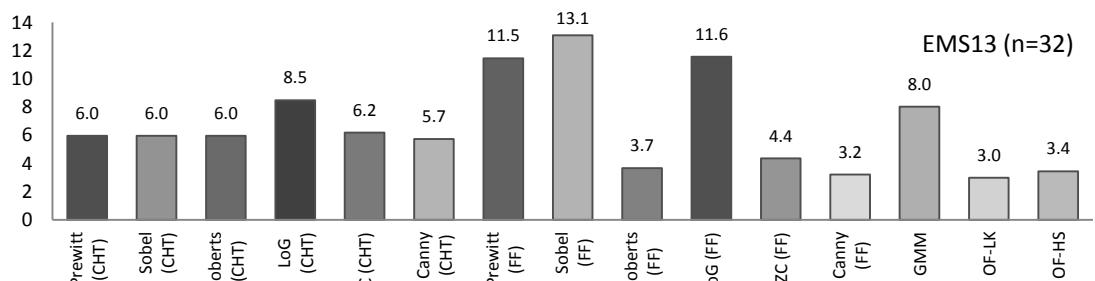


Figure 5-24 Relative μ processing speed (PS)

5.5. REI Result

The EMS was measured against fifteen algorithms and filter combinations for 276 video sequences, incorporating 161,788 individual images. An initial, supervised baseline was made for the thirteen metrics. The fifteen algorithms were then compared unsupervised within TRASE. Presented in Table 5-3 is the REI summary (a low REI indicates improved efficacy).

Table 5-3 Relative Efficacy Index Results

Method	Reference	REI
Prewitt (CHT)	1	995.573
Sobel (CHT)	2	997.573
Roberts (CHT)	3	1046.8
Laplacian of Gaussian (CHT)	5	1054.96
Zerocross (CHT)	5	1137.86
Prewitt (FF)	6	890.025
Sobel (FF)	7	193.599
Roberts (FF)	8	321.532
Laplacian of Gaussian (FF)	9	401.987
Zerocross (FF)	10	251.51
Canny (CHT)	11	231.046
Canny (FF)	12	186.434
Gaussian Mixture Model (GMM)	13	180.967
Optical Flow (Lucas-Kanade)	14	171.556
Optical Flow (Horn-Schunck)	15	163.449

Edge detectors with CHT as the single filter demonstrated to be of little value when detecting the circular ball. A marked improvement is made when the edge detectors used the bespoke feature filter. However, from the experiments within this research, results suggest that OF (Horn-Schunck) having the lowest REI is the most effective algorithm. A software developer could use this reasoning as justification for implementing this as the only CV algorithm for table tennis but this would miss the real value which the EMS measurements offer. Further examination at an event level reveals refined detail within the results which should be used to highlight strengths and weaknesses of each

algorithm for any selected event. In this case study, if we isolate the data for Δ UCDF (undetected frames during collision and deflection), we find that the Canny edge detector with bespoke feature filter performs better than all other algorithms (see Table 5-4).

Table 5-4 REI for Δ UCDF

Method	Δ UCDF
Prewitt (CHT)	100 (FAIL)
Sobel (CHT)	100 (FAIL)
Roberts (CHT)	100 (FAIL)
Laplacian of Gaussian (CHT)	100 (FAIL)
Zerocross (CHT)	100 (FAIL)
Prewitt (FF)	6
Sobel (FF)	8
Roberts (FF)	7
Laplacian of Gaussian (FF)	5.33
Zerocross (FF)	3.5
Canny (CHT)	3.5
Canny (FF)	3
Gaussian Mixture Model (GMM)	5
Optical Flow (Lucas-Kanade)	4.33
Optical Flow (Horn-Schunck)	4.33

For table tennis, this is a particularly important event, as it is the event used to detect a let. Therefore, this enables a powerful justification process for including Canny at this point in the video sequence, if more than one algorithm can be implemented. This can be extended and generalised to ensure the most appropriate algorithm is used for each individual event being processed by the CV application.

5.6. Conclusion

A novel method of measuring and comparing CV algorithms for ball sports at an event level (the REI) has been proposed for ball detection in table tennis. A case study applying the methodology using fifteen CV algorithm and filter combinations and 276 recorded events justifies the use of the Horn-Schunck OF algorithm to detect the table tennis ball. *On average* this has shown to be the most consistent and accurate across all REI events. Others, such as GMM, produced results close to those of Horn-Schunck OF, but had a greater occurrence of false positive detections. Simple edge detectors with a CHT have been shown to be unreliable, difficult to threshold and computationally more exhaustive across a wide range of scenarios, particularly when there is more than one moving object in the frame, or there is a variation in lighting. There is, however, one type of event when edge detection has shown to be beneficial: when the ball's boundary merges with another object's boundary, such as during partial occlusions or at the point of a let. At these events, functionality to momentarily switch algorithm to Canny with feature filter becomes a viable option. Therefore, it is proposed that, when selecting and comparing CV algorithms for ball detection in sports in general, allowance should be given to the notion that a single algorithm may not be the most effective for all events. Instead an optimum solution consisting of a combination of algorithms, switching between them for REI identified events, should be considered.

6. Determination of ball position and dynamics

6.1. Introduction

Accurately determining the motion of a moving table tennis ball in 3D space is non-trivial. With advanced players able to hit the ball at a mean velocity of 18.7ms^{-1} and intermediate players at 16.7ms^{-1} (Iino & Kojima, 2009), measuring the precise location mid-flight often requires high-speed multi-camera installations. The point of reference is arbitrary; the space within which it resides presents no obvious mechanism, or frame of reference, from which it can be measured. Multi camera installations, such as Hawk-Eye (Hawk-Eye Innovations Ltd, 2013) use Computer Vision (CV) processing to triangulate the location of an object hundreds of times per second; even then this location is within a tolerance range. With the limitations of detection and potential for measurement errors, claiming to know exactly where a ball is during match-play could cause naivety in overestimating the capabilities of technologies when reconstructing events (Collins & Evans, 2008).

It is unrealistic to expect a CV algorithm, using monocular video sequences from consumer-grade camera hardware, to determine the 3D position of a ball with any useful degree of certainty. However, a premise of this research hypothesis is that a single, carefully positioned low-cost camera solution can use CV algorithm processing to determine the relative location of the ball within the image and hence detect *changes* in motion (velocity, acceleration and bearing), without the need for an estimation of the ball position in 3D. Data from change detections can then be analysed to detect and classify events¹. However, it has been a finding in Chapter 4 that different CV algorithms generate different location (and ball size) data, within different tolerance ranges. In support of this, the following chapter describes the application of the novel method of measuring the accuracy of monocular CV based table tennis ball motion using a synchronised Vicon installation (Vicon Motion Systems Ltd., n.d.), as described in Chapter 3. Through the EMS (Efficacy Metric Set) analysis in Chapter 5, Horn-Schunck algorithm (Horn & Schunck 1981), based on Optical Flow (OF) techniques has demonstrated to be the most effective algorithm to detect the ball (without consideration for location) and therefore is applied throughout the analysis presented in this chapter. Results indicate that changes in motion can be detected within a 7.5% tolerance with respect to the degree of changed position.

¹ Here an event is defined as a sudden change in motion of the ball which impacts automated officiating support and scoring, or is a key metric for player performance analysis.

6.2. Estimating ball location

When using CV processing, apparent motion is described by examining sequences of frames and calculating changes in an object's location frame by frame. This research for event classification proposes only the change in motion is important; detecting a change in motion in table tennis due to a let, net, bounce or return also coincides with key scoring events. Although the 3D position is not necessary in this case, determining the relative location of the ball between frames within the image with a high degree of confidence is essential. Therefore, event recognition errors caused by the CV algorithm itself, where significant variations in the ball location between frames falsely trigger event detections, must be understood and minimised. This is the primary reason to develop a method to validate the location of the ball within the image. Additionally, there are table tennis events when a measurement of distance travelled between frames is valuable for automated officiating. These are as follows:

- a. **The service.** According to Section 2.6.2 in the International Table Tennis Federation (ITTF) Laws of Tennis Handbook (ITTF, 2013b) the server must throw the ball vertically above the palm of the hand at least 15cm during the beginning of the service. Measuring this using a supervised, manual process during match-play is not practical and is an ideal application for CV processing (P. K. C. Wong, 2008).
- b. **The centre line.** During doubles play the service must land on the correct side of the table, as defined by the line down the centre of the length of the table. The table is divided into two equal half-courts along its length by a white centre line, 3mm wide, running parallel to the length; this centre line is regarded as part of each right half-court (ITTF, 2013b). According to rule 2.6.3, when serving during a doubles match, the ball must land on the right half of the server's and then the right half of the receiver's court to be deemed legal (ITTF, 2014b). Variations in segmentation affect the ability to determine the true location of the ball relative to the line. Comparing the location of this line with the location of the ball provides additional support for umpire decision making.
- c. **Table yield.** According to Section 2.01.03 of the ITTF Laws of Table Tennis Handbook (ITTF, 2013b) the standard ball must have a uniform bounce of about 23 cm off the surface of the table when the dropped to it from a height of 30cm.

CV based assessment of ball location also provides performance measurement capability. Calculating average velocities during a shot (or a sudden velocity differential) provides the coach and players with evidence of skills and abilities. The distance the ball travels, or how close the ball travels to the net, also provides a direct and interactive approach to practising and refining repetitive shots.

6.3. Comparing CV and Vicon

The Horn-Schunck implementation of the OF algorithm has been identified in the algorithm comparison primarily for its success in detecting the ball in the *majority* of images across a range of match-play sequences. As described in Chapter 4, the location of the segmented ball varies between algorithms; therefore there is no way of knowing if this algorithm produces location data close to actual location. Therefore a comparison of the CV generated location data is presented alongside the Vicon data to provide a point of reference and ground truth. As the units of scale are different between the two data sets (pixels and millimetres), a factor for the scale of reference using the dimensions of the ball is applied to the CV output converting the measurements to millimetres. Both the CV and Vicon data sets are then smoothed by applying a Butterworth Filter (Butterworth, 1930). The change in location between frames of the moving ball is evaluated using two methods of calculating distance: Euclidean (Wolfram [1] n.d.) and Manhattan (Wolfram [2] n.d.). The delta between the CV and Vicon data is analysed.

6.3.1. Scale of reference

Table tennis images contain a number of objects of prior known dimensions, e.g. the table, net and ball. Any of these objects could be used, without investigation, to evaluate the ratio of pixels to distance. Assumptions have been made in the hypothesis which suggests the ball itself is the optimum scale of reference when calculating distance travelled or velocity, acceleration and bearing changes between frames. This is because the fixed size of the ball, as it continuously moves, provides a constantly self-optimising scale of reference from which accurate ball dynamics can be measured.

An initial step is required to establish if the image of the ball does, in fact, vary significantly across the immediate area of the table, as viewed by the fixed camera due to spherical aberration and depth of focus. Indeed if the ball size does not change significantly, then a longer dimension, such as the table length, could be more useful as a scale. As such, a number of measurements were taken using a novel positioning rig (see Figure 6-1a) developed for the purpose of measuring the ball diam-

eter around the entire play, as detailed in Chapter 3. The ball was placed on the four heights of the rig (0 cm, 15 cm, 30 cm and 45 cm), directly above a corner, edge and centre line of the table as represented by the positions in Figure 6-1b, thus defining thirty six known locations above the surface of the table. This area is the zone in which the majority of events occur and is most significant during event detection and analysis.

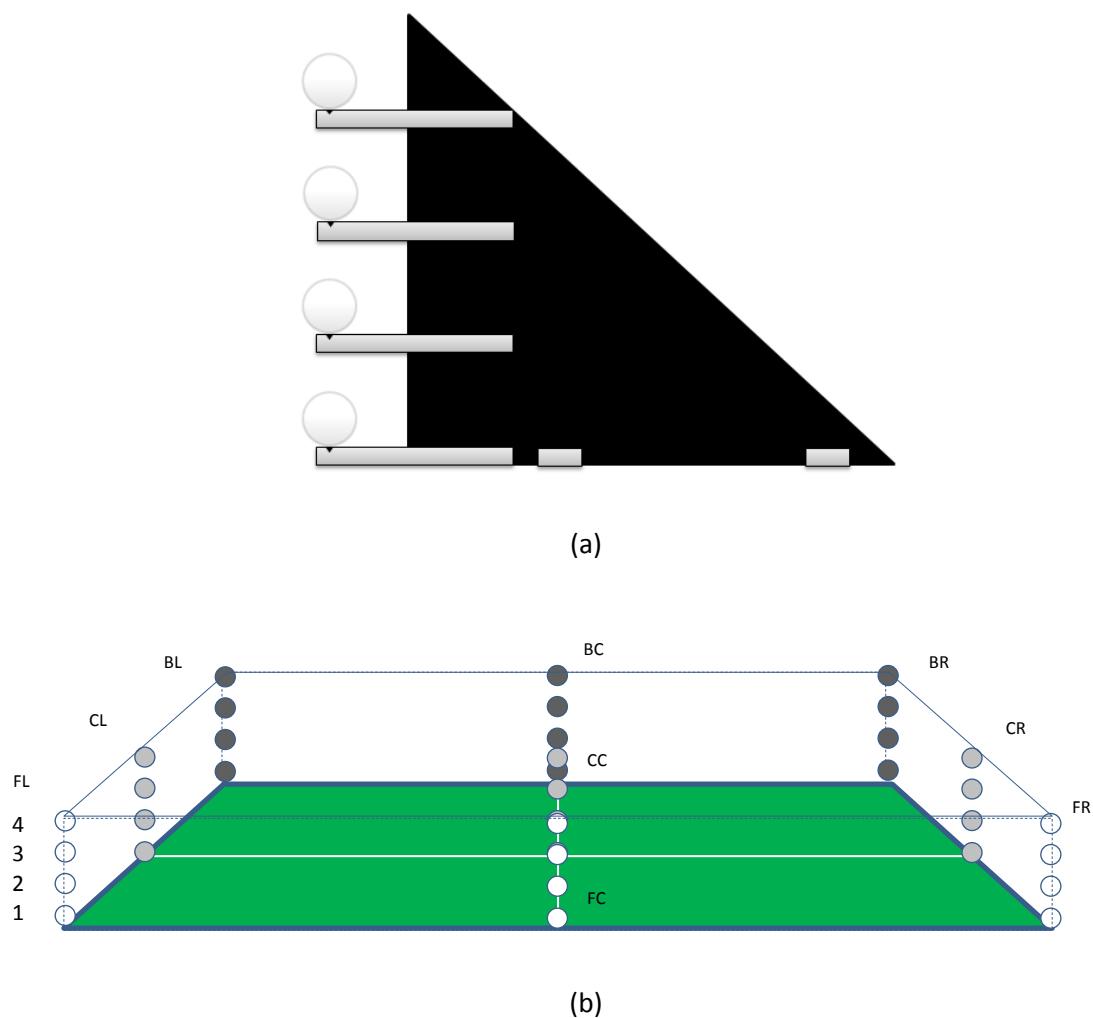


Figure 6-1 Fixed rig locations

A video sequence (containing 100 frames) of the stationary ball² was recorded in each of the locations by a fixed Photron Fastcam Ultima APX camera (Photron USA, 2014) with Sigma 24-70mm f/2.4 lens (Sigma Imaging UK, n.d.), focused at CC4 (the centre of the image) from a distance of 3 m from

² Using a stationary ball eliminates variations in the size of the ball due to spherical aberration from the discussion. Variations in ball size caused by motion blur from slow shutter speeds is the other necessary step when measuring ball size. This analysis is provided in Chapter 4.

the lens to the centre of the table, operating at a resolution of 1024x1024 pixels. Each image from the video sequence was segmented using the Canny edge detector (Canny 1986) and its mean size analysed in a bespoke CV based table tennis detection and motion comparison workbench, TRASE (see Chapter 4). Two images (Figure 6-2a and Figure 6-2b) separated by the largest depth distance (BR1 and FC4), along with the segmented image of the ball, are provided in for comparison of typical ball image size range.

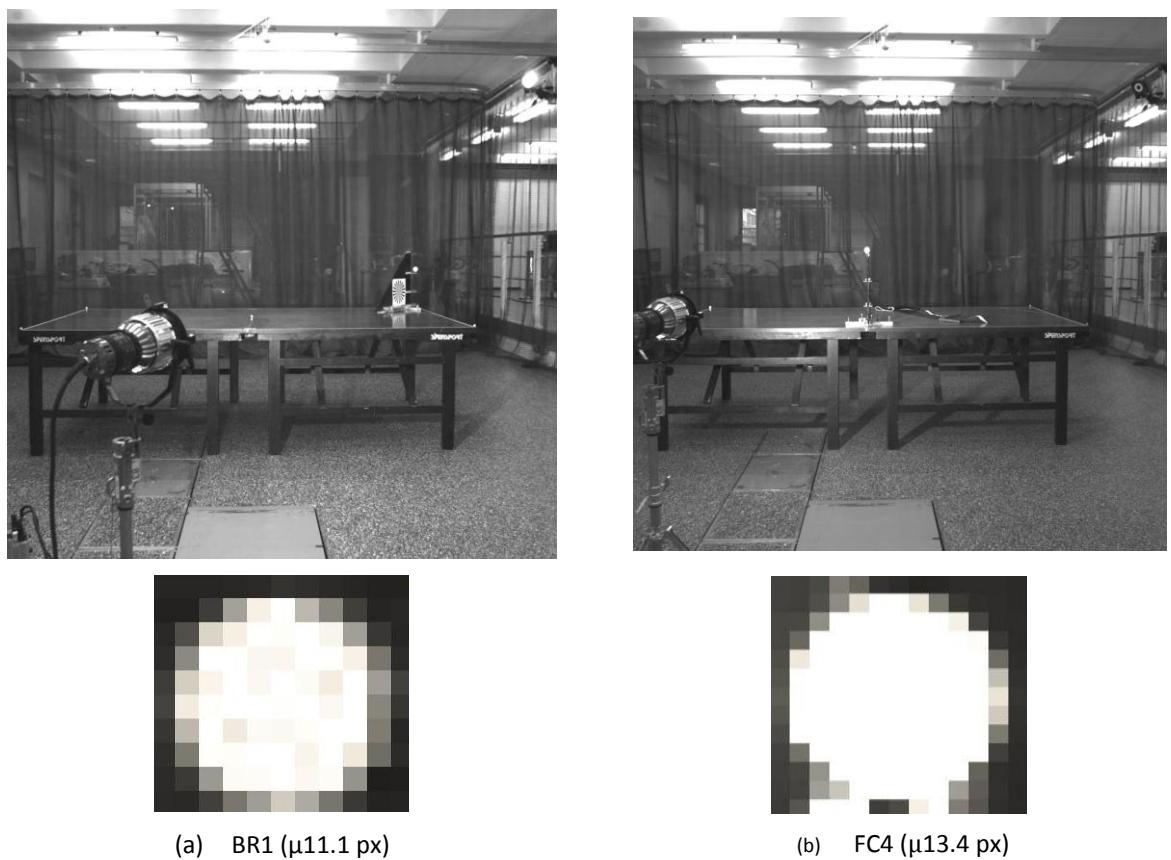


Figure 6-2 Rig based ball metric analysis

The summarised results for the ball positioned 45cm above the table are presented in Table 6-1.

Table 6-1 %Δ Ball image diameter from table centre

Position	Feature extraction diameter (px)	distance from centre of table (mm)	% change in diameter
FR4	12.2	1495.6	6.9
CR4	11.5	1370.0	12.2
BR4	11.2	1495.6	14.5
FC4	13.3	600.0	1.5
CC4	13.1	0.0	0.0
BC4	12.4	600.0	5.3
FL4	12	1495.6	8.4
CL4	11.6	1370.0	11.5
BL4	11.3	1495.6	13.7

The results indicate that, within the limits of resolution of the images, there is no discernible difference in the diameter of the ball within the 45cm height range of the rig. When viewed across the full width of the table, the largest diameter variation was found to be 2.1 pixels (11.2- 13.3 pixels). When measured from the centre of the table (the point of focus as suggested by this research), the largest increase in diameter is 1.9 pixels (FC4 to BR4, or equally BL4). At this resolution it is equivalent to 5.8mm when using a 40 mm table tennis ball. This is a substantial tolerance for calculations of distance, velocity and acceleration and is sufficient justification for continuously measuring the size of the ball as the scale of reference, self-calibrated for each frame, rather than using a fixed known dimension, such as the table length.

6.3.2. Data transformation and smoothing

Each video is processed using the same, identical CV algorithm with the ball characteristic filter settings as shown in Table 6-2:

Table 6-2 Ball characteristics for 2D position validation

Characteristic	Value
Ball size (mm)	40
RFD	10
Maximum eccentricity	0.7

Characteristic	Value
Min OOI area (px)	150
Max OOI area (px)	500
Pi Min	3
Pi Max	3.3

The centre of the ball is determined computationally using the process detailed in Chapter 4. Once the ball has been detected, extracted and its pixel location determined by TRASE, both the Vicon data set (**A**) and the CV derived data set (**B**) contain a co-ordinate matrix describing the position of the ball for each frame. The CV data is transformed as follows:

- a. The y-axis is inverted for each data point to orient the CV trajectory output with the Vicon data. This is performed by subtracting the y-value list from the maximum vertical resolution in the video sequence, equation (6-1).

$$[y]' = y_resolution_{max} - [y] \quad (6-1)$$

- b. Align the data. Vicon and CV data are aligned, setting an initial frame as the frame of reference (F_1). Both data sets are then zeroed with F_1 becoming the origin for subsequent data. Given frame F_k then this distance travelled in each co-ordinate for Vicon is:

$$\Delta A = |F_k(x_v, y_v) - F_{k-1}(x_v, y_v)| \quad (6-2)$$

and the CV data is calculated by applying a similar function (6-3).

$$\Delta B = |F_k(x_c, y_c) - F_{k-1}(x_c, y_c)| \quad (6-3)$$

- c. Convert CV data from pixels to mm, using the ball diameter as the scale of reference, (S_f) (6-4). To determine the current diameter of the ball, it is extrapolated from the average diameter of the previous two frames (configurable in the TRASE) to reduce any error in lighting, over exposure, depth and perspective variability, lens curvature and shutter speed blurring.

$$S_f \text{ (mm)} = \frac{\mu(d_k, d_{k-1})}{40} \quad (6-4)$$

- d. Both sets of data are then smoothed using a Butterworth Filter to mitigate any small errors due to resolution limitation incurred rounding.

The resulting output is two sets of data, frame coincided with equivalent units of length and displacements measured from an initial frame F_1 . To reduce error caused by resolution limitations, different automatic exposures and other influences on the ball's shape and location in the image, the CV data is smoothed by applying the Butterworth filter configured with coefficients in Table 6-3, using the calculation described by Winter (2009). A small sample of the resulting data for CV processing is presented in Table 6-4.

