



Camera calibration and configuration for estimation of tennis racket position in 3D.

ELLIOTT, Nathan.

Available from the Sheffield Hallam University Research Archive (SHURA) at:

<http://shura.shu.ac.uk/20196/>

A Sheffield Hallam University thesis

This thesis is protected by copyright which belongs to the author.

The content must not be changed in any way or sold commercially in any format or medium without the formal permission of the author.

When referring to this work, full bibliographic details including the author, title, awarding institution and date of the thesis must be given.

Please visit <http://shura.shu.ac.uk/20196/> and <http://shura.shu.ac.uk/information.html> for further details about copyright and re-use permissions.

Learning and IT Services
Collegiate Learning Centre
Collegiate Crescent Campus
Sheffield S10 2BP

1446

102 156 567 9



REFERENCE

ProQuest Number: 10700841

All rights reserved

INFORMATION TO ALL USERS

The quality of this reproduction is dependent upon the quality of the copy submitted.

In the unlikely event that the author did not send a complete manuscript and there are missing pages, these will be noted. Also, if material had to be removed, a note will indicate the deletion.



ProQuest 10700841

Published by ProQuest LLC (2017). Copyright of the Dissertation is held by the Author.

All rights reserved.

This work is protected against unauthorized copying under Title 17, United States Code
Microform Edition © ProQuest LLC.

ProQuest LLC.
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106 – 1346

Camera calibration and configuration for estimation of tennis racket position in 3D

Nathan Elliott

A thesis submitted in partial fulfilment of the requirements of
Sheffield Hallam University
for the degree of Doctor of Philosophy

July 2015

Abstract

Previously, stereo camera systems have been used to track markers attached to the racket frame, allowing for racket position to be measured in three-dimensions (3D). Typically, markers are manually selected on the image plane but this can be time consuming and inaccurate. The purpose of this project was to develop and validate a markerless method to estimate 3D racket position using a camera.

The method relies on a silhouette of a racket captured with a camera whose relative pose (rotation and translation) is unknown. A candidate relative pose is used to measure the inconsistency between the silhouette and a set of racket silhouettes captured with a fully calibrated camera (known intrinsic and extrinsic parameters). The measure of inconsistency can be formulated as a cost function associated with the candidate relative pose. By adjusting parameters of the candidate relative pose to minimise the cost, an accurate estimation for the true 3D position of the racket can be made.

Previous studies have found that silhouette-based pose optimisation methods depend on accurate camera calibration and silhouette extraction. Therefore, a repeatable and accurate camera calibration method to provide the relative pose of a camera with respect to a racket was developed. To facilitate silhouette extraction, the racket was painted black and a backlight was used.

Synthetic camera poses and silhouette views associated with a 3D racket model were generated in Blender v2.70 and used to determine the optimum fully calibrated set configuration for a racket. A laboratory-based fully calibrated set (LFCS) consisting of 21 camera poses in a semispheric configuration was created. On average, using this set, racket position was reconstructed to within ± 2 mm. This included systematic error arising from the calibration and error in the segmentation of silhouette boundaries. The maximum reconstruction error was 5.3 mm.

Further synthetic testing demonstrated the methods ability to estimate 3D racket position during simulated real-play conditions. For racket silhouette orientations that simulated strokes expected to occur in tennis between 0 and 90°, mean RMSE for reconstruction of coordinates on the racket face plane was 1.5 ± 1.8 mm. An RMSE of 2 mm was obtained from a camera positioned alongside the net, 14 m from the racket. Finally, this same camera position estimated 3D racket position to an accuracy of 1.9 ± 0.14 mm using a fully calibrated set containing randomly orientated camera poses, during a simulated serve.

This project developed and validated a novel markerless method to estimate 3D tennis racket position. A calibration method to obtain the relative pose of a camera with respect to a racket is presented and an appropriate configuration for a fully calibrated set is determined. The method has potential to be used alongside existing ball trajectory analysis tools to provide unprecedented information about player performance and to enhance tennis broadcasts. Future research should use the recommendations made in this project to inform and assist the development of the method for application during real tennis-play conditions.