Table 6-3 Butterworth filter coefficients

filter coefficients	value	comments
low pass f	10	
Weight time point a_0	0.019888087	
Weight time point a_1	0.039776174	
Weight time point a_2	0.01988809	
Weight time point b_1	1.563332718	
Weight time point b_2	-0.64288507	
K_1	0.22276399	
K_2	0.024811898	
K_3	1.60310889	
Frequency (f)	250	Equivalent to FPS.
ω_c	0.15751793	Calculated as a 2 nd zero phase shift pass

Table 6-4 Data transformation sample output

Frame	Time	x pixels	y pixels	x mm (relative to start)	y mm (relative to start)	x (mm) 1 st pass	x (mm) 2 nd pass	x delta to unfiltered (mm)	y (mm) 1 st pass	y (mm) 2 nd pass	y delta to unfiltered (mm)	velocity (mm s ⁻¹)	acceleration (x10ms ⁻²)
10	-0.008	913	276	0	0	0.000	6.536	6.536	0.000	-0.161	-0.161		
11	-0.004	913	276	0	0	0.000	13.146	13.146	0.000	-0.109	-0.109	1908.07	
12	0.000	913	276	0	0	0.000	21.800	21.800	0.000	-0.014	-0.014	2429.05	1303.1
13	0.004	897	276	20	0	0.398	32.576	12.576	0.000	0.135	0.135	2950.50	1261.6
14	0.008	881	277	40	1.25	2.213	45.401	5.401	0.025	0.351	-0.899	3438.16	1147.2
15	0.012	867	276	57.5	0	6.336	60.077	2.577	0.089	0.648	0.648	3867.93	987.8
16	0.016	851	275	77.5	1.25	13.107	76.337	-1.163	0.172	1.042	-0.208	4227.84	810.4
17	0.020	836	275	96.25	1.25	22.557	93.888	-2.362	0.287	1.547	0.297	4515.08	634.5
18	0.024	819	274	117.5	2.5	34.545	112.440	-5.060	0.462	2.179	-0.321	4733.24	473.8
19	0.028	805	274	135	2.5	48.777	131.727	-3.273	0.712	2.952	0.452	4889.95	337.2
20	0.032	788	273	156.25	3.75	64.860	151.522	-4.728	1.040	3.879	0.129	4994.88	229.9
21	0.036	772	273	176.25	3.75	82.445	171.635	-4.615	1.441	4.972	1.222	5058.36	156.5
22	0.040	757	271	195	6.25	101.187	191.920	-3.080	1.933	6.239	-0.011	5090.71	119.5
23	0.044	740	269	216.25	8.75	120.749	212.271	-3.979	2.592	7.682	-1.068	5101.56	112.4
24	0.048	726	268	233.75	10	140.848	232.618	-1.132	3.481	9.303	-0.697	5099.14	118.9
25	0.052	709	268	255	10	161.235	252.920	-2.080	4.546	11.110	1.110	5089.88	127.0

6.3.3. Measuring the difference

Once the data have been converted to the same units and smoothed, the relative difference between CV and Vicon data can be evaluated. As popular methods for measuring distance in CV based feature extraction (Schulte, 2013), the Euclidean distance (a) and the Manhattan distance (b) are used. Using two methods provides supporting and comparative evidence of accuracy.

- a. **Euclidean distance.** The Euclidean distance is defined (Wolfram, n.d.-a) as the length of the shortest line segment connecting two points (A_f, B_f) (6-5):

$$d(A_f, B_f) = \sqrt{(B_1 - A_1)^2 + (B_2 - A_2)^2} \quad (6-5)$$

A_f is the Euclidean vector starting from F_1 and as direction is not required when measuring the difference between CV and Vicon outputs, this can be refined to the absolute value of the Euclidean length of the distance or displacement vector (6-6).

$$d(A_f, B_f) = | d(A_f, B_f) | \quad (6-6)$$

- b. **Manhattan Distance.** The Manhattan distance (Wolfram, n.d.-b) between two sets of data in n -dimensional space is the sum of the absolute differences of the Cartesian co-ordinates as in equation (6-7):

$$d_1(A_f, B_f) = \sum_{k=1}^n |A_{fk} - B_{fk}| \quad (6-7)$$

The difference in measurement systems is presented in Figure 6-3. The grid on the left (a) shows the Euclidean distance and is calculated as the shortest path between the two points A and B. The grid on the right (b) calculates the Manhattan distance as the sum of the co-ordinate differences, the sum of the total distances $\overrightarrow{AC} + \overrightarrow{CB}$.

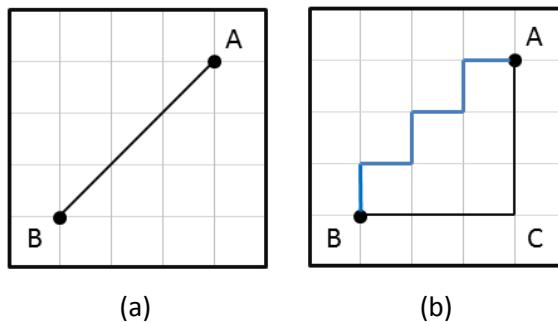


Figure 6-3 Euclidean compared to Manhattan distances

Although when applied to a 2D image from a video sequence this cannot provide an actual difference between positions of a table tennis ball over time, it does provide comparison data across data sets.

6.4. Chalk dust experiment

To support the data received from both Vicon and the CV algorithms, a third experiment was conducted to measure the distance that a ball has travelled. Reducing the chances of computational errors, this method relied on chalk dust placed on the table and a recording of a service from the Butterfly 3000 serving machine (Butterfly North America, 2013) into the freshly coated area of chalk (see Figure 6-4) was made with Vicon and a Photron Fastcam Ultima APX (Photron USA, 2014) video recorder and the Photron Fastcam. The distance (d) travelled from the edge of the table is measured using a steel rule and a comparison is made with the distance measured by the output of TRASE. Additionally a manual pixel based measurement is made in Adobe Photoshop™ Elements 9 using the initial frame in which the ball appears and the single frame at the point of contact with the table surface.

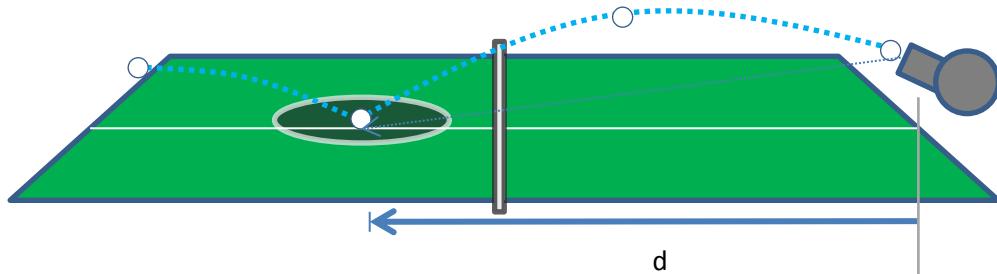


Figure 6-4 Chalk dust experiment

On processing the video from sequence 600-P-12, a complication was discovered in that at the moment the ball makes contact with the surface of the table no extraction algorithm could detect the ball due to the reduced contrast and high glare from the powder behind and underneath the ball. This is apparent on the sequence of images in Figure 6-5. The ball was detected in frames (a)-(d), then again in (f). In frame E there is insufficient contrast around the ball to enable it to be segmented successfully. However, a supervised analysis of the imprinted frame data resulted in the bounce occurring close to frame (e).

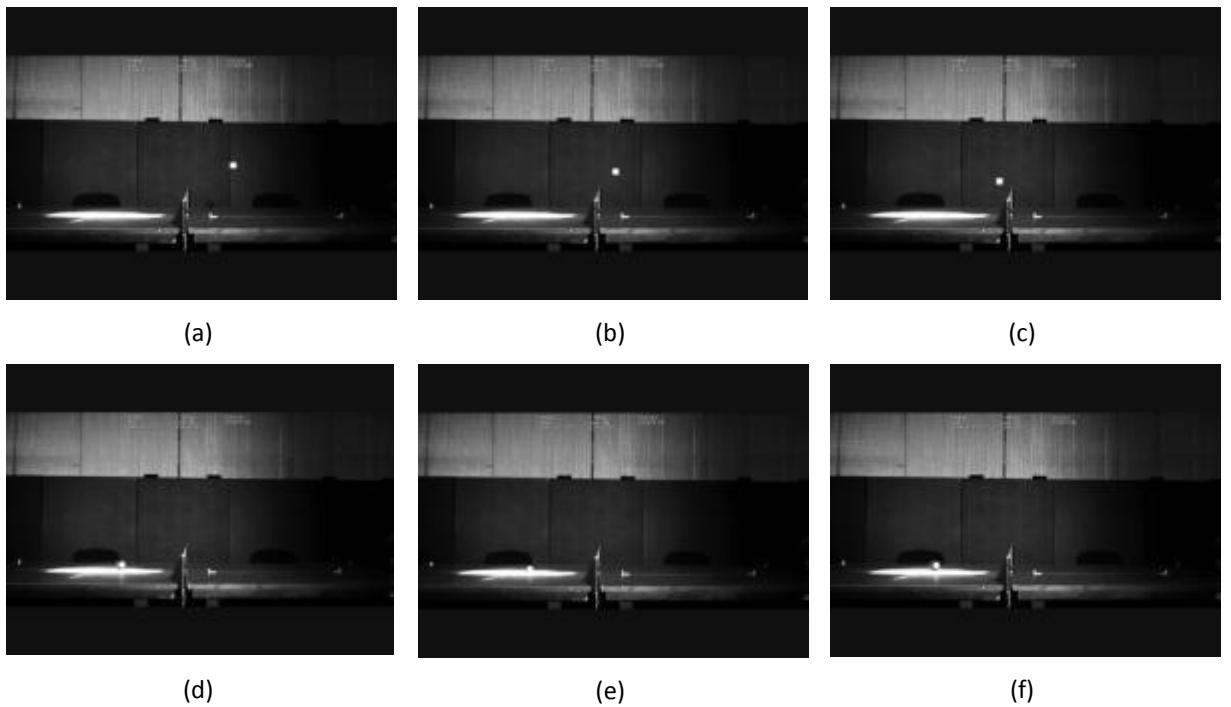


Figure 6-5 Chalk dust contrast in video sequence 600-P-12

In order to determine the distance travelled in millimetres, the mean diameter of the segmented ball (in pixels), was calculated and the distance (in pixels) between the first frame and the last successful detection prior to bounce, was recorded. The known ball diameter of 40mm was then used as the scale of reference to measure the total distance travelled.

6.5. Position results

For the validation of CV positional data, nineteen video sequences were selected. They were chosen for combining cross-table shots (for depth) with varying velocities and events such as lets, bounces and collisions. The image quality was set to enable an image of the ball with minimal motion blur. Two sequences were captured with shutter speeds of 1/500th second, the remaining seventeen at 1/250th second (see Table 6-5).

Table 6-5 Video data for position validation

Video Sequence	FPS	Shutter Speed	Number of frames	Comments
300-P-01	250	1/250 th	294	Serving machine, high speed return
300-P-06	250	1/250 th	207	Serving machine, top spin return

Video	FPS	Shutter Speed	Number	Comments
Sequence	of frames			
300-P-08	500	1/500 th	535	Serving machine, high speed return
400-P-03	250	1/500 th	261	Human, average player, return
500-P-02	250	1/250 th	310	Net collision
500-P-03	250	1/250 th	120	Human, let
500-P-08	250	1/250 th	149	Rally, including top spin, let, bounce
500-P-09	250	1/250 th	193	Rally, including 4 bounces, average players
600-P-06	500	1/500 th	380	Return with bounce
600-P-07	250	1/250 th	184	Human service
600-P-08	250	1/500 th	218	Cross table (diagonal) return
700-P-09	250	1/250 th	174	Human service
700-P-10	250	1/250 th	90	Human service with let
700-P-12	250	1/250 th	215	Human service, let
700-P-13	250	1/250 th	223	Service only (illegal)
700-P-14	250	1/250 th	176	Human service, table edge clip
700-P-15	250	1/250 th	664	Rally with smash shot
700-P-16	250	1/250 th	532	Smash shot
700-P-17	250	1/250 th	276	Human service to net

Frame rate was set to the maximum possible for the given shutter speed (250 and 500 Hz). The maximum resolution of the Fastcam is one megapixel (1MP) for all images. This has the benefit of being easily replicated in consumer-grade video recording hardware. The diameter of the ball's image ranged between 17-21 pixels across all videos, producing a pixel to mm ratio of 0.43-0.53. The results from the location calculations are shown in Table 6-6.

Table 6-6 CV position difference relative to Vicon

Video Sequence	Max ball velocity (ms^{-1}) (Vicon)	Max ball travelled between frames (mm) (Vicon)	CVMan (mm)	CVMan smoothed (mm)	CVEuc (mm)	CVEuc smoothed (mm)
300-P-01	11.4	22.8	4.7	3.9	3.6	3.0
300-P-06	10.7	4.9	4.8	3.6	3.6	2.9
300-P-08	9.7	19.4	4.6	3.8	3.7	3.3
400-P-03	4.9	9.9	4.8	3.7	3.7	2.7

Video Sequence	Max ball velocity (ms ⁻¹) (Vicon)	Max ball travelled between frames (mm) (Vicon)	CVMan (mm)	CVMan smoothed (mm)	CVEuc (mm)	CVEuc smoothed (mm)
500-P-02	10.8	21.6	4.8	3.4	3.7	2.8
500-P-03	16.8	33.6	4.7	3.4	3.6	2.7
500-P-08	12.5	25.1	4.8	3.7	3.7	3.0
500-P-09	15.1	30.2	5.0	3.7	3.5	2.9
600-P-06	19.3	38.6	4.7	3.5	3.6	2.8
600-P-07	18.6	37.3	4.7	4.1	3.6	3.2
600-P-08	19.4	38.8	4.6	3.8	3.6	3.1
700-P-09	16.2	32.5	4.6	3.8	3.6	3.0
700-P-10	9.5	19.1	4.5	3.5	3.5	2.7
700-P-12	7.7	15.3	4.9	3.7	3.7	3.0
700-P-13	8.9	17.9	4.8	3.6	3.7	3.0
700-P-14	11.1	22.1	4.8	4.1	3.7	3.3
700-P-15	12.1	24.2	4.8	3.9	3.7	3.1
700-P-16	28.7	57.5	4.8	4.1	3.7	3.3
700-P-17	2.4	4.9	4.8	3.9	3.7	3.1

The distance between the Vicon location and the CV derived location, using both the Euclidean (*CVEuc*) and Manhattan calculation (*CVMan*) is presented for smoothed and non-smoothed data. For each video sequence, the maximum ball velocity and maximum distance travelled between frames, as measured by Vicon, is provided for comparison. Figure 6-6 highlights the difference between the smoothed and non-smoothed data for the Euclidean calculations. Here the improvement in reducing the mean distance between Vicon data and CV data is clearly visible, with an average displacement reduction of 18.3%.

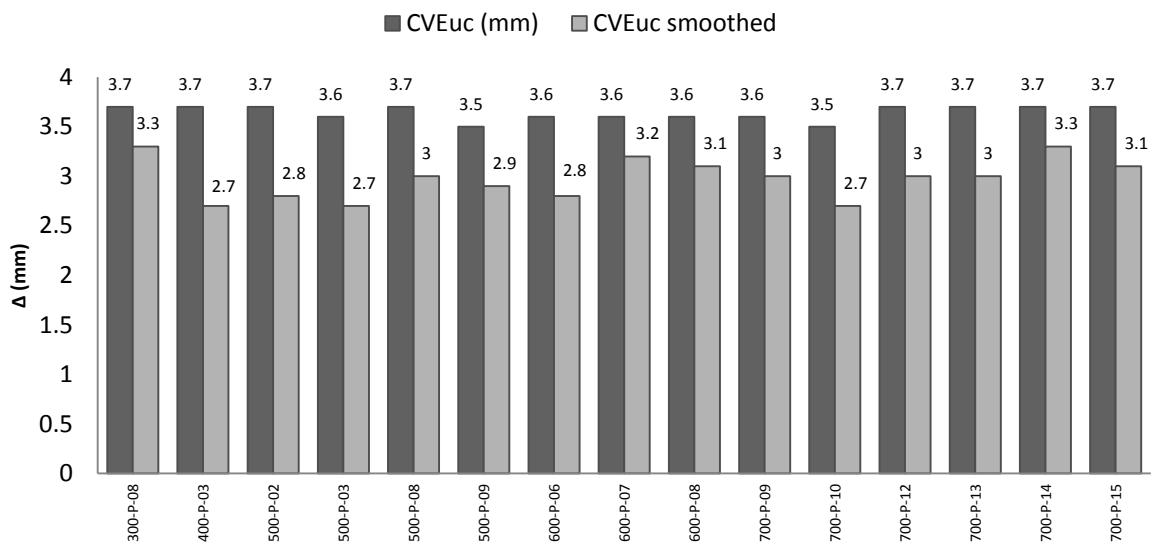


Figure 6-6 Δdisplacement between CV and Vicon (Euclidean)

The mean of the differences between Vicon and CV data for all video sequences, by component vectors for the x-axis and y-axis across the selection of nineteen video sequences, for both smoothed and raw data is presented in Table 6-7 below, together with the mean Manhattan and Euclidean distance calculations.

 Table 6-7 Summary results for the Δ displacement between CV and Vicon

	Non-smoothed (mm)	Smoothed (mm)
Mean Δx	2.2	1.1
Mean Δy	2.5	2.6
Manhattan	4.8	3.8
Euclidean	3.6	3.0

During these experiments the mean distance (measured using the Euclidean calculation) between CV and Vicon, without smoothing, is within 9% of the diameter of the 40mm ball. With smoothing this improves to 7.5%. The velocity of the ball and the distance travelled between frames does not significantly impact the margin of error. There is an increase in error range for the vertical axis compared to the horizontal difference. The indication is that the increase is due to the increased motion of the ball as it travels further over a shorted period of time in the x-axis. Distortion of the ball also occurs

more prominently along this axis, which would have an impact on the accuracy of determining the centre of the ball. Further investigation with a range of velocities in both axes is necessary to confirm this assumption.

When reviewing the differences between the smoothed data (Figure 6-7a) and the non-smoothed data (Figure 6-7b) for an individual video sequence output data (in this example, 700-P-15, N=664), the results for a single vector component (y-axis) indicate that there is a larger error range with reduced variance for the smoothed data. If CV data smoothing is to be applied for table tennis officiating decision support or player performance analysis, it is suggested that a justification for its use would have to be made and any tolerances detailed. Alternatively, this could be a tuneable parameter to be selected for difference scenarios and requirements.

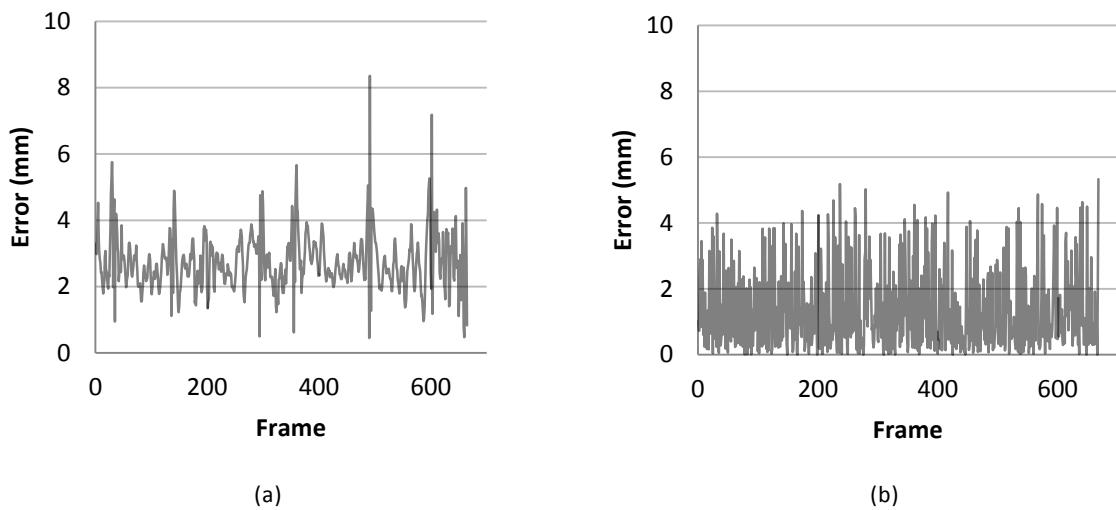


Figure 6-7 Comparison of (a) non-smoothed and (b) Butterworth smoothed data.

In the chalk dust experiment, the resulting output for the distance travelled by the ball was 1852.2mm. The actual measured distance by ruler is 1815mm, representing a difference of 3%. This was also then compared to the Vicon data and a manually supervised measurement of pixels across the image between the two points of interest at several time points during the ball's flight. Table 6-8 directly compares the three measured outputs. It can be seen that using Photoshop™ and aligning the grid to determine the number of pixels horizontally (705) from the serving machine has provided the least accurate result.

Table 6-8 Summary of bounce distances (600-P-12)

Ruler (mm)	TRASE (mm)	Vicon (mm)	Photoshop™ (mm)
1815	1852.2	1822.3	1729.9 (705px)

This demonstrates that the measurement of longer distances over many frames is within a similar range to that found when analysing mean distances between sequential frames. It also supports the use of Vicon as an external, non-intrusive solution for validating the 2D ball motion generated from CV based processing.

6.6. Calculating ball dynamics

Given ball data for 2D distances travelled between frames and the frame frequency, estimates of the components of the velocity and acceleration in each of the two available dimensions are calculated for both sets of data (Vicon³ and CV). During this process the following assumptions are made:

1. x and y are the derived ball location co-ordinates, measured in millimetres, and t is the time interval between subsequent frames, measured in seconds,
2. The time interval step is the shutter speed for the camera; no validation of the shutter speed has been made and is assumed to be accurate. For a true comparison of the Vicon and the CV data, the interval steps of the two systems are equal,
3. Each CV data set generated for each video sequence is processed identically using TRASE and the output stored as a CSV file for further analysis,

The following sections describe the process and results for calculating and validating velocity, acceleration and bearing data from Vicon and CV outputs.

6.6.1. Velocity

Developing a method to estimate the true velocity of the ball with a monocular device represents difficulties due to the projection from the 3D real world to 2D images. However, according to the hypothesis this is not necessary and there is symbolic velocity information which, when retrieved from the 2D images, can be used in event classification. There are several key facts we can assume for this analysis: a) in table tennis the primary direction of travel is across the longest axis of play, b)

³ Although available, the Vicon depth (Z-axis) data has been discarded for these calculations.

the camera is placed parallel to this axis, as defined in Chapter 3, c) the camera is set to use a fixed shutter speed (in this example 250 Hz) and d) the table tennis ball used is always 40mm in diameter. With this information, the proposal is that a representation of sudden changes in velocity can be used for event classification. The locations of a moving ball across two frames is represented within the image by pixel coordinates (x_1, y_1) and (x_2, y_2) , see Figure 6-8.

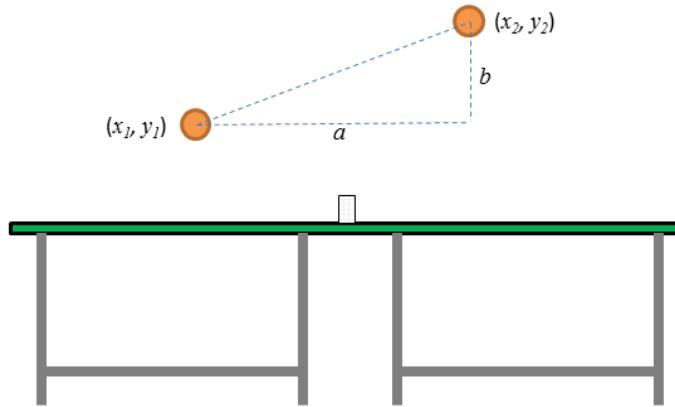


Figure 6-8 Velocity calculation

Given $a = |x_1 - x_2|$ and $b = |y_1 - y_2|$ the estimated velocity (v) in across two frames is calculated as follows (6-8):

$$v = \left(\frac{4FPS}{100d} \right) \sqrt{a^2 + b^2} \text{ ms}^{-1} \quad (6-8)$$

Where d is the diameter of the ball (px) and FPS is the number of frames per second (Hz). Confirmation of accuracy of the calculated velocity is provided by corresponding Vicon data combined with Foresight results.

6.6.2. Acceleration

Acceleration (a_k) is calculated from the change in velocity in the current frame (f_k) and the velocity in the previous frame (f_{k-1}), equation (6-9), where t is the time interval between each frame.

$$a_k = \left(\frac{vf_k - vf_{k-1}}{2t} \right) \text{ ms}^{-2} \quad (6-9)$$

As such, the position of the ball in the previous three frames is required. A measurement of the current acceleration is continually recorded.

6.6.3. Bearing

Monitoring the direction of travel provides an indication of changes of motion. For example, if the ball hits the top of the net (during a let) a deviation in the bearing should be detected, providing supporting evidence for a potential illegal service. The bearing (b_k) is calculated using the inverse tangent of the differences in component vectors, measuring the angle subtended between the location of the ball in the current frame (f_k) and the location in the previous frame (f_{k-1}) as in equation (6-10), where $\text{atan2}(y, x)$ is the angle in radians between the positive x -axis of the plane of the table and the point given by the coordinates (x, y) of the position of ball in the image.

$$b_k = \left| \left(90 - \frac{\text{atan2}(y_{f_{k-1}} - y_{f_k}, x_{f_{k-1}} - x_{f_k}) * 180}{\pi} \right) \right| (\text{mod } 360) \quad (6-10)$$

A running average is continuously compared to the bearing in the current frame. Significant variation between the two suggests a sudden change in direction. Y-axis co-ordinates are mapped to zero based bottom-left notation, where an increase in the y -value corresponds to movement towards the top of the image. The bearing has a value of 0° when $x = 0$ and $y > 0$ and measured clockwise (a vertical line on the image) from which all quadrants describing the ball's bearing are denoted (see Figure 6-9).

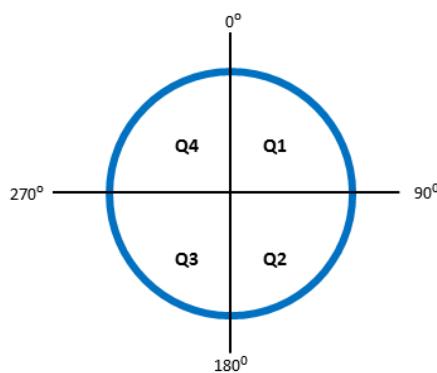


Figure 6-9 Quadrant reference for bearing analysis

This allows ease of visualisation of the ball's motion, where the direction of acceleration of the ball due to the effect of gravity has a bearing of 180° and the table is placed within the field of view (FoV) along the 90-270° plane.

6.7. Dynamics results

The three characteristics of the motion of the ball (velocity, acceleration and bearing) have been recorded for all video. Reviewing the sample data in 700-P-17 ($n=48$), a comparison of velocity and acceleration (Figure 6-10 and Figure 6-11) show a strong correlation. Significant is a sudden change in velocity trend as the ball strikes the net at frame 25. A distinct reduction in velocity is recorded for both CV and Vicon data.

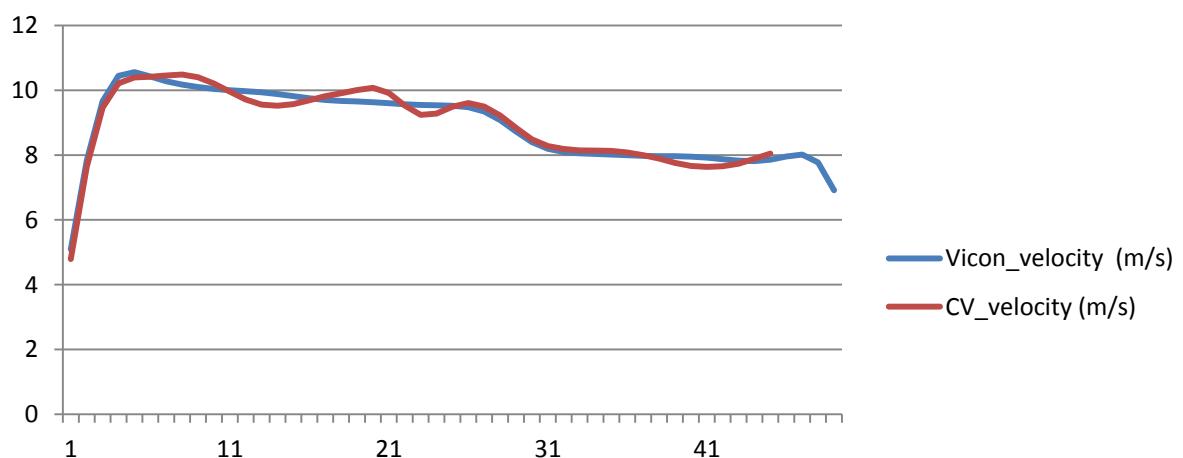


Figure 6-10 Comparison of CV and Vicon Velocity for 700-P-17

The CV determined acceleration data follows general trends when compared to the Vicon data. However in this case, the CV data is more susceptible to noise. The sudden increase in acceleration detected by Vicon is also apparent with the CV data.

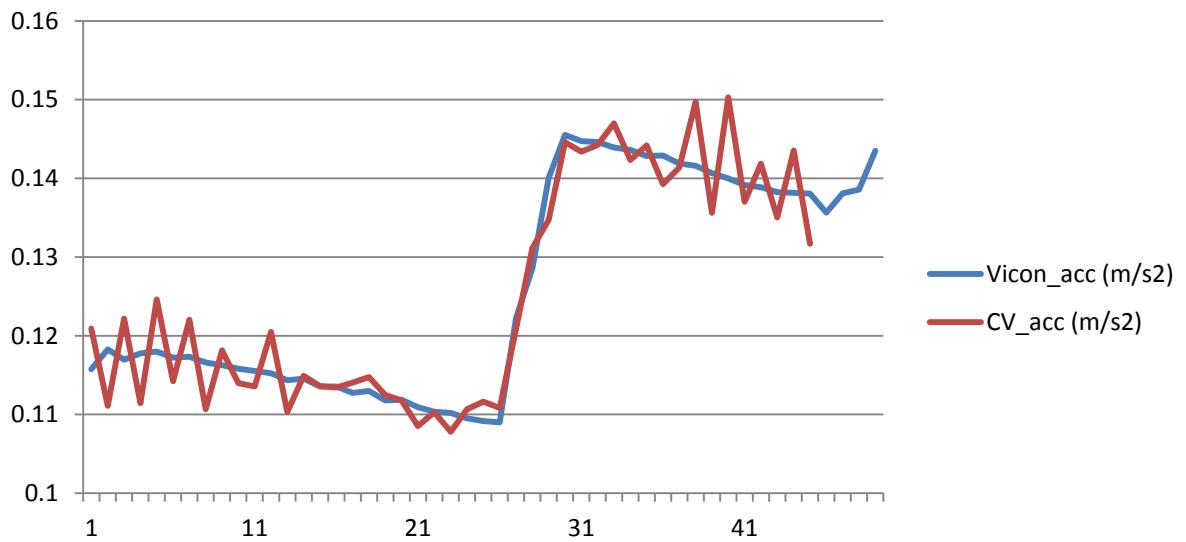


Figure 6-11 Comparison of CV and Vicon acceleration for 700-P-17

The bearing results (Figure 6-12) show a similar output to acceleration. The moment of impact with the net is clearly visible, with the ball increasing its bearing as it strikes the net. This deflection from almost 80° to approximately 100° was one of the smallest deviations observed with the human eye in all the experiments, yet was clearly visible in the chart.

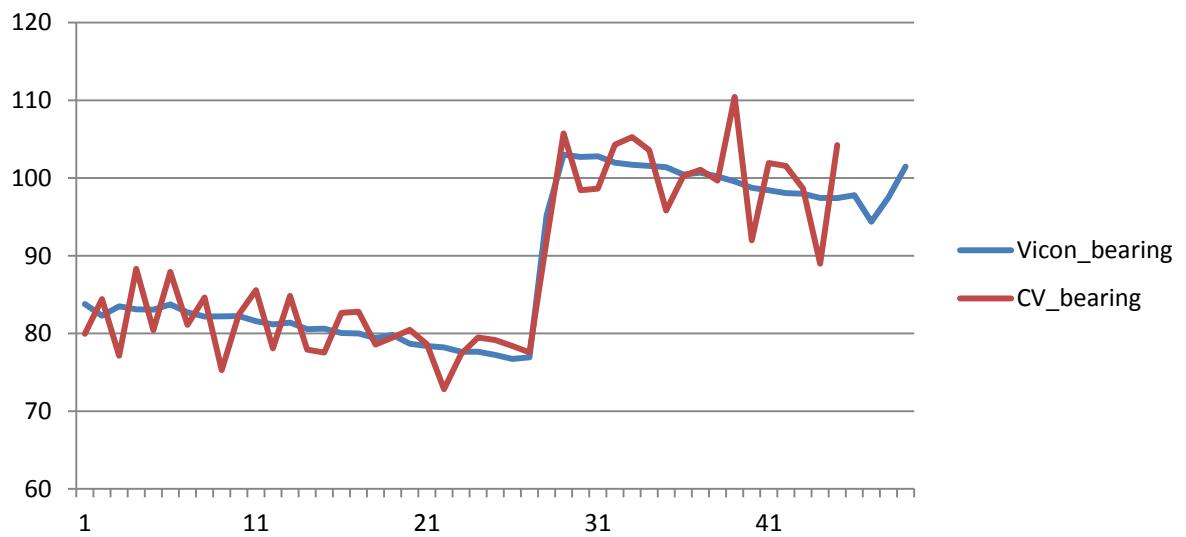


Figure 6-12 Comparison of CV and Vicon bearing data for 700-P-17

Figure 6-13 highlights the statistical difference in greater detail. The relative minor differences in the data would indicate the potential for these three characteristics of motion change to be implemented for event detection.

Figure 6-13 Statistical comparison of CV and Vicon motion characteristics (n=48)

	μv	σv	$\sigma^2 v$	μa	σa	$\sigma^2 a$	$\mu \theta_k$	$\sigma \theta_k$	$\sigma^2 \theta_k$
Vicon	8.89	1.12	1.25	0.1	0.013	0.005	91.5	6.6	42.1
CV	8.87	1.18	1.39	0.1	0.013	0.01	89.0	12.3	150.7

Further research in smoothing the data input could provide a more accurate interpretation of the results.

6.8. Conclusion

It has been presented throughout this research that variations found when determining ball metrics, such as eccentricity and diameter, make it complex to find the centre point of the ball. Finding this point is necessary when establishing the accuracy of a given CV algorithm. Once the centre of the ball can be determined with consistent accuracy, then measurements of distance (and ultimately velocity, acceleration and bearing) can be found. Developing findings in the comparison of CV algorithms (Chapter 5) the Horn-Schunck method of Optical Flow has been used to investigate the accuracy of measurement outputs. To convert from pixels to mm, the proposal here is to use the diameter of the ball as the scale of reference in every frame. Results demonstrate that a 2D CV algorithm can measure distances when using the sum of the mean distance travelled across multiple frames, if the camera is positioned relative to the table to provide image projection representing two dimensions. The minimum mean error between frames calculated by CV and Vicon is 3.01mm, across nineteen video sequences and events.

Evidence for the estimated 2D velocity indicates a good correlation between CV and Vicon, suggesting velocity, acceleration and bearing data from CV can be used in the analysis of event detection. However, there is potential for improvement with modifications to smoothing and calculations of the motion features, as small fluctuations in ball positioning information could cause the misinterpretation of motion changes. The data provided here also show the suitability of user infra-red based technologies to measure location information for table tennis balls; although the modifications made

to the ball prevent it from being used in match play. Therefore, this type of validation is for early proof of concept when validating a CV algorithm, rather than in general calibrating of individual installations.

It has been observed that there is a greater variation of Δ displacement between CV and Vicon data in the x-axis, compared to the y-axis. Early indications are that this is due to the increased motion of the ball in the x-axis. However this necessitates additional study. Further research is required to measure the performance when applying this method to lower FPS. With a decrease in frequency of data of the ball metrics, the accuracy of the measurements may decrease. However, the camera settings used here are within the range of latest consumer-grade camera hardware.

7. Event classification

7.1. Introduction

Detecting a table tennis ball using CV algorithms from monocular video sequences has been demonstrated in Chapters 4 and 5. A method for smoothing data and then calculating and validating the location of the centre of the ball has been proposed in Chapter 6. Sequencing these locations over time provides the information necessary to describe the ball's motion. The limitations of 2D positional data have been mitigated by careful positioning of the video camera (Chapter 3) and the proposed approach for event analysis is to consider, where possible, only the *changes* in object motion. It is the hypothesis of this study that sufficient information is contained within these changes for the automatic detection and classification of some key match-play events, for the ultimate use in a rules based officiating decision support system.

Building on work previously developed for event classification using 2D motion in snooker (Rea et al. 2004), it is proposed that key table tennis events are given a formal definition based on expected changes in motion combined with their pre and post event states stored in a finite-state machine (FSM). Motion is analysed using a running mean of velocity, acceleration and bearing. These three characteristics alone are sufficient for the creation of semantic information required for distinct event identification. A continuous assessment within the Table Tennis Recording Analysis Software Environment (TRASE) of abrupt motion changes triggers an event state check. If the ball's motion matches that of the event description in the FSM then a message is sent to the event engine for evaluation and event state update. This chapter details the definitions for seven match-play event types and provides evidence of the results of event detection when applied to the video sequences generated in Chapter 3, also using the Optical Flow (Horn-Schunck) CV algorithm as justified in Chapter 5 by the Relative Efficacy Index (REI). Results indicate a high degree of success for event detection using a monocular recording. Occlusions during service and table edge are the only cause of event detection failure and methods to mitigate this are provided throughout.

7.2. The proposed classifier

A review of previous research for the automatic classification of table tennis match-play events using CV derived outputs presents no evidence of an available solution. However, during event detection

in snooker carried out by Denman et al. (Denman, H., Rea, N., Kokaram, 2003), events were detected by constantly reviewing semantic information provided by explicit object tracking. Velocity and position graphs show the bounces on the snooker table in the vicinity of the cushion, combining detected motion of the ball with knowledge of game rules. The 2D nature of snooker presents similarities with the 2D representation used here with monocular table tennis video recordings. However with snooker the change in velocity was the only criteria considered. To provide a detailed assessment of events in table tennis, it is proposed that an event is detected by monitoring changes in velocity, acceleration and bearing. A demonstration of snooker ball collision detection has been made by Rea et al. (2004) through the identification of changes in the ratio between the current white ball velocity and the mean previous velocity; in this procedure a ratio threshold of 0.5 was used which would satisfy the need for the physics involved in snooker.

The event detection proposal presented here for table tennis is to take the approach used in snooker but applying improvements through the replacement of a fixed threshold with experimental results unique to individual table tennis events. This proposal also introduces the supporting data of acceleration and bearing measurements. The temporal motion of the ball at a given moment, captured by a frame, is compared to a running average of its component motion. The mean components of velocity (v_p), acceleration (a_p) and bearing (b_p) from the previous three frames are determined using equations (7-1), (7-2) and (7-3) respectively, where k is the current frame identifier.

$$v_p = \frac{1}{3} \left(\sum_{i=k-3}^{i=k-1} v_i \right) \quad (7-1)$$

$$a_p = \frac{1}{3} \left(\sum_{i=k-3}^{i=k-1} a_i \right) \quad (7-2)$$

$$b_p = \frac{1}{3} \left(\sum_{i=k-3}^{i=k-1} b_i \right) \quad (7-3)$$

The characteristics are evaluated and, if the result for any characteristic is a logical TRUE, this triggers the confirmation check followed by the classification of the event. Running averages are reset and the outputs of the previous event become the inputs of the new event. Any missing frame data is not included in the running average so the mean is unaffected by occlusions or non-detection.

Events do not occur across all areas of the frame. During any sequence of match-play, there are many frames which may be considered *inactive*. Inactive frames occur when the ball is in mid-flight, between the players, above and away from the table. The concept of applying inactive frames becomes particularly important to improve processing speeds with higher frame rates, a requirement when ensuring event data is captured. Of equal importance is a focus on hotspot zones within the frame. These are areas which are more likely to have the occurrence of a particular type of event, such as a let or net collision within the vicinity of the location of the net. These hotspot zones are represented in Figure 7-1. Whenever the ball is outside of the zones no event classification is necessary.

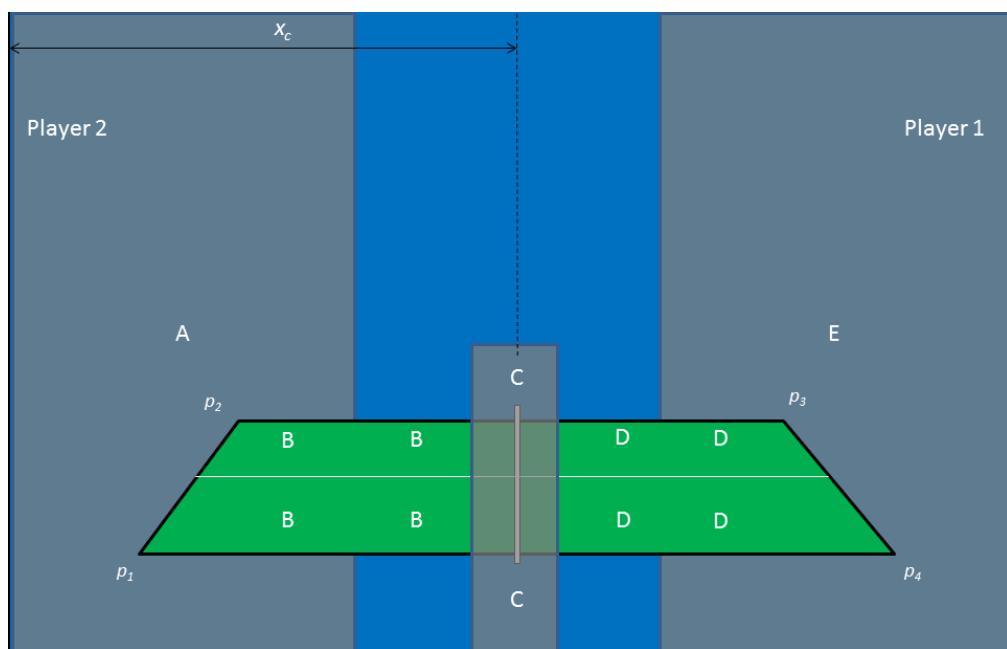


Figure 7-1 Match-playing zones

Due to the pre-defined position of the camera, defining these zones within the image is a straightforward process of applying a simple edge detector¹ and image cropping a single frame, conducted as follows:

- a. Position the camera centrally, in line with the net, at 10° to the plane of the table (as suggested in Chapter 3) and capture a single frame.
- b. **Zone C:** Calculate the horizontal position of the net by dividing the maximum horizontal image resolution by two (Figure 7-1, x_c).

¹ For these experiments, the Canny edge detector was used based on the REI results from Chapter 5.

- c. **Zones B and D:** Determine the co-ordinates of the four corners of the table (p_1, p_2, p_3 and p_4) using the Harris corner detector (Harris & Stephens, 1988). From this the maximum width ($px_1 - px_4$) and height ($px_1 - px_2$) of the table area within the image is calculated. Zones B and D represent the playing surface, either side of x_c , and appear as two halves of a parallel trapezoid.
- d. **Zones A and E:** These are the general locations of the rackets and players. Zones A and E start from the either the left or right extremes of the image to the midpoint between the start of the table and the net² ($px_1 - px_4$) given by equations (7-4) and (7-5) respectively.

$$\text{width (Zone A)} = px_1 + \frac{px_1 - px_4}{4} \quad (7-4)$$

$$\text{width (Zone E)} = px_4 - \frac{px_1 - px_4}{4} \quad (7-5)$$

7.2.1. Event states

An *event* is defined here as a significant change in table tennis ball motion, which affects game statistics, performance analysis results, or officiating decisions. Specific sequences of events form an *event type*. As only one event can occupy the state of the game at any one time, the novel proposal presented here is to incorporate a FSM into the umpire decision support system. Any given event can only be preceded and succeeded by a subset of events, as defined by the rules of the game and predictable motion of the ball. Each event therefore has a unique definition of both changes in temporal motion data and the preceding and succeeding events. Rule checking is implicit; the predictable sequence of events determines which subsequent event(s) can be legal. If the rule based sequence is broken then the event is a ‘fault’ and the score (if recorded) is updated. As an example within the rules of table tennis, if the ball is travelling from left to right then it must not change direction by travelling right to left until the ball has passed the location of the net followed by a table bounce. Anything else would indicate a foul and trigger an update to the score. Furthermore, each defined legal event occurs in a given zone. Twenty four distinct event components have been identi-

² It is an accepted limitation that there are likely to be events occurring outside of the FoV, in particular a return (or missed return). Although the camera could be positioned further away, or the use of a wider angled lens could be incorporated, the reduced image quality negates any potential benefit gain. It is also assumed here that these events are of a lower priority for direct observation and are, instead, inferred by the indirect measurement of a significant delay in the return.

fied for table tennis as part of this research and these events are grouped into seven top-level events as follows:

- a. Table bounce (E7, E8, E9 and E10)
- b. Return (E1, E2)
- c. Net (let) (E11, E12)
- d. Net (collision) (E21, E22, E23, E24)
- e. Service hit (E5, E6)
- f. Table edge (E17, E18)
- g. Over the net (E19, E20)

Related component event IDs are in parenthesis. A full description of all twenty-four component event states and their threshold values is provided in Appendix A.

7.2.2. Event characteristics

It has been established in Chapter 6 that the motion of the ball can be described with sufficient accuracy to observe changes in motion dynamics, in particular by measuring its velocity, acceleration and bearing. To classify each event effectively using these motion changes, it is necessary to understand the different combinations of component motion dynamics for each given event. As such, the following are example descriptions of the TRASE observations of ball motion characteristics for each of the four primary top-level events (a) to (d).

a. Table bounce

Detecting a ball bounce is the most basic event to categorise. In the example output from TRASE (Figure 7-2) of sequence 100-P-02, the ball bounces on the table at frame 32, moving from right to left in the field of view (FoV). The ball changes direction vertically from a minimum y-value of 144, reduces velocity from 5.04 ms^{-1} to 4.19 ms^{-1} , and changes bearing from 238° to 304° . The acceleration is steady in the range $6.50 \times 10^{-5} \text{ ms}^{-2}$ to $1.91 \times 10^{-4} \text{ ms}^{-2}$ (a small variance accountable by the toler-

ance in ball centre calculations). At the instant of the event there is a detectable spike in acceleration as it increases to $1.20 \times 10^{-3} \text{ ms}^{-2}$ (Figure 7-3b).