Keywords: markerless; silhouette; pose; measure of inconsistency; calibration

Acknowledgements

I would like to thank my supervisors Dr Tom Allen, Dr Simon Choppin and Dr Simon Goodwill for their continual support, guidance and enthusiasm throughout this project. Their experience has proved both insightful and motivating. My thanks are also extended to everyone within the Centre for Sports Engineering Research at Sheffield Hallam University, in particular Terry Senior for providing help and advice during data collection.

Finally, I would like to thank my family and Rheannon for their support and encouragement to finish the project.

Contents

1	Introduction.....	1
1.1	Motivation for the research	1
2	Literature review	3
2.1	Introduction.....	3
2.1.1	History of tennis	3
2.1.2	Science and tennis.....	4
2.1.3	Tennis racket performance	5
2.2	Player testing and motion analysis	7
2.2.1	Intrusive techniques	7
2.2.2	Active markers	7
2.2.3	Passive markers	10
2.2.4	Non-intrusive techniques.....	13
2.3	Markerless motion capture.....	16
2.3.1	Visual hulls.....	16
2.3.2	Intrinsic and extrinsic camera parameters	17
2.3.3	Silhouette extraction	20
2.3.4	Quantifying silhouette quality	23
2.3.5	Measuring silhouette consistency.....	24
2.3.6	The epipolar tangency constraint	24
2.3.7	A measure of inconsistency based on epipolar tangents.....	26
2.3.8	Visual hulls for biomechanical applications	28
2.3.9	Single view fitting	32
2.3.10	Conclusion	36
2.4	Chapter findings	37
2.4.1	Conclusion	39
2.5	Aim and objectives	40
3	Intrinsic camera calibration	41
3.1	Introduction.....	41

3.2	Aim and objectives	41
3.3	Calibration procedure.....	42
3.3.1	Checkerboard manufacture.....	42
3.3.2	Camera settings.....	42
3.4	Recording calibration images	43
3.5	Summary	45
3.6	Intrinsic calibration error analysis.....	45
3.7	Method.....	45
3.8	Results.....	46
3.8.1	Image coverage	46
3.8.2	Lens distortion.....	47
3.8.3	The intrinsic model.....	48
3.8.4	Repeatability	48
3.8.5	Calibration error	50
3.9	Discussion.....	51
3.9.1	Repeatability	52
3.9.2	Calibration error	53
3.10	Conclusion	53
4	Calibration of a stereo system.....	54
4.1	Introduction.....	54
4.2	Aim and objectives	54
4.3	Stereo Imaging.....	55
4.4	Method.....	55
4.5	Results.....	57
4.5.1	Image coverage	57
4.5.2	Intrinsic parameters	57
4.5.3	Repeatability of the relative pose parameters	58
4.5.4	Accuracy	60
4.6	Discussion.....	61
4.6.1	Methodological considerations	61

4.6.2	Repeatability	61
4.6.3	Accuracy	62
4.7	Conclusion	63
5	Obtaining the pose of a camera relative to a racket.....	64
5.1	Introduction	64
5.2	Aim and objectives	64
5.3	Method	65
5.4	Relative pose of the master camera to the racket	69
5.5	Accuracy	71
5.5.1	Control point error	71
5.5.2	Relative camera pose error	72
5.5.3	Reconstruction error	73
5.6	Repeatability	73
5.6.1	Repeatability of the slave camera pose	73
5.7	Summary	75
5.7.1	Limitations	76
5.8	Conclusion	76
6	Measuring the consistency of racket silhouettes.....	77
6.1	Introduction	77
6.2	Aims and objectives	77
6.3	The epipolar tangency constraint for a pair of racket silhouettes.....	78
6.4	Epipoles inside silhouettes	79
6.5	Locating the epipolar tangency points	80
6.6	Determining racket tangency point correspondences	81
6.7	Summary	83
6.8	Synthetic data generation	83
6.8.1	The 3D racket model	83
6.8.2	Generation of synthetic camera parameters	84
6.9	Experiments	86
6.9.1	Discrete pixel rendering	86