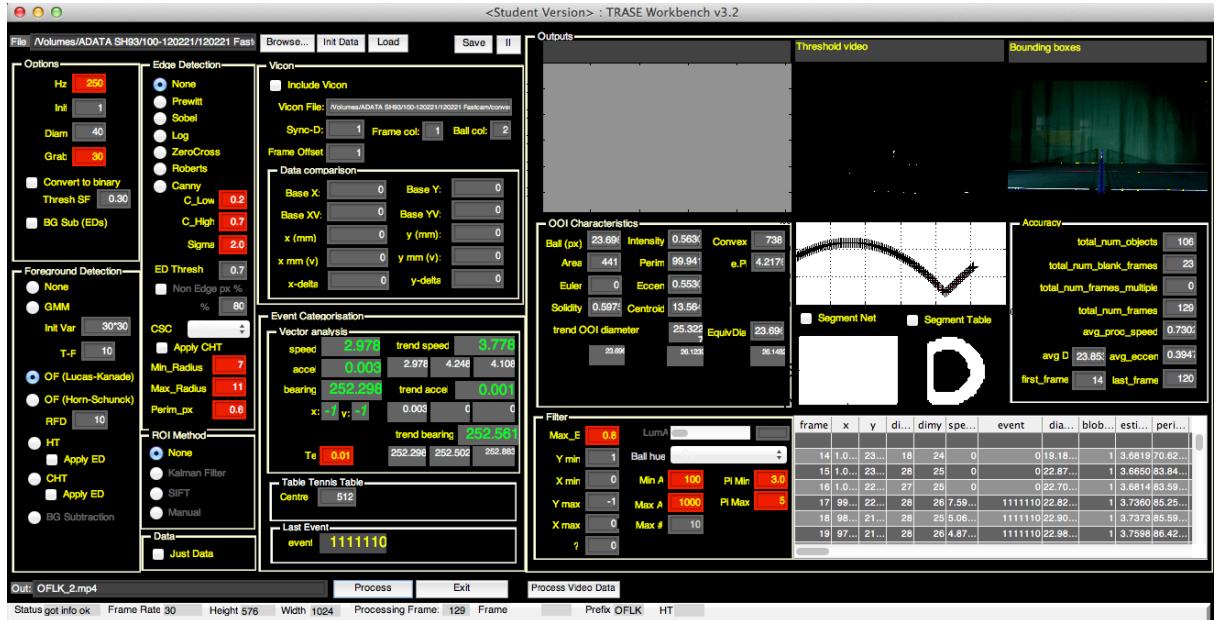


Figure 7-2 Analysis of bounce dynamics in TRASE

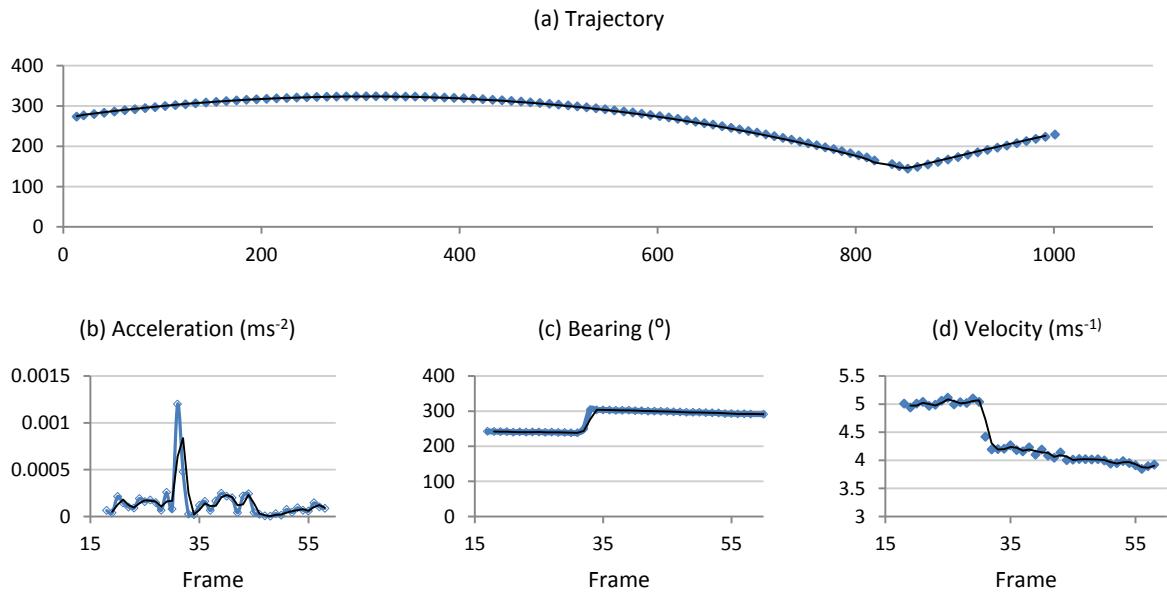


Figure 7-3 Bounce event characteristics

In summary, the expected observations are a sudden change in velocity which then remains steady (Figure 7-3d), combined with a distinct change in bearing (Figure 7-3c) and a momentary spike in acceleration (Figure 7-3a). It is important to state that velocity will not always decrease with a bounce. Analysis of a return with heavy top spin (sequence 500-P-11) travelling at high velocity

shows an increase in velocity from 7.49 ms^{-1} prior to the bounce to a velocity of 8.38 ms^{-1} immediately after the bounce. As an increase in velocity following a table bounce can only happen with top spin, this additional information would be of importance to a coach or player assessing performance when imparting spin during a service or return. Additional CV based measurements of the effects of spin, for example measuring backspin when the horizontal velocity (v_x) is reduced, proposes an interesting area for further research.

b. Return

Data from TRASE processed return events also show characteristics with apparent consistency. Taking sequence 100-P-10 as a typical example (Figure 7-4), the trajectory of the ball is mapped in Figure 7-5a, with a spike in velocity (Figure 7-5d) and acceleration (Figure 7-5b clearly visible). Combining this with the change in bearing (Figure 7-5c) which shows a dramatic reverse in direction from 88° to 286° , provides a strong indication of the return event. This is reinforced when detecting this occurrence in player Zone A or E.

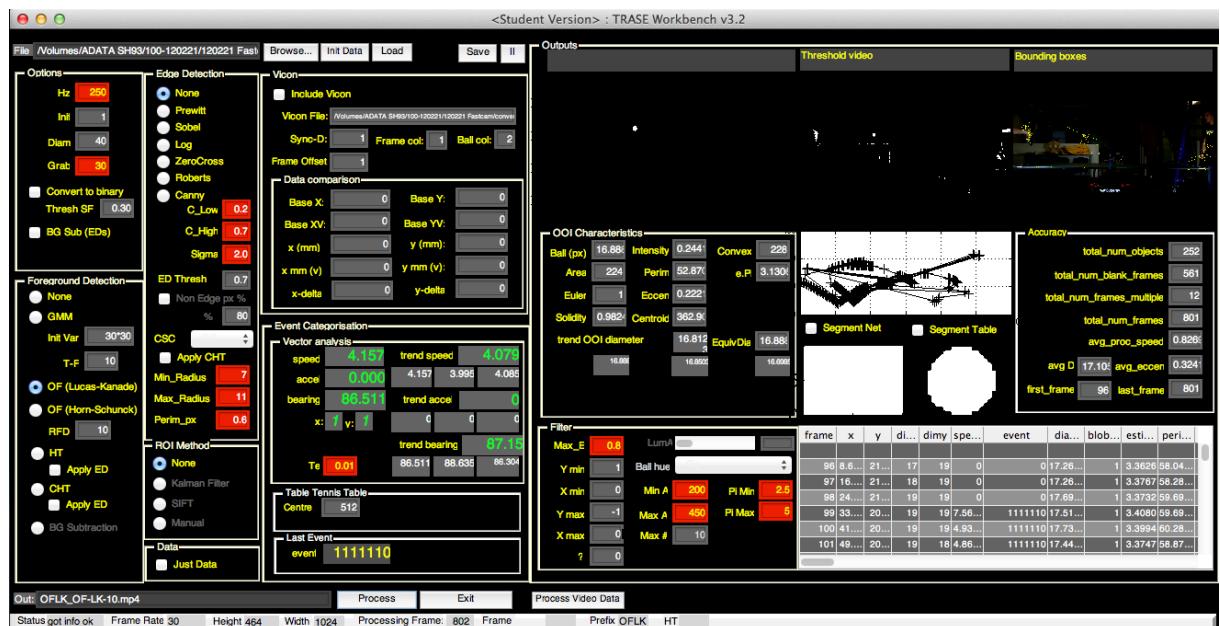


Figure 7-4 Return sequence

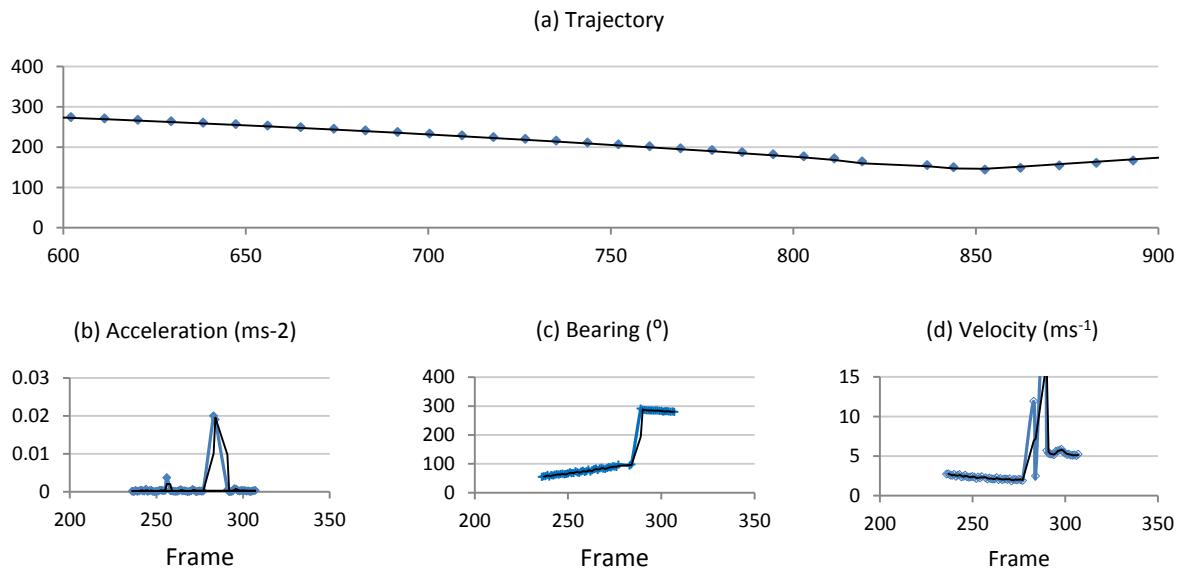


Figure 7-5 Return event characteristics

Video sequence 100-P-10 (presented above) demonstrates one of the common practical issues when using a single recording source to monitor the location of a ball, that of occlusions and transits. The data provides a practical example of missing data during the moment of contact between the ball and racket and the effect upon event classification. In this video sequence, recorded deliberately in low light, a number of ball location data are missing from the output as the ball transits the racket (see Figure 7-6). As a result there is no data between frames 278-282 inclusive. Even with the data missing at the point of contact, it can be deduced that the ball made a sudden change in motion between frame 277 and 283.

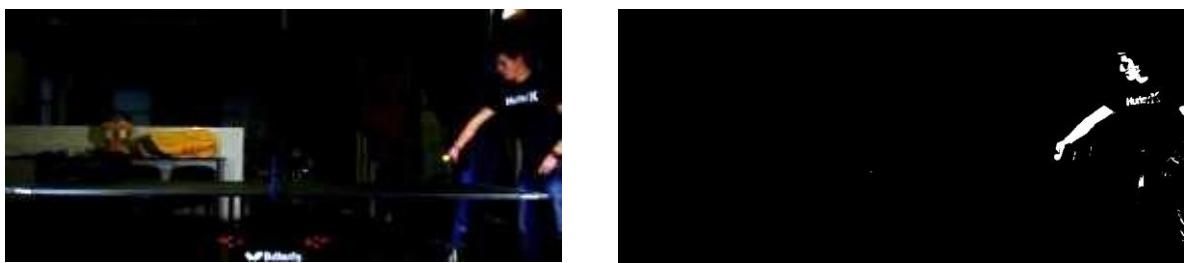


Figure 7-6 Ball boundary merging with the foreground and background

Further processing, such as data extrapolation or the application of a tracking algorithm such as the Kalman Filter, may provide indications of the missing location data. However the results above demonstrate that this extra processing overhead is not necessary, the change in motion provides sufficient information to identify the event.

c. Net (let) detection

Detecting a net (let) stroke during a serve is one of the most challenging events to classify. The observable change in motion is relatively minor when compared to a bounce or a return. When observed by human eye, there appears to be no substantial change in velocity, acceleration or bearing and it requires a trained umpire to be confident to confirm the slightest of let events. However, the slightest of grazes to the top of the net can be detected by examining CV data for motion change. It is necessary to combine all three components which together provide the fingerprint for a let. As an example using the image sequence in Figure 7-7, the slight net collision with the top of the net is processed.

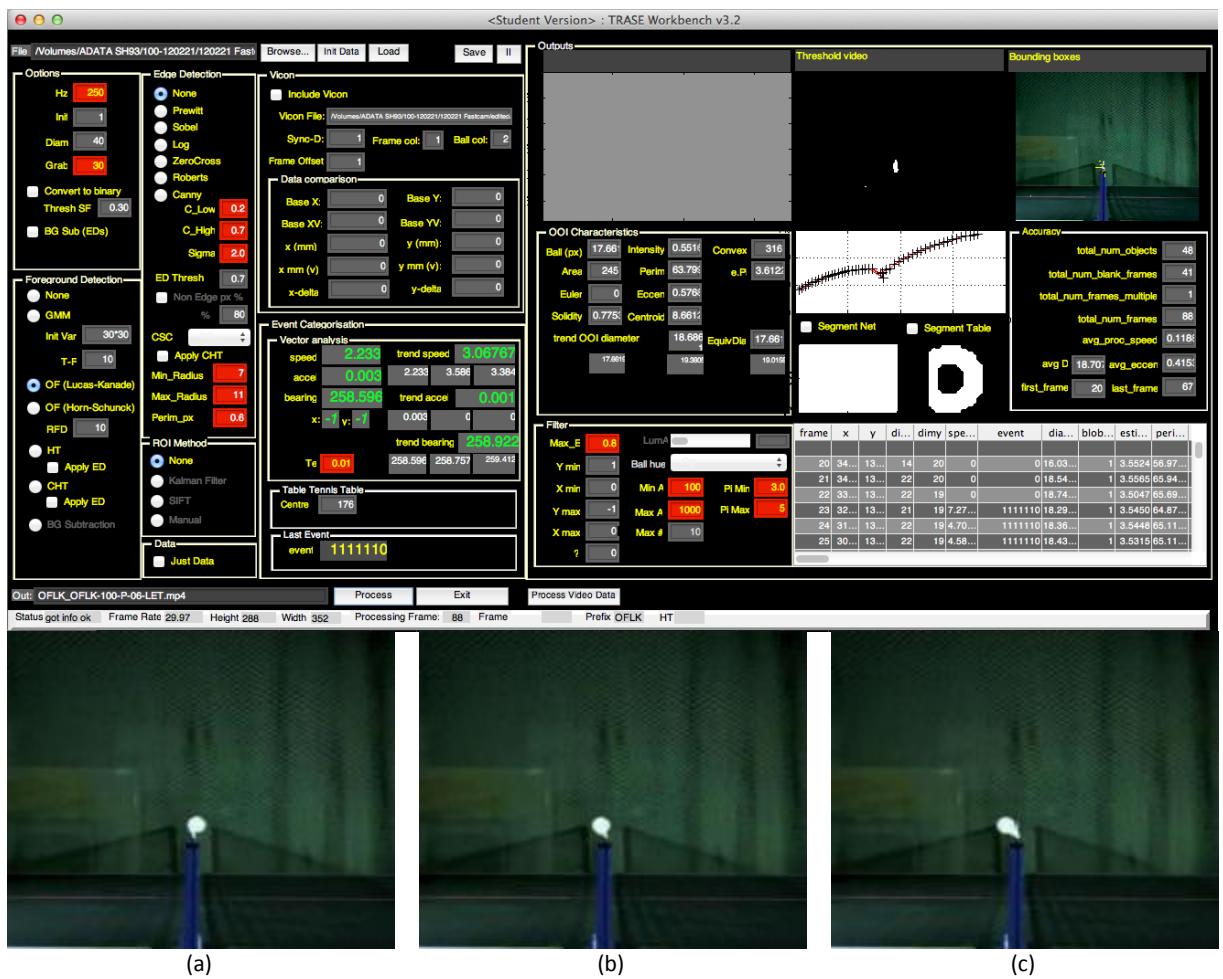


Figure 7-7 Net (let) frame sequence

Extracting positional data from the full video sequence, the motion characteristics of the let can be measured (Figure 7-8).

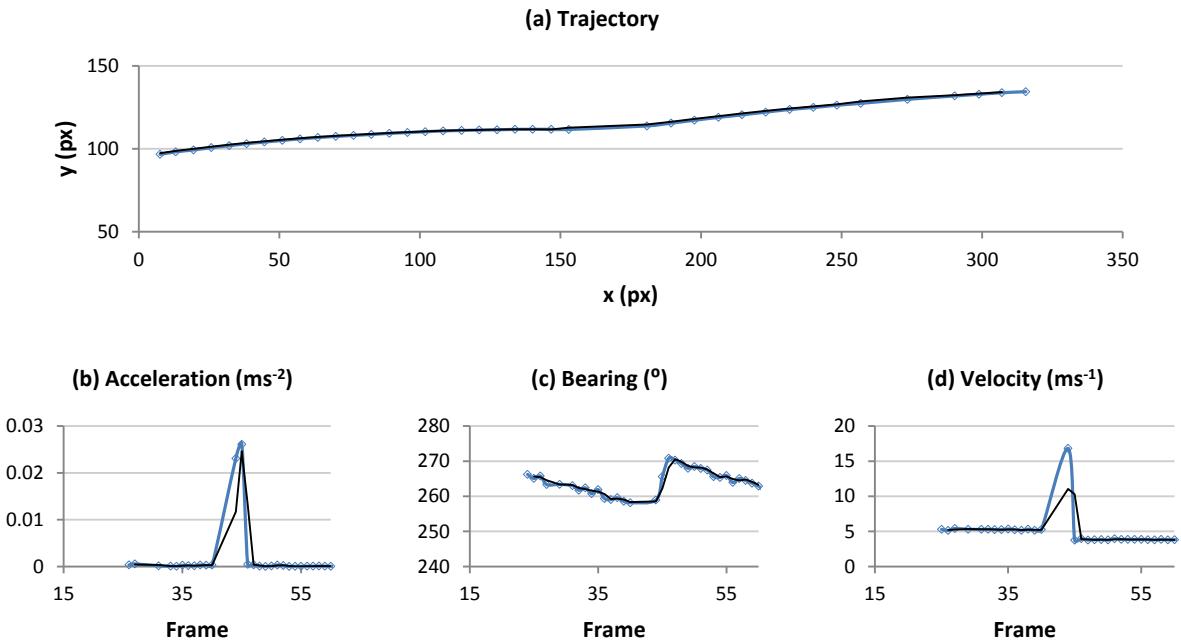


Figure 7-8 Let event characteristics

The ball makes a collision with the top of the net as observed at frame 41. When viewed by human eye alone, this particular video is a difficult net call to confirm. Figure 7-8a shows the trajectory of the ball with, on initial view, an inconclusive deviation in its path. Whereas charts b-d clearly present an event occurring at frame 41, with a spike in acceleration (Figure 7-8b) at this moment, followed by a distinct reduction in velocity (Figure 7-8d) combined with a momentary, abrupt increase in bearing from 258° to 271° (the trend of the bearing had been $< 270^{\circ}$ to this point) before returning approximately to its previous bearing. It is this pattern of characteristics which define a slight trajectory deviation due to let event.

It is useful to compare a successful net (let) detection with a ‘near miss’, in which the ball travels close to the net at high velocity without touching it. Sequence 700-P-06 is such an example where in real time it could be questionable whether or not the ball did hit the net. Applying the process presented in this research, the charts for velocity, acceleration and bearing, as follows, do not show a deviation outside of the acceptable range.

d. Net collision

The ball striking the net and failing to pass the net triggers a fault, irrespective of any preceding events. The primary characteristics used to detect a net collision are (a) occlusion of the ball at the

location of the net, (b) the re-appearance of the ball on the same side of the table as from where it was previously hit, (c) substantial reduction in velocity and (d) unpredictable changes in bearing due to rebound and spin. In sequence 500-P-02 a net collision has been captured in low light levels (see Figure 7-9).



Figure 7-9 Net collision from 500-P-02 showing typical occlusion

When reviewed by the human eye, it can be seen how the ball bounce in front of the net (Figure 7-10a) and strikes the net at frame 462 (Figure 7-9b). However, when this sequence is analysed by TRASE no ball is detected between frames 454 and 473 inclusive due to it becoming partially occluded by the net itself (it becomes ‘trapped’ for several frames within the deformed net). As the ball reappears to TRASE after frame 473 (Figure 7-9c), the velocity of the ball is tending towards zero as represented in Figure 7-10d, then increases and decreases in a sinusoidal oscillation as it bounces on the table surface, often with spin, before coming to rest.

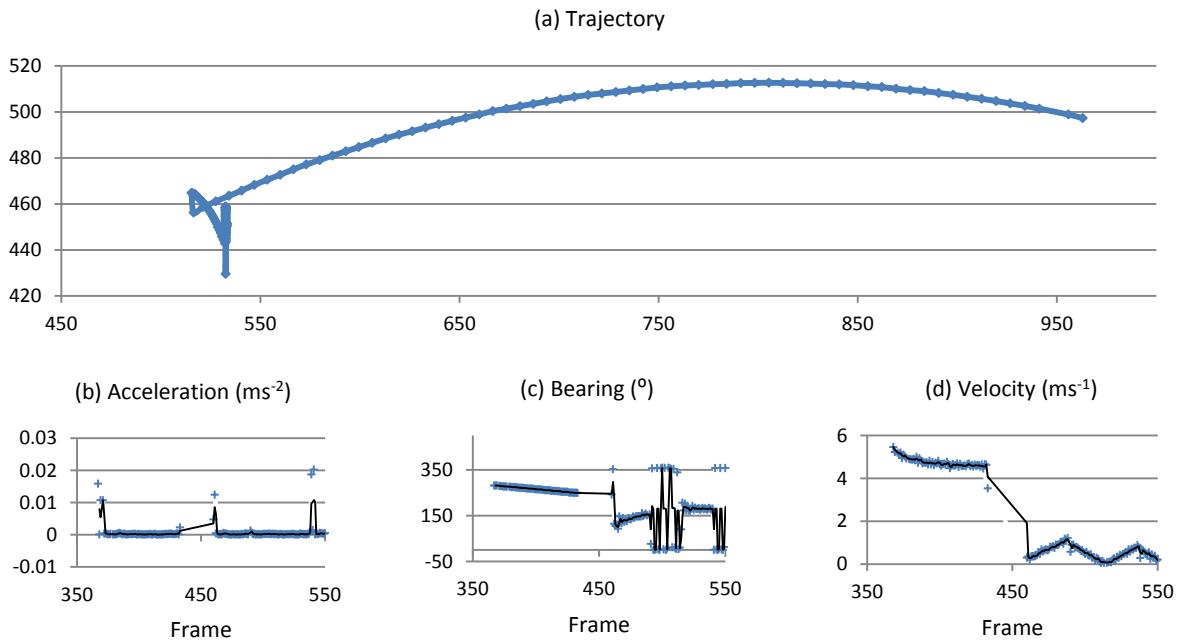


Figure 7-10 Net collision characteristics

Prior to the collision the running average acceleration is $1.81 \times 10^{-4} \text{ ms}^{-1}$ with the next recorded acceleration at frame 460 being $4.72 \times 10^{-4} \text{ m}^{-2}$ (Figure 7-10b). This spike is lower than has been observed in other collisions, such as a bounce, due to the ball losing much of its energy during occlusion with the net³. The net is placed at $x = 512$ (the horizontal resolution of the image is 1024 with the net placed in the centre of the FoV and it can be seen that the ball is never detected beyond this point). Figure 7-10c highlights the oscillations in bearing after hitting the net; it is apparent that the motion before the impact is in quadrants III and IV whereas subsequent to the collision the bearing is in quadrants I and II (the ball has been hit from a right to left direction). The subsequent oscillation in direction change is caused by the ball bouncing almost vertical with slight spin. To validate that this observation is general for all net collisions, a comparison with other sequences has been made; the result is the motion is indeed the same in all sequences. Figure 7-11 presents a further example using sequence 100-P-06.

³ In the example the detected reduction in velocity is from 4.63 ms^{-1} to 0.31 ms^{-1} in $\frac{1}{10}$ second. In reality the deceleration is probably greater, possibly achieving the minimum velocity in just a few frames. However without data for the twenty-five frames during the occlusion this could not be proven with this method.

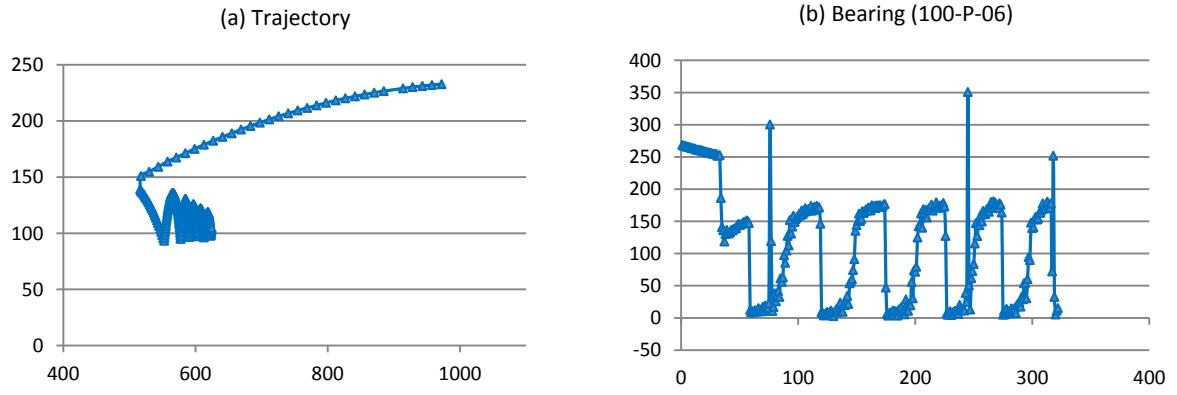


Figure 7-11 100-P-06 net collision bearing and trajectory

In this sequence the ball hits the net without detection beyond location $x = 512$, and then there is a reverse of general direction, followed by the ball bouncing left and right repeatedly due to spin direction change, until losing energy and stopping. This distinctive combination of motion characteristics has only been observed during a net collision and therefore is applied to the event description.

7.2.3. Event motion thresholds

There are apparent (small) changes in the motion of the ball caused by the quality of the video and segmentation procedures determining the centre of the ball. The continually changing trajectory of the ball in free flight, due to gravity and air flow around the ball, naturally creates changes in velocity, acceleration and bearing. When detecting changes in motion caused by a match-play event, the sensitivity of the motion detector must be calibrated to filter these motion changes, which are *not* caused by these events. The practical examination of a sample of video sequences provides evidence of the amount of sudden change in motion characteristics detectable for any given event. By applying the dynamics measurements in Chapter 6 to the top-level event sequences, a mean measurement of expected minimum change (%) in motion for velocity $\text{min}\Delta v$, acceleration $\text{min}\Delta a$ and bearing $\text{min}\Delta b$ has been calculated (Table 7-1).

Table 7-1 Event threshold values

Event	Sample size	$\text{min}\Delta v(\%)$	$\text{min}\Delta a(\%)$	$\text{min}\Delta b(^{\circ})$
Table bounce	140^4	49	0.1%	58

⁴ Camera settings: 94xCS13, 11xCS18, 20xCS02, 5xCS14.

Event	Sample size	minΔv(%)	minΔa(%)	minΔb(°)
Net collision (let)	39 ⁵	9	134	12
Net collision (net)	75 ⁶	84	160	140
Return	91 ⁷	146	189	111
Service hit	38 ⁸	560	412	205
Table edge	6 ⁹	24	196	53

As these are the *minimum* values detected during the experiments for the available sample size, an assumption is made that this sample size is sufficient to provide accurate motion threshold values. These threshold values are then used in each individual event state to test for an event. A number of events require no threshold value in the event test (for example, ‘Over the net’ E19 and E20). Instead, these events use current location data combined with knowledge of the preceding events in the sequence.

7.3. Classification results

A summary of the results of this novel classifier is presented in Table 7-2. Where events occur in both left-to-right and right-to-left play directions (such as the returns E1 and E2) the results have been combined.

Table 7-2 Summary of classification results

Event	Sample size	Classification success rate (%)		Occlusions	Classification success rate in video without occlusions (%)
Table bounce (E7, E8, E9 and E10)	140	100		0	100
Return (E1, E2)	39	79		6	95
Net collision (let) (E11, E12)	75	97		0	97
Net collision (net) (E21, E22, E23, E24)	91	100		0	100

⁵ Camera settings: 28xCS13, 6xCS18, 1xCS02, 4xCS14.

⁶ Camera settings: 65xCS13, 7xCS18, 3xCS02.

⁷ Camera settings: 71xCS13, 7xCS04, 2xCS05, 11xCS08.

⁸ Camera settings: 14xCS13, 13xCS02, 8xCS14, 3xCS14.

⁹ Camera settings: 4xCS13, 2xCS14.

Event	Sample size	Classification success rate (%)	Occlusions	Classification
				success rate in video without occlusions (%)
Service hit (E5, E6)	38	79	13	95
Table edge (E17, E18)	6	67	2	100
Over the net (E19, E20)	126	100	0	100

The results show a promising application of 2D based CV processing to event classification with an overall success rate of 95.9%. The majority of failures occur when the ball, during returns and services¹⁰, is partially occluded by either the player or racket, when using a monocular recording device. Removing occlusion sequences from the ‘return’ and ‘service hit’ events improves success rates for both to 95% (giving a total success rate for event detection of 98.8%). However, as Wong (2007) notes in his assessment of ball occlusion during a service ‘the view of the ball is blocked and rule 2.06.04 is violated’ inferring that a failure for the camera to have line of sight of the ball during a serve is equivalent to an umpire not seeing the ball. With the suggested positioning of the camera in a location analogous to that of the umpire, this would immediately indicate the service was a fault, prior to any ball detection being made. Table edge events without occlusions were detected with a success rate of 100%, however occlusions occurred in 66% of all edge events. This is primarily due to the camera position only having clear sight of one edge of the table (the nearest length to the camera) and is an issue for the application of a single camera. Positioning a second camera opposite the first on the other side of the table would reduce but not eradicate occlusions when the ball hits the edge closest to the player zones.

7.4. Classification for officiating

The automatic CV based classification of events during a table tennis match is useful for performance assessment, video annotation and highlighting and the first stages in officiating. The event state model, as presented, is able to automatically determine if a given event was in the correct sequence and hence within the rules. However, there are a number of events in table tennis which require additional information to determine if they are within the rules of the game. In particular, these include the service and the doubles line rule. A further application of the service measurement algo-

¹⁰ There is also the occlusion caused when the ball strikes the *edge* of the table opposite the camera. This is a relatively rare event, and can be easily mitigated by a second camera opposite the first, if considered a requirement.

rithm, to assist the umpire, is table yield (the bounce given by the combination of ball and table). Details of novel proposals for automatically classifying these events and measurements are provided in the following sections.

7.4.1. Service height (apogee)

According to ITTF rule 10.3.1, during a service the ball must travel ‘near vertically’ upwards, without imparting spin, so that it rises at least 16cm from the palm of the hand.’ (ITTF 2014a, p.14). Applying a monocular recording device to the service event presents complications for ball detection, extraction and measurement. Detecting the service event and checking for its legality appears in much of the work by Wong (2007) and is given further consideration here. Wong’s suggestion is to measure the angle between the location at the start of the service and the location at its maximum height, from which the height may be determined. In this method no consideration is given to (a) the location of the camera with respect to the service, (b) the image scale of reference (S_f), or (c) how to ensure that an image of the ball has been captured at its maximum height. The novel proposal presented here aims to improve on Wong’s work by solving these three complexities.

Placing the camera in the location defined in Section 3.2 enables the *majority* of events to be detected successfully. It also positions the camera *in situ*, level with the general height of the ball above the ground, as it leaves the player’s palm. This minimises perspective error and provides a plane (the y-axis) for which a near linear measurement can be taken¹¹. Furthermore, placing the camera in a similar FoV to that of the umpire and staying within the rules of the game, the ball must be visible to the umpire at all times and therefore will also be visible to the camera.

It is important to measure the ball rising to a maximum height above the surface of the player’s hand (the ball’s apogee) during a service to confirm its legality. As the palm is open and flat, detecting the start of travel is relatively straightforward; this is the point at which the ball is first identified. Prior to this the proximity of the hand prevents the feature filter from detecting the ball. Capturing the ball at exactly the apogee of its trajectory presents a greater challenge. It would be ideal to capture the ball at precisely the point of maximum height of the throw, using the change in direction of velocity as the indicator. This solution, however, requires a frame rate tending towards infinity. Therefore, another solution is necessary.

¹¹ For a discussion on the accuracy of the location of the ball when represented in 2D, see Chapter 6.

The apogee occurs at the point where there is a zero vertical velocity and the trajectory gradient is zero. The motion of the ball on *either side* of the apogee is exactly equivalent to the motion of the ball having been *dropped* from the height of the apogee. Given sufficient data points, with defined time intervals between each, the suggestion is to extrapolate the data in reverse, to find the maximum¹². Given the frame rate, it is possible to calculate the mean distance of the two frames containing the imaged ball either side of the apogee and the predicted apogee (see Figure 7-12).

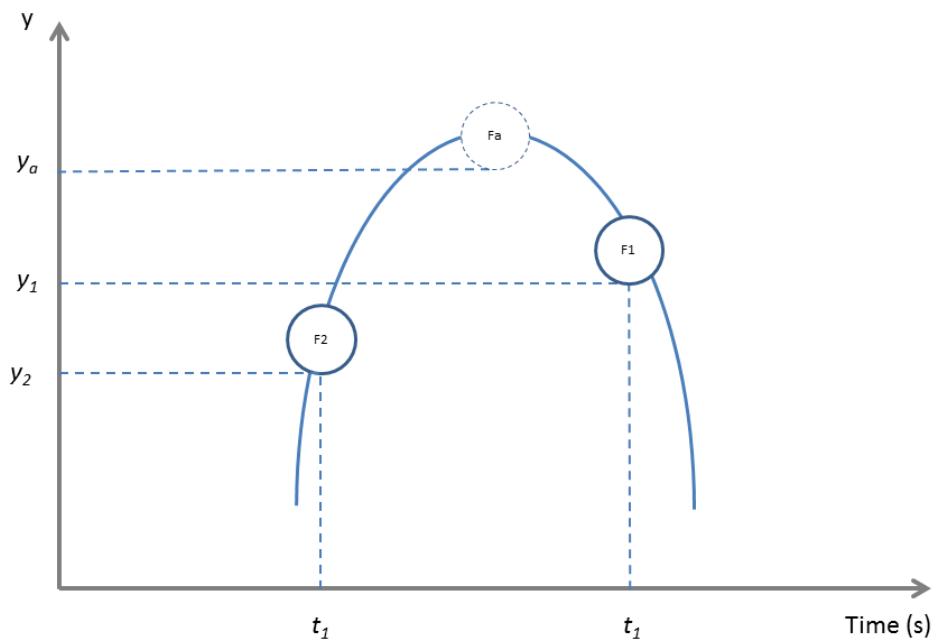


Figure 7-12 Apogee calculation

By applying the gravity equivalence principle (Encyclopedia Britannica Inc., 2014) the time taken (t) for any object to fall under gravity without air resistance is known and can be used to measure how far a table tennis ball falls in time between frames (7-6):

$$t = \sqrt{\frac{2d}{g}} \quad (7-6)$$

Where $g = 9.81 \text{ ms}^{-2}$ and d is the distance travelled by the ball in seconds (t).

Throughout the motion either side of the apogee, the ball is in symmetrical flight. At the apogee of a service (or the apogee of any shot) we know that the vertical velocity immediately before and after

¹² Although no documentation can be found, it is assumed that the height is measured from the base of the ball in the palm to the base of the ball at its apogee.

the apogee is equal, due to the constant effect of gravity and the approximate parabolic trajectory taken by the ball over this short distance¹³. For example, the ball will always take 1/20th second to travel 1 cm both to the apogee (up) and also away from the top of the apogee (down). As the recording device will be imaging the trajectory of the ball either side of the apogee, two frames will capture the ball closest to the apogee, either the frame on the way up (F1) or the frame on the way down (F2). The two heights above the palm of the hand can be calculated directly from analysis of the centre location of the ball and the image scale of reference (S_f). The mean of the two heights y_1 and y_2 (μy) can then be calculated to give a height measurement, for which the time taken to reach this point *from* the apogee (Δy) is exactly half the time interval between frames ($t_2 - t_1$). Placing these known values into equation (7-7):

$$\frac{t_2 - t_1}{2} = \sqrt{\frac{\left(2\Delta y - \left(\frac{y_1 + y_2}{2}\right)\right)}{g}} \quad (7-7)$$

$$\Rightarrow \Delta y = 2 \left(g \left(\frac{t_2 - t_1}{2} \right)^2 + \left(\frac{y_1 + y_2}{2} \right) \right) \quad (7-8)$$

Where $g = 9.81 \text{ ms}^{-2}$. Therefore the height of the ball, given the height of the two frames closest to the apogee, is:

$$y_a = \mu y + \Delta y \quad (7-9)$$

It is important to note that a frame rate of sufficient frequency is required to capture the ball somewhat close to the apogee. Experiments here indicate that the service throw takes approximately 0.8 seconds to reach the minimum height required for a legal service. Even a low speed camera operating at a frame rate of 60 Hz would provide sufficient data points for service analysis. This method does not classify service faults due to the ball not resting in the palm of the hand at the start of the event. Further research and algorithm development is required to categorise the shape of the hand at the beginning of the service. A suggestion would be to apply a bounding box to the hand image. An eccentricity (e) tending towards $e \leq 1$ with the ball not detected in the frame would indicate the hand is not being held flat and is obstructing the ball from view.

¹³ The low velocities at the point of apogee lead to an assumption that the air resistance is negligible and there is no effect of spin.

7.4.2. Table yield

This equation used for calculating the height of the service throw (7-9) can also be used in assessing table yield for construction materials of the table and ball. In this rule a standard table tennis ball dropped from a height of 30 cm must uniformly bounce about 23 cm (ITTF, 2013b). The moment of the bounce can be detected, together with the y-coordinate of the centre of the ball. As the ball reaches its apogee, two frames either side of this point are identified (through the change in direction) and the mean height of the ball in these frames are determined. The difference between the mean height in the two frames and the apogee (Δy) is calculated by equation (7-6) and the total height of the bounce is recorded.

7.4.3. Centre line

The table is divided into two equal half-courts along its length by a white centre line, 3mm wide, running parallel to the length; this centre line is regarded as part of each right half-court (ITTF, 2013b). According to rule 2.6.3, when serving during a doubles match, the ball must land on the right half of the server's and then the right half of the receiver's court to be deemed legal (ITTF, 2014b). Presented here are details for calculating the location of the centre line and determining the location of the ball with respect to the line, or whether the ball has landed *on* a line, using a monocular recording device as the imaging source.

During the experiments, detecting this line with any degree of accuracy or consistency has shown to be a non-trivial task due to low contrast and reflection 'hotspots' caused by the low perspective of the camera. The Harris corner detector (Harris & Stephens, 1988) used in the proposed classifier to determine event zones is again used here to detect the corners of the table. Re-using the data for identifying the corners of the table (p_1 to p_4), a parallel trapezoid may be defined to present the table surface (see Figure 7-13). A characteristic of any trapezoid is that its centre may be found by the intersection of its diagonals $\overrightarrow{p_1p_3}$ and $\overrightarrow{p_2p_4}$.

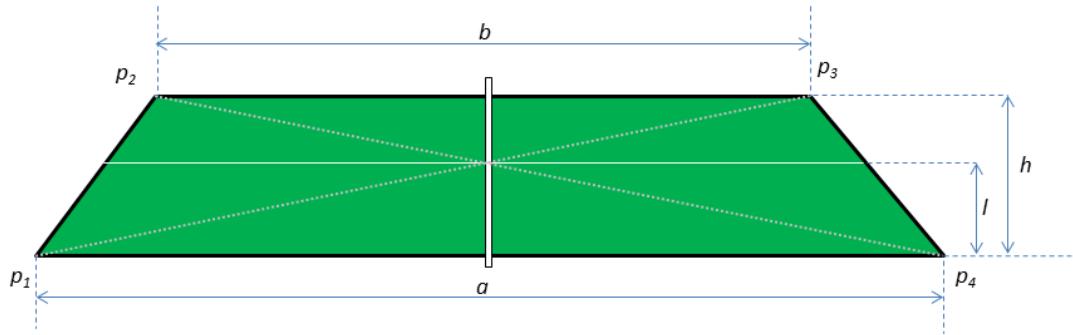


Figure 7-13 Centre line location

With the camera positioned along the line of the net, the trapezium is practically a parallel trapezoid. Allowing for off-centre camera alignment causing small parallax errors, the distance between the two parallel edges of the table (h) is calculated as $\frac{1}{2}(py_1 - py_2 + py_4 - py_3)$, where py_n is the y-coordinate of the corner location. Applying the formula for the centre of a trapezium (eFunda Inc., n.d.), the perpendicular distance from table edge $\overrightarrow{p_1p_4}$ to the line through the centre of the table at distance l is calculated using equation (7-10):

$$l = \frac{h(2a + b)}{3(a + b)} \quad (7-10)$$

Where $a = (px_4 - px_1)$, $b = (px_3 - px_2)$ and px_n is the x-coordinate of the corner location. This distance (l) from the leading parallel edge forms the basis of y-coordinate reference for doubles line decision calls

According to the ITTF the width of the doubles line on the table is 3 mm. Due to the transformation of the image by the camera placed at an angle \emptyset to the table surface plane it will appear narrower (only perpendicular to the table plane does the line resolve to the full 3mm in width). The width (w) as the line appears to the camera lens focal plane can be determined using the equation (7-11).

$$w = 3\tan\emptyset \quad (7-11)$$

At a camera angle of 10° the width of the line resolves to approximately 0.53 mm, which in most video sequence recordings in this research is less than 1 pixel in width; this explains part of the reasoning for the difficulty in detecting the line using image processing techniques. Generalising equation (7-10) for the potential increased precision of higher resolution recordings, the distance (l) in

pixels between $\overrightarrow{p_1 p_4}$ and the doubles line is given by equation (7-12) where S_f is the image scale of reference and py_1 is the distance in pixels from the bottom of the image to the table.

$$l = \frac{h(2a + b)}{3(a + b)} - 3S_f(\tan\theta) + py_1 \quad (7-12)$$

Using the algorithm proposed in Section 5.6 the y-axis value of the centre of the ball at the point of a bounce *near* the doubles line may be determined. The ball's centre location in the y-axis (y_c) as seen from the camera is not the point at which the ball bounces on the table (see Figure 7-14). The point of contact of the ball on the table is at the lowermost edge of its circumference and, when the camera is at an angle α to the plane of the table, is therefore occluded from the camera by the ball itself.

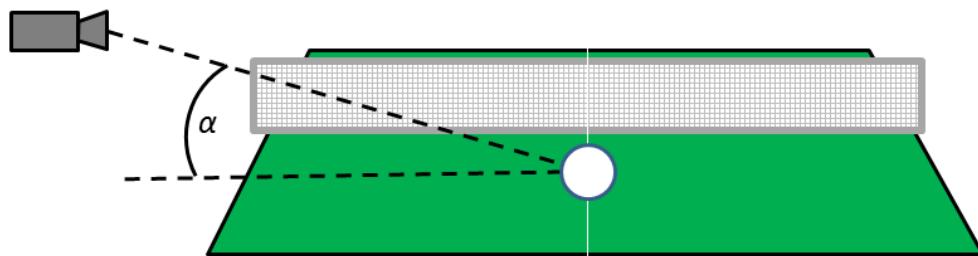
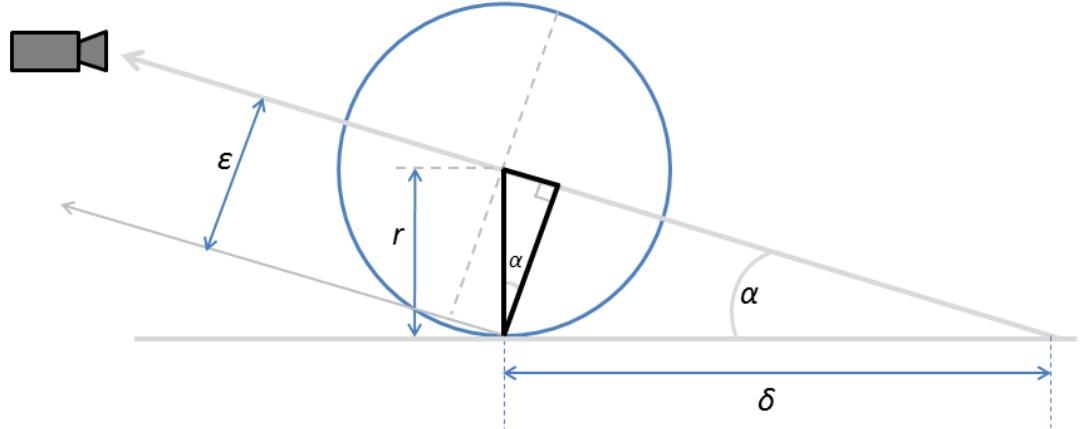


Figure 7-14 Occlusion of the doubles line

To improve the detection of a line fault of a service in a doubles game from a monocular image, an adjustment to the calculation of the bounce location (y_c) triggering the event must be made. To calculate this adjustment, assume the 40 mm diameter ball hits the centre line precisely on the left edge of the 3mm doubles line. When viewed from the angle of the camera, the centre of the ball is projected along the table a distance of 54.95 mm (δ). See Figure 7-15 below.