6.9.2	Comparing racket silhouettes with simpler shapes	88
6.9.3	Silhouette resolution	89
6.10	Discussion.....	90
6.11	Conclusion	92
7	Optimising the configuration of a fully calibrated set of racket silhouettes using synthetic data.....	93
7.1	Introduction.....	93
7.2	Aim and objectives	93
7.3	Formulating the ETE as an objective function	94
7.4	Candidate relative poses	94
7.5	Method.....	98
7.6	Results.....	100
7.6.1	Dome shaped fully calibrated set configurations	101
7.6.2	Semi-circular configurations	103
7.6.3	Reconstructing the racket face plane	108
7.6.4	Quantification of removing a pose from a fully calibrated set.....	110
7.7	Discussion.....	110
7.7.1	Conclusion	112
8	Validation of racket pose estimates using a fully calibrated set created in the laboratory.....	113
8.1	Introduction.....	113
8.2	Aim and objectives	113
8.3	Creating the LFCS.....	113
8.3.1	Method.....	113
8.3.2	Results.....	115
8.4	Assessing the accuracy of the LFCS	116
8.4.1	Synthetic data generation	116
8.4.2	Method.....	116
8.4.3	Results.....	117
8.5	Racket position validation.....	118

8.5.1	Method.....	118
8.5.2	Results.....	119
8.6	Discussion.....	121
8.7	Conclusion	125
9	Estimating 3D racket position from different camera positions	127
9.1	Introduction	127
9.2	Aim and objectives	127
9.3	The fully calibrated set	127
9.4	Method.....	128
9.4.1	Camera orientation	128
9.4.2	Camera distance.....	130
9.5	Results.....	133
9.5.1	Orientation	133
9.5.2	Distance.....	135
9.5.3	Serve simulation	136
9.6	Discussion.....	138
9.6.1	Orientation	138
9.6.2	Distance.....	139
9.6.3	Application to tennis.....	143
9.7	Conclusion	144
10	Conclusions	145
10.1	Introduction.....	145
10.2	Summary of research	145
10.2.1	Camera calibration.....	145
10.2.2	Common local reference frame	146
10.2.3	Measuring silhouette consistency.....	146
10.2.4	Determining the configuration of a fully calibrated set	146
10.2.5	Laboratory-based fully calibrated set (LFCS)	147
10.2.6	Simulation of real play conditions.....	147

10.3	Conclusion	149
10.4	Future research.....	151
10.4.1	Future development in the laboratory	151
10.4.2	Application in real-play conditions	151
10.4.3	Programming language.....	153
10.4.4	Markerless 3D position estimation tool.....	153
10.4.5	Incorporation with existing tennis software	154
11	References	155
12	Personal bibliography	167
13	Other publications	168
14	Appendix	169
14.1.1	Calculation of internal camera parameters from Blender (v2.70)	169
14.1.2	MATLAB function to calculate the internal parameters for a camera in Blender	169

Table of figures

Figure 2.1 <i>Jeu de paume</i> (taken from Haake et al. 2007).	3
Figure 2.2 Wingfield's lawn tennis (taken from Haake et al. 2007).	4
Figure 2.3 Average speed of 20 fastest serves at Grand Slam events between 2002 and 2005 (taken from Miller 2006).	5
Figure 2.4 CODA active marker system on the racket (taken from Mitchell, Jones and King 2000).	8
Figure 2.5 The wireless 3D motion tracker produced by Xsens, showing a single sensor and multiple sensors being used to analyse the biomechanics of a baseball pitch.	8
Figure 2.6 The Babolat Play Pure Drive (Babolat 2014) with embedded sensors in the handle and Bluetooth and wired connections in the butt.	9
Figure 2.7 Approved PAT products: (a) SONY Smart Tennis Sensor 'pod' (adapted from SONY 2015), (b) ZEPP Tennis 'pod' (adapted from ZEPP 2015) and (c) Artengo Personal coach 'pod' (adapted from ARTENGO 2015).	10
Figure 2.8 Stereo camera setup used at the 2006 Wimbledon qualifying tournament (taken from Choppin et al. 2007a).	11
Figure 2.9 Photo showing the location of the 5 reflective tape markers on the racket frame (taken from Choppin et al. 2007a).	12
Figure 2.10 Example of a camera from the Motion Analysis Corporation with red light sensor and the Raptor system in action outdoors being used to track a golf shot.	13
Figure 2.11 An example of a line call review from Hawk-Eye (2015).	14
Figure 2.12 Camera locations used for analysis by SAGIT (taken from Martínez-Gallego et al. 2013).	15
Figure 2.13 The visual hull concept: (a) a duck viewed by two cameras, (b) two silhouette views of a duck, (c) visual cones corresponding to the two silhouettes, (d) the visual hull corresponding to the two silhouettes and (e) the visual hull corresponding to ten silhouettes (taken from Forbes 2007).	17
Figure 2.14 Pinhole camera geometry illustrating focal length and principal point which govern the projection of real world point Q in the image plane as q (taken from Bradski and Kaehler 2008).	18
Figure 2.15 The normalised 2D image coordinate P_c (camera coordinate) is related to the 3D (object) coordinate P_o by applying the rotation matrix R and translation vector t (taken from Bradski and Kaehler 2008).	20
Figure 2.16 (a) A B-spline snake is initialised close to the target silhouette, (b) points are sampled from each spline segment and a search for the silhouette edge along the direction normal to the local tangent at each sample point is performed. (c) The points	