 Figure 7-15 Doubles line correction factor (ϵ)

When using the centre of the ball as the location descriptor in a 2D image, an equivalent location for the centre of the ball to be above the doubles line requires moving the ball by δ towards the camera. Using the scale factor (S_f), the correction factor (ϵ) can be calculated using equation (7-13).

$$\cos(\alpha) = \frac{S_f r}{\epsilon} \quad (7-13)$$

The angle subtended by the camera to the plane of the surface of the table is α and s is the image scale of reference (ratio of mm to pixels). The radius (r) of the ball is a constant 20mm. Following the proposal in this research for the standard camera placement and configuration, α will be approximately 10° . The correction factor to the y-axis coordinate of the centre of the ball of $+18.8S_f$ (px) for right to left play, or $-18.8S_f$ (px) for left to right play, is to be applied. In the test sequences presented during these experiments, the scale of reference is in the range 3-2px to 1 mm. This requires a shift in centre ball detection above the doubles line of 6.3-9.4 px. With a doubles line of only 3mm (1-1.5 px) the tolerance is small and requires this degree of accuracy for an automated officiating system to provide confidence in its use. It is also of importance to consider future improvements in video recording resolutions and frame rates where this becomes increasingly relevant.

7.5. Application to umpiring and coaching

An automated solution for performance analysis and match rule contravening, as presented in this chapter, benefits the player, umpire and coach. This is through the immediate feedback of either comprehensive ball dynamics data (such as ball location, deviation, bounce height etc.) or by meas-

uring the accuracy and consistency of a player's shot (location of the bounce, speed of return, reaction times, etc.). Furthermore, a specific event of interest would be identified and the relevant video (with supporting data) replayed to the players, coaches or umpires, letting their expert knowledge contextualise the CV outputs. It is this additional human experience and knowledge which provides the final constructional analysis.

Moreover, solely depending on such an automated system to infer match play decisions and performance coaching would be unreliable. As demonstrated by the Indian Wells 2009 Murry V Llubicic tennis match (Hawk-Eye Innovations Ltd., 2009), even sophisticated, high-cost, dedicated technologies are prone to occasional error. The TRASE system and its generated outputs are intended to be used alongside a skilled coach or umpire, providing additional supporting evidence. It should not be considered as a replacement for the highly qualified match personnel.

7.6. Conclusion

All event classification using a monocular recording device requires accurate determination of the change in distance of the ball over time. Using 2D data in itself is sufficient for the *majority* of event classifications in table tennis. However, it requires a carefully positioned imaging device capable of recording sufficient amount quality data. Multiple cameras assist with occlusions, particularly those observed during table edge events but they do not necessarily benefit the detection of a bounce, a let, service or net event.

All identified table tennis events may be classified using derived velocity, acceleration and bearing characteristics. Absolute real-world velocity and acceleration calculations of the ball are not essential for detecting faults or automatic scoring. Evidence provided here indicates that of greater importance are *relative changes* in ball motion, within tested threshold values, detected by CV processing. It will not falsify results if the calculated velocity and acceleration of the ball is incorrect, as long as the error factor is constant throughout all calculations. Comparing a running average of smoothed data for velocity, acceleration and bearing for the previous three frames offers sufficient accuracy for a high success rates in event detection, with failure only occurring due to occlusions from the table or player. For any given sequence which does not require world unit measurements in event detection, knowing the frame interval rate and ball dimensions is not necessary. For these events the ball diameter can be reduced to a unitary value for direct comparison with other video

sequences. However, determining a relative real-world locations and velocities of the ball become significant in events which require distance measurements, such as the service and doubles line fault. Novel proposals for an improved measurement and fault detection application for both the service and doubles line faults have been detailed. Although not a direct match-play event, the measurement of the table yield as an officiating activity, based on the service height validation, has also been proposed.

An event state analysis can be used to validate the ‘legality’ of sequences (or combinations) of events. Novel event inputs and outputs have been documented and evaluated within TRASE to enable automatic scoring and fault detection direct from event classification. Areas within the scene have been identified as ‘inactive’ for event detection. Conversely, focussing only on zone ‘hotspots’ has been shown to be useful in reducing processing overheads, without degrading the accuracy of event detection.

8. Conclusion and further research

8.1. Introduction

There has been a recent increase in the use of sports video sequence processing to automatically generate object of interest (OOI) trajectory analysis and motion prediction. The desire from spectators for closer involvement, the need from governing bodies for unbiased decision support systems and the increased value broadcasters can offer from such solutions is becoming an expectation of many sports. However, not all sports are benefiting from this technology, due to complications in installation and prohibitive costs. With this background, the research presented here has been established to investigate, using known Computer Vision (CV) algorithms, the suitability of consumer-grade monocular recording devices for officiating support and player performance analysis in low-investment sports. As a popular sport with global reach and without an established CV based video analysis solution, within the research context presented in this thesis, table tennis has been selected as the representative sport in this domain.

The focus has been to test the hypothesis that a monocular camera is capable of successfully detecting match-play events in table tennis using only 2D ball motion data. These events may then be applied to automatic officiating and player performance systems. Throughout the research, it has been demonstrated that well-understood and repeatable experiment-based data is necessary to justify the CV algorithm selection for successful detection of a ball and its location within the image. A CV based method for the initial stage of detecting a ball in table tennis video sequence had already been suggested (P. Wong, 2009) and this, along with the suggested feature filter, has been enhanced and implemented throughout this research. However, a thorough investigation into the necessary camera location, hardware requirements and complications caused by environmental conditions (Chapter 3), was also required. The CV algorithm must not only be capable of detecting the ball for the majority of frames, but it must do so during key events for automatic classification. This research discovered a variation in the detection capabilities and determined ball location outputs of different CV algorithms (Chapter 4), causing false event detection triggers. Therefore, a novel method of measuring event centric detection efficacy, called the Efficacy Metric Set (EMS), resulting in the Relative Efficacy Index (REI) has been developed (Chapter 5), alongside a novel method of confirming the relative accuracy of the motion of the detected ball between frames (Chapter 6). Finally, using this motion data it has been shown that there are repeatable observed ball trajectory patterns, which

have been successfully applied to a novel solution for automatic event classification (Chapter 7). These combined stages present a new framework for the future evaluation of monocular based CV algorithms in table tennis and ball sports in general.

8.2. Discussion

Central to the proposed framework has been the depth of understanding and rigour in analysis. However, even with carefully controlled experimentation, a single, monocular recording device will never be able to determine an accurate 3D position of the ball, nor will it capture all events. There is also no accurate determination of depth and there is no single location in which the camera can be placed to avoid the potential for occlusions. Furthermore, the field of view (FoV) of single camera is compromised by a) its distance from the table necessary to capture all events, and b) spherical aberration caused by the wide angle lens. The only complete solution is to provide extra asynchronous cameras aimed specifically at the players in such an orientation as to avoid occlusions.

Yet, there are useful outputs if the question is changed: rather than attempting to accurately determine the location and hence trajectory of the ball, instead examine changes in *relative* ball motion and use a knowledge model to generate useful information for performance analysis, automatic scoring or officiating support. To enable this, first ensure the focus of play is the table. The position of the camera is critical in obtaining suitable trajectory data and avoiding occlusions. Through experimentation, it has been shown that the optimum location for the camera¹ is 2m-3 m from the edge of the table in line with the net at approximately 10° to the playing surface. As might be expected, this coincides with the location of the umpire. The location is sufficient for the majority of observations. A single camera placed at the side of the table can record necessary trajectory deviations. However, the occurrence of the ball striking the table edge directly opposite the camera is the single event for which the camera will never have direct line of sight of the ball. For these events, a second camera, placed directly opposite the first, will ensure all events are captured.

It is a proposal of this research that, due to the high speeds of the ball travelling relatively short distances, a frame rate of 200 Hz is necessary to capture sufficient data points for player performance considered to be at club level or above². This is most relevant during a let event, where trajectory

¹ Set the camera to manual focus on the centre of the table and adjust the FoV wide enough to capture the playing zone.

² For beginners or players of low proficiency, frame rates of 100 Hz would be adequate.

deviation tends towards a minimum for all events. These configurations preclude the use of current webcam and mobile phone technology recording devices, due to focus and frame rate limitations. It does allow for high speed, dedicated cameras intended for consumer use, such as the Casio Exilim™ range (CASIO Europe GmbH, n.d.).

For the CV algorithm comparison presented, a set of 276 video sequences of varying qualities and player performance were applied to fifteen combinations of CV based algorithms. The algorithms included edge detectors or foreground segmentation processing, combined with a simple Circular Hough Transform (CHT) (Duda & Hart, 1973) or a bespoke feature filter. Each CV process was evaluated using the EMS to derive an REI. Using this method of evaluation, it was concluded that the Horn-Schunck Optical Flow (OF) algorithm (Horn & Schunck 1981) performed most successfully whilst simple edge detectors were shown to be generally unreliable. However, the REI also indicated that Horn-Schunck OF was not the most effective for one particular event: the let. During this event, the Canny edge detector with the bespoke filter was deemed the most effective. This demonstrates how the use of the EMS should be used to offer further insights and justification when selecting algorithms for CV processing. In this case, the proposal is to use the Horn-Schunck OF algorithm as the default method, whilst automatically switching to the Canny algorithm as the ball approaches the zone around the net.

Previous research for CV within table tennis has often focused on the algorithm, or the detection capability, without consideration of the accuracy of the location of the ball. The key findings in this research demonstrate the need to understand the accuracy when calculating ball features and ultimate motion of the ball. It must not be assumed that all CV algorithms determine the ball's size, shape and location with parity, as Chapter 4 has shown. Validation of the CV method for location and relative motion of the ball between frames was made using a parallel Vicon installation (Vicon Motion Systems Ltd., n.d.). Comparison data results in a mean accuracy, with an incorporated scale factor, of 3.01mm when applied to a selection of nineteen video sequences and events. This tolerance is within 10% of the diameter of the ball and accountable by the limits of image resolution. The scale factor was dynamically calculated and updated (for every frame) using the known size of the ball and dividing that to its measured size, in pixels, within the image. Due to the positioning of the camera, no depth data is provided, and so the continually moving ball has been shown to provide the most accurate method of determining the scale factor.

Given this proposed installation and configuration of the monocular camera, combined with the use of the Horn-Schunck OF and Canny algorithms, event detection using pattern matching and event states to validate the ‘legality’ of event sequences, has been demonstrated to be a viable option. The results show a promising application of 2D based CV processing to event classification with an overall success rate of 95.9%. The majority of failures occur when the ball, during returns and services, is partially occluded by either the player or racket, when using a monocular recording device. Removing occlusion sequences from the ‘return’ and ‘service hit’ events improves success rates for both to 95% (giving a total success rate for event detection of 98.8%). Furthermore, a novel method of determining the height of the ball at apogee during its trajectory has been presented. The suggested uses for such a method are in determining the legality of the throw of the ball during a service and in evaluating the suitability of the table through the measurement of its ‘yield’. A final novel practical application using event and trajectory analysis has been determining the legality of a centre line fault, by calculating its projected location during a bounce on the table from only the 2D data available.

8.3. Contribution highlights

As summary highlights, this research contributes to the domain of CV in sport, in particular as referenced to table tennis as the case study, with the presentation, analysis and discovery of the following:

1. A comprehensive root analysis of CV processing capabilities and evidence based optimal CV application, using empirical data from monocular recordings of table tennis match play events (see Chapter 2). The result is a CV algorithm review and selection framework extendable to other ball sports (**iv**).
2. A database of reusable, designed, event based monocular table tennis video sequences (see Chapter 3). A software environment within which to directly and automatically compare multiple CV algorithms against this video library (**vi & vii**).
3. Presented in sections 4.4 to 4.6, the discovery of variations in the measurements of ball motion detection and shape characteristics between different CV algorithms (**ii**).
4. A rigorous comparison of CV algorithms against a novel event based measurement, the Efficacy Metric Set (EMS), as discussed in Chapter 5. The creation of a unique, unbiased, event based

comparison index, the Relative Efficacy Index (REI) (see section 5.2). Application of the index additionally includes the justification for automatic switching of CV algorithms for specific events (section 5.5) (ii).

5. In Chapter 6 an approach for the non-intrusive validation of CV-based detection of changes in ball motion using measurements of velocity, acceleration and bearing is presented (i).
6. Feasibility studies for using the table tennis ball as the continually optimised scale of reference to measure real world distances from 2D images (see section 6.3.1) (v).
7. Chapter 7 proposes a novel method of unsupervised table tennis (or similar ball sports) event classification based on 2D ball motion data and the sequence of expected events (iii).

A list of existing and planned publications as a direct result of this research can be found in Appendix D. References to these publications are highlighted in parenthesis above.

8.4. Further research and improvement

8.4.1. The EMS as a measure of predictability

The EMS provides a measure of the success of a particular algorithm in detecting an object during extreme changes in motion. Although designed here for use in the evaluation of CV algorithms in ball sports, the EMS and resulting REI also provide a basis for event detection evaluation in other scenarios where a moving object must be detected. For example, a security solution, designed to track human movements, would implement the EMS *in reverse* as a measure of predictability. In this way the EMS will automatically generate a co-efficient of sudden and unexpected changes in motion, as the detector ‘loses’ track of its object of interest.

8.4.2. Automatic algorithm switching

It has been shown that the EMS (combined with the ultimate REI) provides a ranking of evaluated CV algorithms at any given event. Therefore, for any given event there will always be a justification for implementing an optimum algorithm. A future project would be to refine TRASE (or any other CV comparison framework) to use the output of the event categorised REI to automatically switch to the optimum algorithm for any given event. When combined with a knowledge base containing the pre-

dicted sequence of events, this would ensure the highest success rate of ball detection within each frame whilst also reducing CV algorithm selection identification and processing overheads.

8.4.3. Let detection

The focus of this research has been to detect events using the changes in the motion of the ball as the primary indicator. However, additional observations have been made which, with further analysis, could be integrated to support the event identification process. One such example is in the detection of a let event. It is conceivable that a ball with significant velocity, only marginally strikes the net, causing a motion deviation which is beyond the limitations of the image resolution and therefore becomes undetected. On these occasions, it is more likely that detection is made through the movement of the net. A simple net segmentation will highlight the change, as can be seen in the following sequence of frames (a-i) taken from 100-P-06 (Figure 8-1).

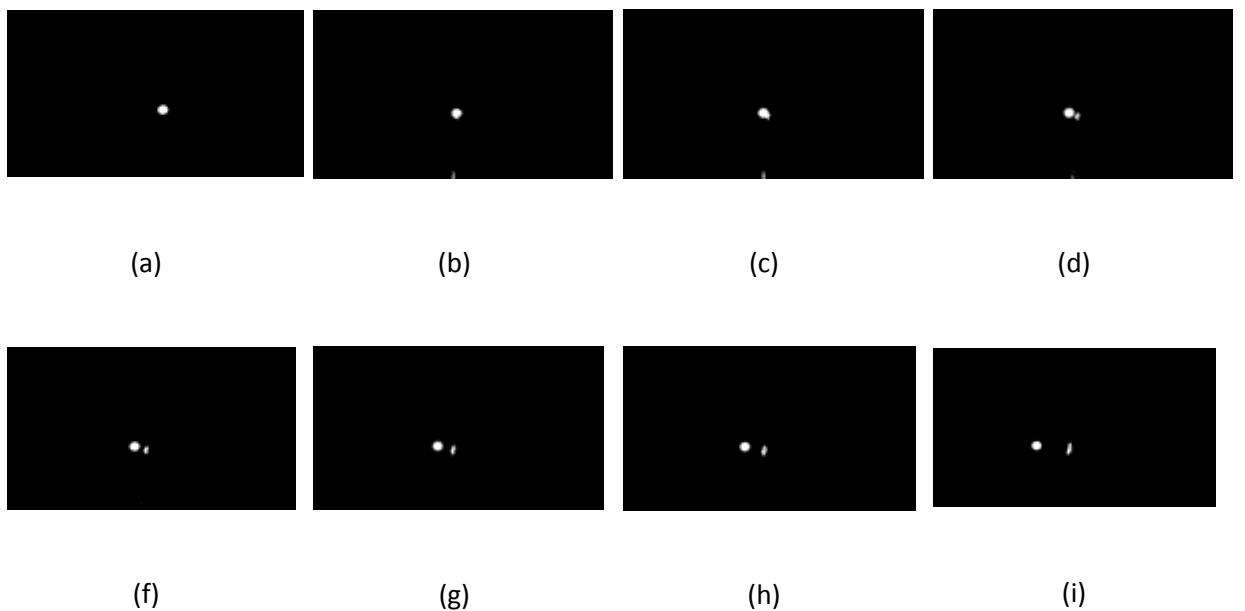


Figure 8-1 Net motion detection

The net is visible behind the ball as a shimmering contour, appearing to move with the ball for three frames, before continuing to oscillate for over 60 frames, long after the ball has left the FoV. The suggestion here is that this analysis would provide supporting evidence of a let event, even when a change in motion of the ball remains undetected. Similar confirmations may be made for other events by observing the movements of the players and the rackets during a service or return.

8.4.4. Human vs. computer officiating

A full comparison of the abilities of skilled human officiating when compared to the automated CV solutions suggested by this research would enable a detailed study into the strengths and weaknesses of each. For example, would the human estimation of the service throw height be as accurate as the measurement taken by the automated system? What are the limits of detection of a let for both the human and CV software? It is proposed that a blinded experiment, consisting of several expert umpires, simultaneously officiate a match. Each decision for each umpire is recorded and annotated with the time taken for the decision. The match is also recorded and is then used as the source video sequence for CV algorithm comparison and software development. Any results could be used as a justification for the use of a CV based solution.

8.4.5. Other sports

The presented framework for detecting changes in a ball's motion may be generalised to be applied to many other racket sports, such as squash and badminton. As an example, there is little research in the application of CV for ball detection in squash. However, a short discussion is included here as there are a number of similarities between squash and table tennis from a CV processing perspective. The ball size, high speed, player and racket occlusions of the ball, the fixed surface boundaries, rule sequencing, numbers of players and court markings all have common features. With squash there are increased challenges due to higher off-the-racket velocities, and increased ball deformation. In some respects, however, squash does not have some of the complexities of table tennis. The ball to background contrast, for example, is generally high with reduced confusion from the background. Also the limits of match-play area are clearly defined within the boundary of the squash court. Hardware installations are considerably less complicated by the permanent structures within which squash is played. Extensions into more motion restrictive sports, such as snooker or curling, could also be made with the inclusion of multi-object detection and tracking

8.4.6. Machine learning based classification

The framework presented for the determination of ball dynamics uses three characteristics of the ball motion namely, the velocity, acceleration and bearing. For all positions of a ball within its trajectory, these three features are recorded and were used in the determination of ball dynamics, based on a rule based approach widely accepted to be more computationally efficient as compared to

learning based approaches. Potential exist to use machine learning that will use data from known ball trajectories to predict the detailed behaviour of ball dynamics. This is an idea that will need a significant research effort and hence was considered outside the scope of the research context presented in this thesis. Nevertheless it is a very useful investigation that could lead to research findings of significant practical importance.

9. Bibliography

- 3M Industrial Adhesives and Tapes Division, 2003. 3 Scotchlite™ High Gain Reflective Sheeting 7610. Available at:
<http://multimedia.3m.com/mws/mediawebserver?mwslid=66666UF6EVsSyXTlx248TaEVtQEVS6EVs6E666666--&fn=7610.pdf> [Accessed May 12, 2012].
- Adidas, 2014. micoach Smart Ball. Available at: <http://micoach.adidas.com/gb/smartzball/> [Accessed June 14, 2014].
- Aggas, L.T.N.S., 2013. Hawk-Eye in sport - Luke Aggas, Hawk-Eye Innovations Ltd - The Naked Scientists. Available at:
<http://www.thenakedscientists.com/HTML/content/interviews/interview/1000345/> [Accessed September 12, 2013].
- America's Cup Event Authority LLC, 2013. America's Cup / Virtual Eye. Available at:
<http://americascup.virtualeye.tv/> [Accessed December 10, 2013].
- Animation Research Ltd, 2013a. Cricket 2010 - Ball tracking system. Available at:
<http://www.youtube.com/watch?v=XAX2mIXxjiU> [Accessed February 9, 2012].
- Animation Research Ltd, 2013b. The Story of Virtual Eye Cricket. Available at:
<http://www.virtualeye.tv/index.php/the-sports/virtual-eye-cricket> [Accessed October 23, 2013].
- Animation Research Ltd, 2013c. Virtual Eye. Available at: <http://virtualeye.tv/index.php/the-sports/virtual-eye-golf> [Accessed October 23, 2013].
- Apple Inc., iPhone 5 - Technical Specifications. Available at:
https://support.apple.com/kb/SP655?locale=en_GB [Accessed October 17, 2014].
- Arbabmir, M.V. et al., 2014. Improving night sky star image processing algorithm for star sensors. *Journal of the Optical Society of America*, 31(4), pp.794–801. Available at:
<http://josaa.osa.org/abstract.cfm?URL=josaa-31-4-794>.
- Atlanta Georgia Table Tennis Association, 2002. Training and Demonstrations. Available at:
http://www.ittf.com/Committees/Umpires_Referees/Video/training/1-Service/..%5Cindex.html [Accessed January 24, 2013].
- Ballard, D.H. & Brown, C.M., 1982. *Computer Vision*, Englewood Cliffs, New Jersey: Prentice-Hall Inc.
- Butterfly North America, 2013. Amicus 3000 Plus. Available at:
http://www.butterflyonline.com/inventoryD.asp?item_no=A3000P&CatId=%7BF9CA9690-D732-49DB-9AB5-EC463BB9C860%7D [Accessed January 9, 2013].
- Butterworth, S., 1930. On the theory of filter amplifiers. *Wireless Engineer*, 7(6), pp.536–541.

- Canny, J., 1986. A computational approach to edge detection. *IEEE transactions on pattern analysis and machine intelligence*, 8(6), pp.679–698.
- Canon, Canon HG10. Available at: http://www.canon-europe.com/for_home/product_finder/camcorders/high_definition_hd/hg10/ [Accessed May 20, 2013].
- CASIO Europe GmbH, Technical Specifications - EXILIM EX-ZR300 - EXILIM ZR-Series - Digital Cameras - Products - CASIO. Available at: <http://www.casio-europe.com/euro/exilim/exilimrserie/exzr300/specifications/> [Accessed May 20, 2014].
- Chakraborty, B. & Meher, S., 2011. 2D Trajectory-based Position Estimation and Tracking of the Ball in a Basketball Video. ... *Conference on Trends in Optics and* Available at: <http://dspace.nitrkl.ac.in:8080/dspace/handle/2080/1569> [Accessed August 18, 2012].
- Collins, H. & Evans, R., 2008. You cannot be serious! Public understanding of technology with special reference to "Hawk-Eye." *Public Understanding of Science*, 17(3), pp.283–308. Available at: <http://pus.sagepub.com/cgi/content/abstract/17/3/283>.
- Cyganek, B. & Siebert, J.P., 2009. *An Introduction to 3D Computer Vision Techniques and Algorithms*,
- D'Orazio, T. et al., 2009. A visual system for real time detection of goal events during soccer matches. *Computer Vision and Image Understanding*, 113(5), pp.622–632. Available at: <http://linkinghub.elsevier.com/retrieve/pii/S1077314208000441> [Accessed August 12, 2012].
- Denman, H., Rea, N., Kokaram, A., 2003. Content-based analysis for video from snooker broadcasts. *Computer Vision and Image Understanding*, 92(2-3), pp.176–195. Available at: <http://linkinghub.elsevier.com/retrieve/pii/S1077314203001188> [Accessed January 25, 2014].
- Duda, R.O. & Hart, P.E., 1973. *Pattern Classification and Scene Analysis*, Available at: <http://www.jstor.org/stable/1573081?origin=crossref>.
- Duda, R.O. & Hart, P.E., 1972. Use of the Hough Transform to detect lines and curves in pictures. *Communications of the ACM*, 15, pp.11–15.
- Earl, E., 2015. MediData, Garmin to partner to integrate wearables in clinical trials. *Becker's Healthcare*. Available at: <http://www.beckershospitalreview.com/healthcare-information-technology/mediadata-garmin-to-partner-to-integrate-wearables-in-clinical-trials.html> [Accessed March 26, 2015].
- eFunda Inc., General Trapzoid. Available at: <http://www.efunda.com/math/areas/trapezoid.cfm> [Accessed December 9, 2013].
- Encyclopedia Britannica Inc., 2014. The principle of equivalence. *Encyclopedia Britannica*. Available at: <http://www.britannica.com/EBchecked/topic/242523/gravity/61490/The-principle-of-equivalence> [Accessed July 10, 2014].
- EPFL, 2013. CVLAB | EPFL. Available at: <http://cvlab.epfl.ch/> [Accessed October 3, 2013].

- Erdem, C.E., Sankur, B. & Tekalp, A.M., 2004. Performance measures for video object segmentation and tracking. *IEEE Transactions on Image Processing*, 13(7).
- FIFA, 2014. Updated goal-line technology testing manual published - FIFA Quality Programme. Available at: <http://quality.fifa.com/en/News/Updated-goal-line-technology-testing-manual-published/> [Accessed June 4, 2014].
- Foresight Sports, 2013. Launch Monitors - Golf | Foresight Sports. Available at: <http://www.foresightsports.com/catalog/launch-monitor> [Accessed May 12, 2013].
- Fry, M.J. & Ohlmann, J.W., 2012. Introduction to the special issue on analytics in sports, part I: General sports applications. *Interfaces*, 42, pp.105–108.
- Fullen, R., 2010. Table Tennis: Fast Ball Sport? *Pro Table Tennis*. Available at: <http://protabletennis.net/content/table-tennis-fastest-ball-sport>.
- H. Ballard, D. & Brown, C.M., 1982. *Computer Vision*, Englewood Cliffs, New Jersey: Prentice-Hall Inc.
- Harris, C. & Stephens, M., 1988. A Combined Corner and Edge Detector. *Proceedings of the Alvey Vision Conference 1988*, pp.147–151. Available at: <http://www.bmva.org/bmvc/1988/avc-88-023.html>.
- Hart, S., 2014. Athletics to adopt Hawk-Eye after successful trials at the World Relay Championships in Nassau. *Telegraph Media Group*. Available at: <http://www.telegraph.co.uk/sport/othersports/athletics/10860908/Athletics-to-adopt-Hawk-Eye-after-successful-trials-at-the-World-Relay-Championships-in-Nassau.html> [Accessed May 30, 2014].
- Hawk-Eye Innovations, 2013. Cricket. Available at: <http://www.hawkeyeinnovations.co.uk/page/sports-officiating/cricket> [Accessed July 10, 2013].
- Hawk-Eye Innovations Ltd, 2013. Hawk-Eye. Available at: <http://www.hawkeyeinnovations.co.uk/> [Accessed May 28, 2013].
- Hawk-Eye Innovations Ltd, 2015. Tennis :: Hawk-Eye. Available at: <http://www.hawkeyeinnovations.co.uk/page/sports-officiating/tennis> [Accessed December 2, 2014].
- Hawk-Eye Innovations Ltd., 2009. Indian Wells 2009. Available at: [http://www.hawkeyeinnovations.co.uk/UserFiles/File/Ljubicic Challenge.pdf](http://www.hawkeyeinnovations.co.uk/UserFiles/File/Ljubicic%20Challenge.pdf) [Accessed August 23, 2012].
- Hawkins, P., 2013. Cricket may have embraced technology too quickly, says Hawk-Eye founder | Sport | The Guardian. *The Guardian*. Available at: <http://www.theguardian.com/sport/2013/aug/08/cricket-embraced-technology-hawkeye-founder> [Accessed March 31, 2014].

- Hoffmann, G., 2013. CIELab Color Space. Available at: <http://docs-hoffmann.de/cielab03022003.pdf> [Accessed March 6, 2014].
- Holt, N., 2011. India force ICC to “water down” Hawk-Eye system ahead of series in England. *The Telegraph*. Available at: <http://www.telegraph.co.uk/sport/cricket/international/england/8601423/India-force-ICC-to-water-down-Hawk-Eye-system-ahead-of-series-in-England.html> [Accessed June 23, 2012].
- Horn, B.K.P. & Schunck, B.G., 1981. Determining optical flow. *Artificial Intelligence*, 17(1-3), pp.185–203. Available at: <http://linkinghub.elsevier.com/retrieve/pii/0004370281900242>.
- Hu, J. et al., 2013. HDR Deghosting: How to deal with saturation? In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. pp. 1163–1170.
- Iino, Y. & Kojima, T., 2009. Kinematics of table tennis topspin forehands: effects of performance level and ball spin. *Journal of sports sciences*, 27(12), pp.1311–21. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/19746298> [Accessed January 25, 2014].
- Intel Corporation, 2009. ARK | Intel® Core™2 Duo Processor P7550 (3M Cache, 2.26 GHz, 1066 MHz FSB). Available at: http://ark.intel.com/products/42014/Intel-Core2-Duo-Processor-P7550-3M-Cache-2_26-GHz-1066-MHz-FSB [Accessed May 19, 2014].
- Itseez, 2015. OpenCV | OpenCV. Available at: <http://opencv.org/> [Accessed January 4, 2015].
- ITTF, ITTF Photo Gallery. Available at: http://www.ittf.com/_front_page/ittf1.asp?category=photo_gallery [Accessed February 10, 2014].
- ITTF, 2011. The Ball (Version for 40mm Balls) Technical Leaflet T3. *ITTF*. Available at: www.ittf.com/stories/pictures/T3_Ball40mm_BoD2011.PDF [Accessed March 5, 2012].
- ITTF, 2013a. 2013 List of ITTF Approved Balls. *List of ITTF Approved Balls*. Available at: http://www.ittf.com/stories/Pictures/Balls_03_13.pdf [Accessed October 24, 2013].
- ITTF, 2013b. *International Table Tennis Federation Handbook 2013-2014* 2013th–2014th ed., Available at: http://www.ittf.com/ittf_handbook/ittf_hb.html [Accessed March 3, 2013].
- ITTF, 2014a. *Handbook for MATCH OFFICIALS* 15th ed., Available at: http://www.ittf.com/ittf_handbook [Accessed February 12, 2014].
- ITTF, 2014b. THE LAWS OF TABLE TENNIS. In *ITTF Handbook 2013 - 2014*. ITTF, pp. 24–31, Available at: http://www.ittf.com/ittf_handbook/ittf_hb.html [Accessed September 22, 2014].
- Joblove, G.H. & Greenberg, D., 1978. Color spaces for computer graphics. *ACM SIGGRAPH Computer Graphics*, 12(3), pp.20–25. Available at: <http://portal.acm.org/citation.cfm?id=807362> [Accessed May 18, 2014].

- JVC Professional Europe Ltd, 2012. Siemens Star Chart. Available at:
<http://whitfors.myweb.port.ac.uk/wp-content/uploads/Siemens-Star-Chart-Focus.pdf>
[Accessed December 18, 2013].
- Kaewtrakulpong, P. & Bowden, R., 2001. An Improved Adaptive Background Mixture Model for Real-time Tracking with Shadow Detection. *Advanced Video Based Surveillance Systems*, pp.1–5.
- Kalman, R.E., 1960. A New Approach to Linear Filtering and Prediction Problems. *Journal of Basic Engineering*, 82, p.35.
- Kanungo, T. et al., 2002. An efficient k-means clustering algorithm: analysis and implementation. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 24(7), pp.881–892.
- Koninklijke Philips N.V., 2014. PCVC820K/P00 PC USB Video Camera, ToUcam II - Philips Support. Available at: http://www.p4c.philips.com/cgi-bin/dcbint/cpindex.pl?ctn=PCVC820K/P00&scy=hu&slg=en&mid=Link_Software [Accessed October 12, 2014].
- Lee, K.T. & Xie, W., 2004. THE VARIATION IN SPINS PRODUCED BY SINGAPORE ELITE TABLE TENNIS PLAYERS ON DIFFERENT TYPES OF SERVICE. *ISBS*.
- Leveaux, R., 2010. Facilitating Referee's Decision Making in Sport via the Application of Technology. *Communications of the IBIMA*, pp.1–8.
- Lucas, B.D. & Kanade, T., 1981. An iterative image registration technique with an application to stereo vision. *Proceedings of the 7th international joint conference on Artificial intelligence (IJCAI'81)*, Volume 2, pp.674–679. Available at:
http://www.ri.cmu.edu/pub_files/pub3/lucas_bruce_d_1981_1/lucas_bruce_d_1981_1.ps.gz
[Accessed January 18, 2012]
- Marr, D. & Hildreth, E., 1980. Theory of edge detection. *Proceedings of the Royal Society of London. Series B, Containing papers of a Biological character. Royal Society (Great Britain)*, 207, pp.187–217.
- McGuinness, K., 2009. Image segmentation, evaluation, and applications. *Electronic Engineering*. Available at: <http://doras.dcu.ie/14998/> [Accessed June 30, 2013].
- McHugh [1], S., Understanding Depth of Field in Photography. Available at:
<http://www.cambridgeincolour.com/tutorials/depth-of-field.htm> [Accessed April 26, 2014].
- McHugh [2], S., Digital Camera Image Noise: Concept and Types. Available at:
<http://www.cambridgeincolour.com/tutorials/image-noise.htm> [Accessed September 9, 2013].
- Mezaris, V., Kompatsiaris, I. & Strintzis, M.G., 2004. Video Object Segmentation Using Bayes-Based Temporal Tracking and Trajectory-Based Region Merging. *IEEE Transactions on Circuits and Systems for Video Technology*, 14(6), pp.782–795. Available at:
<http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=1302160> [Accessed January 18, 2012].

- Microsoft, Nokia Lumia 900 Specifications. Available at: <http://www.microsoft.com/en-gb/mobile/phone/lumia900/specifications/> [Accessed February 10, 2014].
- Modi, K.P., Sahin, F. & Saber, E., 2005. An Application of Human Robot Interaction: Development of a Ping-Pong Playing Robotic Arm. *2005 IEEE International Conference on Systems, Man and Cybernetics*, 2, pp.1831–1836. Available at: <http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=1571413> [Accessed January 9, 2013].
- Morris, T., 2004. *Computer Vision and Image Processing* First., Basingstoke: Palgrave Macmillan.
- Munsell, A.H., 1905. *A Color Notation*, G. H. Ellis Company. Available at: <http://books.google.com/books?id=PgcCAAAAYAAJ> [Accessed June 17, 2014].
- Munsell Color, 2013. ANSI Color Standards | Munsell Color System; Color Matching from Munsell Color Company. Available at: <http://munsell.com/color-products/color-standards/munsell-ansi-color-standards/> [Accessed December 3, 2013].
- NCH Software, 2013. Prism Video Converter Software. Available at: <http://www.nchsoftware.com/prism/index.html> [Accessed June 30, 2013].
- Nixon, Mark; Aguado, A., 2008. *Feature Extraction and Image Processing* Second Edi., Oxford: Academic Press.
- O'Donovan, P., 2005. Optical Flow: Techniques and Applications. *International Journal of Computer Vision*, pp.1–26. Available at: <http://www.dgp.toronto.edu/~donovan/stabilization/opticalflow.pdf> [Accessed May 18, 2014].
- Okumura, K., Oku, H. & Ishikawa, M., 2011. High-speed gaze controller for millisecond-order pan/tilt camera. *Robotics and Automation* (..., 1, pp.6186–6191. Available at: http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=5980080 [Accessed August 20, 2012].
- Oldham, K. et al., 2015. Experiments in the Application of Computer Vision for Ball and Event Identification in Indoor Sports. In S. Phon-Amnuaisuk & T. W. Au, eds. *Computational Intelligence in Information Systems SE - 27*. Advances in Intelligent Systems and Computing. Springer International Publishing, pp. 275–284. Available at: http://dx.doi.org/10.1007/978-3-319-13153-5_27 [Accessed February 22, 2014].
- Olympus, E-PL1. Available at: http://www.olympus.co.uk/site/en/archived_products/cameras_1/pen_1/e_pl1/e_pl1_main.pdf [Accessed June 11, 2013a].
- Olympus, ZUIKO DIGITAL ED 14-42mm 1:3.5-5.6 - Digital SLR Lenses - Olympus. Available at: http://www.olympus.co.uk/site/en/c/lenses/digital_slr_lenses/standard_lenses/zuiko_digital_ed_14_42mm_13556/index.html [Accessed May 20, 2014b].
- Olympus America Inc., Archived Products > C-5050 Zoom. Available at: http://www.olympusamerica.com/cpg_section/cpg_archived_product_details.asp?id=890 [Accessed March 9, 2013].

- Pallavi, V. et al., 2008. Ball detection from broadcast soccer videos using static and dynamic features. *Journal of Visual Communication and Image Representation*, 19(7), pp.426–436. Available at: <http://linkinghub.elsevier.com/retrieve/pii/S1047320308000527> [Accessed August 12, 2012].
- Pantofaru, C., 2005. A Comparison of Image Segmentation Algorithms. *Robotics*, 18, pp.123–130. Available at: http://www.ri.cmu.edu/pub_files/pub4/pantofaru_caroline_2005_1/pantofaru_caroline_2005_1.pdf [Accessed March 24, 2012].
- Photron USA, 2014. Photron - High Speed Camera Products. Available at: <http://www.photron.com/?cmd=products> [Accessed August 12, 2014].
- Pingle, K.K., 1969. Visual Perception by a Computer. In A. Grasselli, ed. *Automatic Interpretation and Classification of Images*. Academic Press.
- Ponce, J. & Forsyth, D., 2012a. *Computer vision: a modern approach*, Pearson Education Limited. Available at: <http://www.inria.fr/centre/paris-roquencourt/actualites/computer-vision-a-modern-approach> [Accessed May 28, 2013].
- Ponce, J. & Forsyth, D., 2012b. *Computer vision: a modern approach*, Harlow: Pearson Education Limited.
- Prewitt, J.M.S., 1970. Object enhancement and extraction. *Picture Processing and Psychopictorics*, pp.75–149.
- Rea, N., Dahyot, R. & Kokaram, A., 2004. Modelling High Level Structure in Sports with Motion Driven HMMS. *ICASSP*, III, pp.621–624.
- Roberts, L.G., 1965. Machine perception of three-dimensional solids. In *Optical and Electro-optical Information Processing*. pp. 159–197. Available at: <http://oai.dtic.mil/oai/oai?verb=getRecord&metadataPrefix=html&identifier=AD0413529> [Accessed March 18, 2012].
- Ruiz, N.A., 2010. *Tracking of a Basketball Using Multiple Cameras*,
- Rusdorf, S. & Brunnett, G., 2005. Real time tracking of high speed movements in the context of a table tennis application. *Proceedings of the ACM symposium on Virtual reality software and technology - VRST '05*, p.192. Available at: <http://portal.acm.org/citation.cfm?doid=1101616.1101657> [Accessed October 1, 2014].
- Schulte, M., 2013. *Real-time feature extraction from video stream data for stream segmentation and tagging*.
- Sight Machine Inc, SimpleCV. Available at: <http://simplecv.org/> [Accessed August 18, 2012].
- Sigma Imaging UK, Sigma Imaging UK | 24-70mm F2.8 IF EX DG HSM. Available at: <http://www.sigma-imaging-uk.com/24-70mm-f28-if-ex-dg-hsm> [Accessed May 20, 2014].

- Smith, A., 2012. CC Sabathia: Analyzing Why the New York Yankees Ace Is One of MLB's Elite Arms | Bleacher Report. *Bleacher Report*. Available at: <http://bleacherreport.com/articles/1418495-cc-sabathia-analyzing-why-the-new-york-yankees-ace-is-one-of-mlbs-elite-arms> [Accessed November 28, 2012].
- Sony Corporation, HDR-AS15 Specifications | Action Cam Video Cameras | Sony UK. Available at: <https://developer.sony.com/devices/cameras/sony-action-cam-with-wi-fi-hdr-as15/> [Accessed February 4, 2014].
- SportVision Inc., 2012. <http://www.sportvision.com/>. Available at: <http://www.sportvision.com/> [Accessed September 22, 2012].
- SportVU, 2012. SportVU. *Stats LLC*. Available at: <http://www.sportvu.com> [Accessed January 5, 2014].
- STATS LLC, 2015. NBA Player Speed. Available at:
http://stats.nba.com/tracking/#!/player/speed/?sort=AV_SPD&dir=1 [Accessed April 2, 2015].
- STATS LLC, 2013. STATS. Available at: <http://www.stats.com/> [Accessed April 2, 2013].
- Stauffer, C. & Grimson, W.E.L., 1999. Adaptive background mixture models for real-time tracking. *Proceedings 1999 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Cat No PR00149*, 2(c), pp.246–252. Available at:
<http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=784637> [Accessed February 12, 2012].
- Sun, L. et al., 2009. Ball's flight trajectory prediction for table-tennis game by humanoid robot. In *2009 IEEE International Conference on Robotics and Biomimetics ROBIO*. Ieee, pp. 2379–2384. Available at: <http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=5420828> [Accessed May 3, 2014].
- Table Tennis England, 2010. Umpire Location. Available at: <http://tabletennisengland.co.uk/our-sport/our-sport-subpages/volunteering/about-volunteering/> [Accessed February 10, 2014].
- Takeuchi, T., Kobayashi, Y., Hiruta, S., & Yuza, N., 2002. The effect of the 40 mm diameter ball on table tennis rallies by elite players. *International Journal of Table Tennis Sciences*, 4(5), pp.267–277.
- Tan, R.T., Nishino, K. & Ikeuchi, K., 2004. Separating reflection components based on chromaticity and noise analysis. *IEEE transactions on pattern analysis and machine intelligence*, 26(10), pp.1373–9. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/15641724> [Accessed June 3, 2013].
- Teachabarikiti, K., Chalidabhongse, T.H. & Thammano, A., 2010. Players tracking and ball detection for an automatic tennis video annotation. In *11th International Conference on Control, Automation, Robotics and Vision, ICARCV 2010*. pp. 2491–2494 [Accessed May 18, 2014].
- Technical Standardization Committee on AV & IT Storage Systems and Equipment, 2002. *Exchangeable Image File Format, Version 2.2*, Japan Electronics and Information Technology Industries Association. Available at:

- <http://web.archive.org/web/20131018091152/http://exif.org/Exif2-2.PDF> [Accessed November 16, 2014].
- The Iffmpeg Team, 2013. iMediaHUD. Available at: <http://www.ffmpeg.com/iMediaHUD/> [Accessed January 12, 2013].
- The International Table Tennis Federation, 2014. *Handbook for Match Officials* 15th Editi., ITTF. Available at: http://www.ittf.com/URC/PDF/HMO_15th_edition.pdf [Accessed June 23, 2014].
- The MathWorks, I., Convert RGB colormap to HSV colormap - MATLAB rgb2hsv - MathWorks United Kingdom. Available at: <http://uk.mathworks.com/help/matlab/ref/rgb2hsv.html> [Accessed March 23, 2014].
- The MathWorks Inc., Computer Vision System Toolbox - MATLAB & Simulink - MathWorks United Kingdom. Available at: <http://uk.mathworks.com/products/computer-vision/index.html> [Accessed November 18, 2014a].
- The MathWorks Inc., MATLAB - The Language of Technical Computing. Available at: <http://www.mathworks.com/products/matlab/> [Accessed June 14, 2014b].
- The MathWorks Inc., regionprops. Available at:
<http://www.mathworks.co.uk/help/releases/R2012a/toolbox/images/ref/regionprops.html#bqkf8ih> [Accessed November 13, 2014].
- The Moving Picture Experts Group, 2014. MPEG-4 | MPEG. Available at:
<http://mpeg.chiariglione.org/standards/mpeg-4> [Accessed May 15, 2014].
- Theobalt, C., Albrecht, I. & Magnor, M., 2004. Pitching a Baseball — Tracking High-Speed Motion with Multi-Exposure Images. *Control*, pp.540–547.
- Turberville, J., 2004. How fast does a Table Tennis ball really go? Available at:
<http://www.jayandwanda.com/tt/speed.html> [Accessed May 30, 2013].
- Velammal, B.L. & Kumar, P.A., 2010. An Efficient Ball Detection Framework for Cricket. *IJCSI International Journal of Computer Science*, 7(3), pp.30–35. Available at:
<http://www.ijcsi.org/papers/7-3-2-30-35.pdf> [Accessed July 2, 2013].
- Vicon Motion Systems Ltd., Vicon | Systems. Available at: <http://www.vicon.com/System/TSeries> [Accessed May 5, 2013].
- Wang, D., 1998. on Watersheds and Temporal Tracking. *IEEE Transactions on Circuits and Systems*, 8(5), pp.539–546 [Accessed August 23, 2013].
- Weisstein [1], E.W., Eccentricity. Available at: <http://mathworld.wolfram.com/Eccentricity.html> [Accessed April 13, 2014].
- Weisstein [2], E.W., Rose. Available at: <http://mathworld.wolfram.com/Rose.html> [Accessed April 14, 2015].