are updated to attach to the silhouette edge that was found (adapted from Wong 2001).	21
Figure 2.17 An example of sub-pixel resolution silhouette boundary extraction. The pixel and sub-pixel resolution boundaries are shown in green and red respectively. Each vertex of the sub-pixel boundary lies on a pixel-length segment (shown in blue) associated with each edge of the pixel resolution boundary (taken from Forbes 2007).	22
Figure 2.18 The epipolar tangency constraint for a pair of fully calibrated silhouettes: (a) shows a frontal view and (b) a side view (taken from Forbes 2007).	24
Figure 2.19 The epipolar tangency constraint for a pair of duck silhouettes. Projection of a tangency point in (a) the left image is constrained to lie on the epipolar tangent line in (b) the right image, vice versa (taken from Forbes 2007).	26
Figure 2.20 Error associated with real camera parameters and segmented silhouette boundaries resulting in the epipolar tangency lines not projecting exactly through the tangency points in (a) the left and (b) the right images. The reprojection error is used to measure the degree of inconsistency for a silhouette pair (taken from Forbes 2007).	27
Figure 2.21 Positions of the eight synchronised cameras surrounding the player (taken from Sheets et al. 2011).	29
Figure 2.22 (a) Footage from camera four of a representative service trial, (b) corresponding visual hulls, (c) tracked subject-specific segment models and (d) correspondence between visual hull and model surfaces (taken from Sheets et al. 2011).	30
Figure 2.23 A 6-view fully calibrated silhouette set associated with a six-camera setup, the resulting visual hull model of the particle created by intersecting all the visual cones. The 'unknown' viewpoint associated with the particles pose is shown with respect to the fully calibrated set (adapted from Price and Morrison 2007).	33
Figure 2.24 Arbitrary camera pose numbers 12 to 15 estimated from the circular fully calibrated silhouette views 1 to 11 (taken from Wong and Cipolla 2004).	34
Figure 2.25 Images 1 to 3 are examples from the first checkerboard sequence used to obtain the intrinsic parameters. Images 4 to 6 are examples from the second sequence and were used to determine camera pose parameters (adapted from Tavares, Azevedo and Vaz (2008).	35
Figure 2.26 Camera calibration (intrinsic and extrinsic), determination of a common local reference frame and silhouette extraction are the main contributors to error in view fitting methods.	39
Figure 3.1 Chapter 3 calculated camera intrinsic parameters and quantified the associated error.	41

Figure 3.2 A checkerboard image (a) after semi-automatic intersection detection and (b) at a higher zoom level showing the intersections as red points within the search region.	44
Figure 3.3 Image showing the coverage achieved; each red dot represents a checkerboard intersection.	46
Figure 3.4 Illustration of the (a) radial and (b) tangential lens distortion; the numbers indicate the pixel displacement which is highest towards the image periphery.	47
Figure 3.5 Image showing the fit of the intrinsic model (red circle) with respect to the point detected by the software (green crosshair).	48
Figure 3.6 Pixel discrepancy between all reprojected intersection points and points detected by the software for one of the calibrations.	48
Figure 3.7 The distribution of focal length values for (a) f_x and (b) f_y across the ten intrinsic calibrations, showing the mean (circle), median, upper and lower quartiles and upper and lower deciles.	49
Figure 3.8 The distribution of principal point values for (a) c_x and (b) c_y across the ten intrinsic calibrations, showing the mean (circle), median, upper and lower quartiles and upper and lower deciles.	50
Figure 3.9 Pixel projection error in the u and v direction for each of the 10 intrinsic calibrations, showing the mean (circle), median, upper and lower quartiles and upper and lower deciles.	50
Figure 3.10 Mean reconstruction error \pm standard deviation error across the ten calibrations for coordinates on the calibration plane in the X, Y and resultant (R) directions.	51
Figure 4.1 Chapter 4 calculated the extrinsic parameters for a stereo pair and quantified the associated error.	54
Figure 4.2 A stereo system. The 3D position of the point P can be determined by triangulation of the matched 2D coordinates p_l and p_r (adapted from Bradski and Kaehler 2008).	55
Figure 4.3 An example of the image coverage achieved for one of the ten stereo calibration repetitions in the field of view of (a) the master camera (b) the slave camera.	57
Figure 4.4 Typical pixel displacement due to radial distortion for (a) the master camera and (b) the slave camera for one of the ten stereo calibrations.	58
Figure 4.5 Schematic of one of the ten completed stereo calibrations , showing the relative position of the slave with respect to the master camera and the checkerboards	58

Figure 4.6 Translation in the (a) X, (b) Y and (c) Z direction of the slave camera with respect to master camera estimated using the three intrinsic models, showing the mean (circle), median, upper and lower quartiles and upper and lower deciles.....	59
Figure 4.7 Rotation about the axis' of (a) X, (b) Y and (c) Z of the slave camera with respect to master camera calculated using the three intrinsic models, showing the mean (circle), median, upper and lower quartiles and upper and lower deciles.....	59
Figure 4.8 Mean reconstruction error \pm standard deviation for the geometry of the checkerboard in the X, Y and Z directions using a complete and 4th order radial distortion model.	60
Figure 4.9 Mean \pm standard deviation checkerboard square size (horizontal and vertical), measured using the MATLAB camera calibration software for the three intrinsic models. The black dashed line indicates the known square size of 50 mm. ...	61
Figure 5.1 Chapter 5 placed a camera and a racket in common local reference frame and quantified the associated error of the cameras pose.	64
Figure 5.2 The lower calibration plane was defined by coordinates P_1 to P_4 and P_5 to P_7 defined the upper calibration plane which aligned with the racket face plane.	66
Figure 5.3 Illustration of the angles measured showing (a) the intersecting planes, the angle offsets of the planes (b) parallel and (c) perpendicular with the racket face plane.	67
Figure 5.4 Laser scanning of the racket and calibration planes.....	67
Figure 5.5 Rendered laser scan of the racket and calibration planes. The long axis of the racket scan was intersected by two virtual perpendicular planes used to measure the angles (Plane YZ and Plane XZ).....	68
Figure 5.6 The cameras and the racket were defined in a common local reference frame using the seven orthogonal control points, P_1 to P_7	70
Figure 5.7 Seven manually digitised 2D image coordinates in the view of the master camera.	70
Figure 5.8 Bar graphs showing error in relative pose of the slave with respect to the master camera obtained through manual digitisation of 7, 4 and 3 control points.....	73
Figure 6.1 Chapter 6 used synthetic data to quantify the influence of discrete pixel rendering and extraction of silhouettes on the epipolar tangency error (ETE).	77
Figure 6.2 The epipolar tangency constraint for a pair of fully stereo calibrated racket silhouettes.	78
Figure 6.3 A tangency point p_{ijk} is constrained to lie on the line $F_{ij}p_{ijk}$ in the image plane of (a) C_1 and (b) C_2 (Figure 6.1).	78
Figure 6.4 Error associated with real camera parameters and silhouette boundaries results in the epipolar tangency lines not projecting exactly through the tangency points in the opposite image for (a) C_1 and (b) C_2 in Figure 6.2.	79