- Weszka, J., 1978. A survey of threshold selection techniques. *Computer Graphics and Image Processing*, 7(2), pp.259–265. Available at:
<http://www.sciencedirect.com/science/article/pii/0146664X78901168> [Accessed June 5, 2012].
- Winter, D.A., 2009. *Biomechanics and Motor Control of Human Movement* Fourth Edi., John Wiley & Sons.
- Wolfram [1], Euclidean Distance. Available at:
<https://reference.wolfram.com/language/ref/EuclideanDistance.html> [Accessed April 13, 2013a].
- Wolfram [2], Manhattan Distance. Available at:
<https://reference.wolfram.com/language/ref/ManhattanDistance.html> [Accessed April 13, 2013b].
- Wong, Patrick, 2009. Identifying table tennis balls from real match scenes using image processing and artificial intelligence techniques. *International Journal of Simulation Systems, Science & Technology*, 10(7), pp.6–14. Available at: <http://oro.open.ac.uk/22642/> [Accessed October 18, 2012].
- Wong, Patrick K.C., 2007. Developing an Intelligent Assistant for Table Tennis Umpires. *First Asia International Conference on Modelling & Simulation (AMS'07)*, pp.340–345. Available at:
<http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=4148683> [Accessed August 12, 2012].
- Wong, Patrick K.C., 2008. Developing an intelligent table tennis umpiring system: identifying the ball from the scene. *AICMS 08. Second Asia International Conference* Available at:
http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=4530517 [Accessed October 18, 2013].
- Wong, Patrick and Dooley, Laurance, 2010. High-motion table tennis ball tracking for umpiring applications. In *IEEE 10th INTERNATIONAL CONFERENCE ON SIGNAL PROCESSING PROCEEDINGS*. ieee, pp. 2460–2463. Available at:
<http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=5657001> [Accessed August 12, 2012].
- Wood, R.J., 2008. Snicko-Meter. *Topend Sports*. Available at:
<http://www.topendsports.com/sport/cricket/equipment-sncko-meter.htm> [Accessed December 10, 2012].
- Xu, W. et al., 2013. A novel star image thresholding method for effective segmentation and centroid statistics. *Optik-International Journal for Light and Electron Optics*, 124(20), pp.4673–4677.
- Yan, F., Christmas, W. & Kittler, J., 2005. A Tennis Ball Tracking Algorithm for Automatic Annotation of Tennis Match. *Proceedings of the British Machine Vision Conference 2005*, pp.67.1–67.10. Available at: <http://www.bmva.org/bmvc/2005/papers/paper-150.html> [Accessed December 10, 2014].

- Yan, F., Christmas, W. & Kittler, J., 2008. Layered data association using graph-theoretic formulation with applications to tennis ball tracking in monocular sequences. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 30, pp.1814–1830.
- Yoshida, K. et al., 2014. The rotation speed of the service ball delivered by world-class table tennis players. *Taiikugaku kenkyu (Japan Journal of Physical Education, Health and Sport Sciences)*, 59(1), pp.227–236.
- Yu, X. et al., 2004. A TRAJECTORY-BASED BALL DETECTION AND TRACKING ALGORITHM IN BROADCAST TENNIS VIDEO Xinguo Yu”, Chern-Horng Simh, Jenny Ran Wang’, and Loong Fah Cheongb. *Processing*, pp.1049–1052.
- Yu, X. et al., 2003. Trajectory-based ball detection and tracking with applications to semantic analysis of broadcast soccer video. *Proceedings of the eleventh ACM international conference on Multimedia - MULTIMEDIA ’03*, pp, p.11. Available at: <http://portal.acm.org/citation.cfm?doid=957013.957018> [Accessed April 12, 2012].
- Yu, X. & Farin, D., 2005. Current and Emerging Topics in Sports Video Processing. *Event (London)*.
- Zhang, H., Fritts, J.E. & Goldman, S.A., 2008. Image segmentation evaluation: A survey of unsupervised methods. *Computer Vision and Image Understanding*, 110, pp.260–280.
- Zhang, Y.J. & Gerbrands, J.J., 1994. Objective and quantitative segmentation evaluation and comparison. *Signal Processing*, 39, pp.43–54.
- Zhang, Z.Z.Z., Xu, D.X. De & Tan, M.T.M., 2010. Visual Measurement and Prediction of Ball Trajectory for Table Tennis Robot. *Ieee Transactions On Instrumentation And Measurement*, 59(12), pp.3195–3205. Available at: <http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=5454397> [Accessed October 5, 2012].

10. Appendices

A. Event state table

ID	State Name	Conditions	Action(s)
E1	Return right→left	Entry: E10 Exit: {E14, E20} Velocity check $(vx_k < 0) \wedge (vx_p > 0) \wedge (\Delta v > 146\%)$ Acceleration $ ax_k > ax_p \wedge (\Delta a > 189\%)$ Bearing $(180 < b_k < 360) \wedge (0 < b_p < 180) \wedge (\Delta b > 111^\circ)$ X location $x_k > x_n$ Zone D	Reset vx_p, ax_p, bx_p Increment rally count Player/Team 1
E2	Return left→right	Entry: E7 Exit: {E13, E19} Velocity check $(v_k > -v_p) \wedge (vx_p < 0) \wedge (\Delta v > 146\%)$ Acceleration $ ax_k > ax_p \wedge (\Delta a > 189\%)$ Bearing $(0 < b_k < 180) \wedge (180 < b_p < 360) \wedge (\Delta b > 111^\circ)$ X location $x_k < x_n$	Reset vx_p, ax_p, bx_p Increment rally count Player/Team 2
E3	Service throw right→left	Entry: Null Exit: E5 Velocity check $(vy_p = 0) \wedge (vy_k > 0)$ Acceleration $ ay_k > ay_p \wedge ax_k = 0$ Bearing $(b_k = 0) \wedge (b_p = 0)$ X location $x_k < (x_n + \frac{l}{2})$	Apogee check (>16cm) Reset vx_p, ax_p, bx_p

E4	Service throw left→right	Entry: Null Exit: {E3, E6} Velocity check ($\text{vy}_p = \mathbf{0}$) \wedge ($\text{vy}_k > \mathbf{0}$) Acceleration ($ ay_k > ay_p $) \wedge ($ax_k = 0$) Bearing ($b_k = 0$) \wedge ($b_p = 0$) X location $x_k < (x_n - \frac{l}{2})$	Apogee check (>16cm) Reset $vx_p, ax_p bx_p$
E5	Service hit right→left	Entry: E3 Exit: E8 Velocity check ($\text{vx}_k < \mathbf{0}$) \wedge ($\text{vx}_p > \mathbf{0}$) \wedge ($\Delta v > 560\%$) Acceleration ($ ax_k < ax_p $) \wedge ($\Delta a > 412\%$) Bearing ($180 < b_k < 360$) \wedge ($b_p = 0$) \wedge ($\Delta b > 205^\circ$) X location $x_k > (x_n + \frac{l}{2})$	Reset $vx_p, ax_p bx_p$
E6	Service hit left→right	Entry: E3 Exit: E9 Velocity check ($\text{vx}_k > \mathbf{0}$) \wedge ($\text{vx}_p = \mathbf{0}$) \wedge ($\Delta v > 560\%$) Acceleration ($ ax_k < ax_p $) \wedge ($\Delta a > 412\%$) Bearing ($0 < b_k < 180$) \wedge ($b_p = 0$) \wedge ($\Delta b > 205\%$) X location $x_k < (x_n - \frac{l}{2})$	Reset $vx_p, ax_p bx_p$
E7	Left table bounce right→left,	Entry: E20 Exit: E2 Velocity check ($\text{vy}_k < \mathbf{0}$) \wedge ($\text{vy}_p > \mathbf{0}$) \wedge ($\Delta v > 49\%$) Acceleration No test Bearing ($270 < b_k < 360$) \wedge ($180 < b_p < 270$) \wedge ($\Delta b > 58\%$) X location $(x_n - \frac{l}{2}) < x_k < x_n$	Reset $vx_p, ax_p bx_p$

E8	Right table bounce right→ left	Entry: E5 Exit: {E19, E12} Velocity check $(vy_k < 0) \wedge (vy_p > 0) \wedge (\Delta > 560\%)$ Acceleration No test Bearing $(270 < b_k < 360) \wedge (180 < b_p < 270) \wedge (\Delta b > 58\%)$ X location $x_n < x_k < (x_n + \frac{l}{2})$	Reset $vx_p, ax_p bx_p$
E9	Left table bounce left→ right	Entry: E6 Exit: {E2, E11} Velocity check $(vy_k < 0) \wedge (vy_p > 0) \wedge (\Delta > 560\%)$ Acceleration No test Bearing $(0 < b_k < 90) \wedge (90 < b_p < 180) \wedge (\Delta b > 58\%)$ X location $(x_c - \frac{l}{2}) < x_k < x_c$	Reset $vx_p, ax_p bx_p$
E10	Right table bounce left→right	Entry: E19 Exit: E1 Velocity check $(vy_k < 0) \wedge (vy_p > 0) \wedge (\Delta v > 560\%)$ Acceleration No test Bearing $(0 < b_k < 90) \wedge (90 < b_p < 180) \wedge (\Delta b > 58\%)$ X location $x_c < x_k < (x_c + \frac{l}{2})$	Reset $vx_p, ax_p bx_p$
E11	Net (let) service left→right	Entry: E9 Exit: {E3, E4, Fault (let)} Velocity check $\Delta v > 560\%$ Acceleration $\Delta a > 134\%$ Bearing $(0 < b_k < 180) \wedge (0 < b_p < 180) \wedge (b_k > b_p) \wedge (\Delta b > 12^\circ)$ X location $x_k \approx x_c$	Reset $vx_p, ax_p bx_p$ Increment Player/Team 2 let fault count. Check let fault count >1, Player/Team 1 score

E12	Net (let) service right→left	Entry: E8 Exit: {E3, E4, Fault (let)} Velocity check $\Delta v > 560\%$ Acceleration $\Delta a > 134\%$ Bearing $(180 < b_k < 360) \wedge (180 < b_p < 360)$ $\wedge (b_k < b_p) \wedge (\Delta b > 12^\circ)$ X location $x_k \approx x_c$	Reset vx_p, ax_p, bx_p Increment Player/Team 1 let fault count. Check let fault count >1, Player/Team 2 score
E13	Net (rally) left→right	Entry: E2 Exit: {E3, E4, Fault (net)} Velocity check $\Delta v > 560\%$ Acceleration $\Delta a > 134\%$ Bearing $(0 < b_k < 180) \wedge (0 < b_p < 180)$ $\wedge (b_k > b_p) \wedge (\Delta b > 12^\circ)$ X location $x_k \approx x_c$	Reset vx_p, ax_p, bx_p Increment Player/Team 1 score
E14	Net (rally) right→left	Entry: E1 Exit: {E3, E4, Fault (net)} Velocity check $\Delta v > 560\%$ Acceleration $\Delta a > 134\%$ Bearing $(180 < b_k < 360) \wedge (180 < b_p < 360)$ $\wedge (b_k < b_p) \wedge (\Delta b > 12^\circ)$ X location $x_k \approx x_c$	Reset vx_p, ax_p, bx_p Increment Player/Team 2 score
E15	Doubles line fault left→right	Entry: E4 Exit: {E3, E4, Fault (doubles line)} Velocity check $(vy_k < 0) \wedge (vy_p > 0) \wedge (\Delta v > 560\%)$ Acceleration No test Bearing $(0 < b_k < 90) \wedge (90 < b_p < 180) \wedge (\Delta b > 58\%)$ Y location <i>tbc</i> X location $(x_c - \frac{l}{2}) < x_k < x_c$	Reset vx_p, ax_p, bx_p Increment doubles fault count. Check for increment Player/Team 1 score

E16	Doubles line fault right→left	Entry: E3 Exit: {E3, E4, Fault (doubles line)} Velocity check $(vy_k < 0) \wedge (vy_p > 0) \wedge (\Delta v > 560\%)$ Acceleration No test Bearing $(180 < b_k < 360) \wedge (180 < b_p < 360)$ $\wedge (b_k > b_p) \wedge (\Delta b > 58\%)$ Y location tbc X location $x_c < x_k < (x_c + \frac{l}{2})$	Reset vx_p, ax_p, bx_p Increment doubles fault count. Check for increment Player/Team 1 score
E17	Table edge right→left	Entry: E20 Exit: {E1, E2, E3, E4} Velocity check $\Delta v > 24\%$ Acceleration $\Delta a > 196\%$ Bearing $(180 < b_k < 360) \wedge (180 < b_p < 360)$ $\wedge (b_k > b_p) \wedge (\Delta b > 53\%)$ X location $x_k \neq x_n$	Reset vx_p, ax_p, bx_p
E18	Table edge left→right	Entry: E19 Exit: {E1, E2, E3, E4} Velocity check $\Delta v > 24\%$ Acceleration $\Delta a > 196\%$ Bearing $(0 < b_k < 180) \wedge (0 < b_p < 180)$ $\wedge (b_k > b_p) \wedge (\Delta b > 53\%)$ X location $x_k \neq x_n$	Reset vx_p, ax_p, bx_p
E19	Over the net left→right	Entry: {E2, E9} Exit: E18, E10 Velocity check No test Acceleration No test Bearing $(0 < b_k < 180) \wedge (0 < b_p < 180)$ X location $x_k \geq x_c \wedge (x_p < x_c)$	

E20	Over the net right→left	Entry: {E1,E8} Exit: {E17, E7} Velocity check No test Acceleration No test Bearing $(180 < b_k < 360) \wedge (180 < b_p < 360)$ X location $x_k \leq x_c \wedge (x_p > x_c)$	
E21	Service net collision left→right	Entry: E6 Exit: E3, E4, Fault (let) Velocity check $\Delta v > 560\%$ Acceleration $\Delta a > 134\%$ Bearing $(0 < b_p < 180) \wedge (180 < b_k < 360)$ $\wedge (b_k < b_p) \wedge (\Delta b > 12^\circ)$ X location $x_k < x_c$	Reset $v x_p, a x_p b x_p$ Fault (net) Player 2
E22	Service net collision right→left	Entry: E5 Exit: E3, E4, Fault (let) Velocity check $\Delta v > 560\%$ Acceleration $\Delta a > 134\%$ Bearing $(0 > b_p > 180) \wedge (180 > b_k > 360)$ $\wedge (b_k > b_p) \wedge (\Delta b > 12^\circ)$ X location $x_k > x_c$	Reset $v x_p, a x_p b x_p$ Fault (net) Player 1
E23	Rally net colli- sion left→right	Entry: Return left→right Exit: {E3, E4, Fault (net)} Velocity check $\Delta v > 560\%$ Acceleration $\Delta a > 134\%$ Bearing $(0 < b_p < 180) \wedge (180 < b_k < 360)$ $\wedge (b_k < b_p) \wedge (\Delta b > 12^\circ)$ X location $x_k < x_c$	Reset $v x_p, a x_p b x_p$ Increment Player/Team 2 score

E24	Rally net Collision right→left	Entry: Return right→left Exit: E3, E4, Fault (net)} Velocity check $\Delta v > 560\%$ Acceleration $\Delta a > 134\%$ Bearing $(0 > b_p > 180) \wedge (180 > b_k > 360)$ $\wedge (b_k > b_p) \wedge (\Delta b > 12^0)$ X location $x_k > x_c$	Reset vx_p, ax_p, bx_p Increment Player/Team 1 score
-----	--------------------------------	---	--

B. Defining a FAIL

A FAIL occurs when either no candidates are detected even though the object of interest is in the image, or too many candidate objects are detected to making conclusions from the data unreliable. When applying the EMS, either of these instances is critical to the success of the method, due to its implicit focus on key match-play events. As such, a FAIL is given a relatively large weighting (100) and must be used consistently.

Determining when no candidate is successfully detected is straightforward: the candidate array will have zero dimensions. However, algorithms may be implemented to deliberately detect multiple candidates (usually for further filtering or, on occasion, when there is intentionally more than one candidate to detect). It has been found, during experimental analysis for this paper, that on occasions a large number of candidates are detected and no amount of careful tuning of the filter parameters reduces the number of candidates being detected. Take, for example, the results from an extreme case 700-P-10 when an image with a ball in a complex environment (Figure 10-1) is processed by initially converting to grey-tone, then applying a binary threshold, followed by the Zero Cross edge detector, filling holes in the boundaries and then finally ball classification using CHT. The resulting ball location detection output for this frame is presented in Figure 10-3.

It can be clearly seen that there are too many candidate objects detected for useful data (in this single representative image, over 630 candidate balls were found). Furthermore, over the entire video sequence from which this image was taken, the processing speed was an average of 239 seconds per frame. This is clearly too long for real time processing, no matter how efficient the algorithm code could be improved. Manually adjusting the parameters through several iterations in an attempt to find a suitable solution did not yield an improved result. The combination of source data and algorithm used was not successful. The outcome for this algorithm with this video sequence was a FAIL and recorded as such for the EMS review.



Figure 10-1 Original unprocessed frame containing a single ball



Figure 10-2 700-P-10 Failed detection using Zerocross and CHT

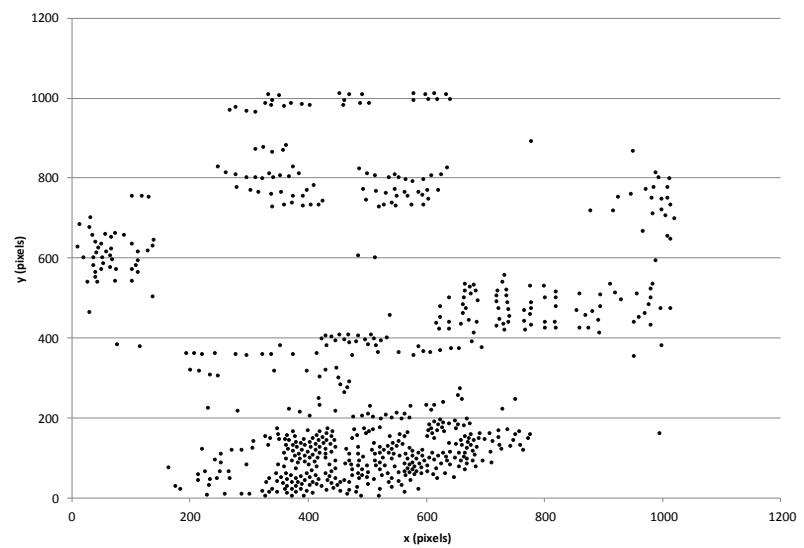


Figure 10-3 Failed detection using Zerocross and CHT

Experimental results from algorithms used to validate the rigor of the EMS have shown that this only occurs with edge detectors in complex scenes (foreground extraction techniques are not vulnerable to this problem).

C. Core camera settings

Throughout this experimental research, many combinations of camera configurations, often running in parallel, have been used to generate the substantial video data set. However, the number of primary camera settings (including frame rate and shutter speed) has been minimised to enable ease of direct comparison and the development of future research. These core camera settings are listed in Table 10-1 below.

Table 10-1 Core camera settings

ID	Camera	FPS	Shutter	Optional Lens
CS01	Canon HG10	60	1/200	Built-in
CS02	Casio Exilim™ ZR-300	120	1/300	Zuiko Digital ED 14-42mm
CS03	Casio Exilim™ ZR-300	240	1/300	Zuiko Digital ED 14-42mm
CS04	Olympus E-PL1	30	1/300	Zuiko Digital ED 14-42mm
CS05	Olympus E-PL1	30	1/200	Zuiko Digital ED 14-42mm
CS06	Philips ToUcam II	15	1/100	Built-in
CS07	Photron Fastcam Ultima APX	50	1/1000	Sigma 24-70mm
CS08	Photron Fastcam Ultima APX	125	1/300	Sigma 24-70mm
CS09	Photron Fastcam Ultima APX	250	1/1000	Sigma 24-70mm
CS10	Photron Fastcam Ultima APX	250	1/2000	Sigma 24-70mm
CS11	Photron Fastcam Ultima APX	250	1/10000	Sigma 24-70mm
CS12	Photron Fastcam Ultima APX	250	1/500	Sigma 24-70mm
CS13	Photron Fastcam Ultima APX	250	1/250	Sigma 24-70mm
CS14	Photron Fastcam Ultima APX	500	1/500	Sigma 24-70mm
CS15	Photron Fastcam Ultima APX	1000	1/1000	Sigma 24-70mm
CS16	Photron Fastcam Ultima APX	20000	1/20000	Sigma 24-70mm
CS17	Sony HDR AS-15	30	1/200	Built-in
CS18	Sony HDR AS-15	60	1/200	Built-in

NB: Both the Apple iPhone™ 5 and Nokia Lumia™ 900 have non-configurable settings within dynamic ranges, with values set by proprietary software depending on the subject being imaged. As such the configurations for these devices are different for each video sequence and cannot be generalised.

Thirty three video sequences (of the 276 total, generated in the work described in Chapter 3) are individually identified in this report to aid with differentiating data outputs. For these references to individual video sequences, their corresponding camera core configuration ID is cross referenced in Table 10-2 below as a summary. Where multiple video sequences are discussed and resulting data is combined, the details of the proportions of grouped camera configurations are provided in the body text.

Table 10-2 Video sequence to camera configuration ID cross reference

Video	Core camera configuration ID
100-P-01	CS14
100-P-02	CS12
100-P-03	CS12
100-P-06	CS12
100-P-10	CS09
100-C-14	CS01
200-P-03	CS08
300-P-01	CS13
300-P-03	CS13
300-P-04	CS13
300-P-06	CS13
300-P-08	CS14
400-P-03	CS12
500-P-02	CS13
500-P-03	CS13
500-P-08	CS13
500-P-09	CS13
500-P-11	CS13
600-P-06	CS14
600-P-07	CS13
600-P-08	CS12
600-P-09	CS15
600-P-10	CS14
600-P-12	CS13
700-P-06	CS13
700-P-09	CS13

700-P-10	CS13
700-P-12	CS13
700-P-13	CS13
700-P-14	CS13
700-P-15	CS13
700-P-16	CS13
700-P-17	CS13

D. Scholarly contributions

- i. Kevin M. Oldham, Paul Wai Hing Chung, Eran A. Edirisinghe, Benjamin J. Halkon. A validated approach to automatically quantifying table tennis dynamics from a single video source (abstract). 9th International Symposium on Computer Science in Sport, Marmara University, Istanbul, Turkey. 19-23 June 2013.
- ii. Kevin M. Oldham, Paul Wai Hing Chung, Eran A. Edirisinghe, Benjamin J. Halkon. Experiments in the Application of Computer Vision for Ball and Event Identification in Indoor Sports. The Fourth International Neural Network Society Symposium, Institut Teknologi Brunei, 7-9 November 2014. In S. Phon-Amnuaisuk & T. W. Au, eds. *Computational Intelligence in Information Systems SE - 27*. Advances in Intelligent Systems and Computing. Springer International Publishing 2015, pp. 275–284. Available at: http://dx.doi.org/10.1007/978-3-319-13153-5_27

Conference papers submitted:

- iii. Submitted the paper entitled “Table tennis and Computer Vision: A monocular event classifier” to be published for the 10th International Symposium on Computer Science in Sport (ISCSS 2015), September 09-11, 2015, Loughborough, UK.

Journal papers and online access materials in preparation include the following:

- iv. A summary paper containing the full research overview of the event detection framework.
- v. The paper entitled “Using the ball as a scale of reference in monocular computer vision interrogations of table tennis dynamics”.
- vi. The TRASE software is to be made available online. Currently written as a Matlab application, a web interface is to be developed to enable TRASE to be used online as open source. Specifications of public modular interface requirements are also to be detailed to allow for new CV algorithms to be written and uploaded by external researchers.
- vii. The library of video sequences, along with the design of experiments, camera settings and event descriptions, is to be made available in an open access online database.

{intentionally blank}