Figure 6.5 Upper and lower tangency point location for a racket silhouette.....	80
Figure 6.6 Illustration of the method used to locate the two tangencies. The angle on the image between the positive x-axis and the tangency line associated with all vertices in a silhouette boundary is calculated and placed in the respective quadrant. Vertices corresponding to the highest and lowest angles are the tangencies.	81
Figure 6.7 Handedness of the epipole with respect to the tangency point is used to decipher whether tangency correspondence is (a) opposite or (b) the same.....	82
Figure 6.8 The 3D racket model created in Blender (v2.70) showing (a) the template image used to form the shape of the frame and (b) the final rendered model.....	84
Figure 6.9 Synthetic relative master camera and slave camera poses with respect to the model racket in Blender scene.	85
Figure 6.10 Epipolar lines plotted using synthetic camera parameters and synthetic racket silhouette boundary.	86
Figure 6.11 Epipolar lines plotted using synthetic camera parameters and rendered silhouette boundaries. The epipolar lines did not project exactly through (a) the upper and (b) the lower tangency points due to the square shape of the pixels.	87
Figure 6.12 Silhouette outlines of (a) a single pixel, (b) a 2D plane with the same height and width as a racket and (c) the racket.	88
Figure 6.13 The RMS ETE (mm) across synthetic fully calibrated consisting of 72, 64, 40 and 20 camera poses using racket silhouettes rendered at successively lower resolution. The RMS ETE (mm) increases as resolution decreases.....	90
Figure 7.1 Chapter 7 used synthetic camera poses to quantify the influence of discrete pixel rendering and extraction of silhouettes on 3D racket position estimation.	93
Figure 7.2 A spherical coordinate (adapted from WolframMathWorld 2015).	96
Figure 7.3 Method used to assign each spherical coordinate a reference frame with respect to the racket model. (a) shows a side view and (b) shows a 3D view.	97
Figure 7.4 Candidate relative poses generated using spherical coordinates showing (a) a 3D view and (b) a top view.....	98
Figure 7.5 Illustration of a sample optimisation using showing (a) positions of the ground truth left camera (red), the initial candidate relative pose (blue) and the camera pose estimate (green), (b) the initial ETEs and (c) the optimised ETEs.....	100
Figure 7.6 Camera pose estimate error for translation in the (a) X, (b) Y and (c) Z directions using an 80 pose synthetic set. The largest errors corresponded to side views of the racket.	101
Figure 7.7 Silhouettes associated with (a) a frontal view and (b) a side view of the racket model.	102
Figure 7.8 Translation errors in the (a) X, (b) Y and (c) Z directions for camera pose estimates made using a 72 pose dome shaped fully calibrated set configuration.....	102

Figure 7.9 Translation errors in the (a) X, (b) Y and (c) Z directions for camera pose estimates made using a circular 18 pose fully calibrated set configuration.....	103
Figure 7.10 Translation errors in the (a) X, (b) Y and (c) Z directions for camera pose estimates made using a semi-circular fully calibrated set configuration containing 36 poses.....	104
Figure 7.11 Translation errors in the (a) X, (b) Y and (c) Z directions for camera pose estimates made using a semi-circular fully calibrated set configuration containing 27 poses.....	105
Figure 7.12 Translation errors in the (a) X, (b) Y and (c) Z directions for camera pose estimates made using a semi-circular fully calibrated set configuration containing 18 poses.....	106
Figure 7.13 Translation errors in the (a) X, (b) Y and (c) Z directions for camera pose estimates made using a semi-circular fully calibrated set configuration containing 9 poses.....	106
Figure 7.14 RMSE for reconstructed coordinates on the racket face plane in the X, Y and R directions using semi-circular fully calibrated sets with (a) 36, (b) 27, (c) 18 and (d) 9 poses.....	108
Figure 7.15 Resultant RMSE for coordinates on the racket face plane increases with camera pose estimate translation error in the (a) X, (b) Y and (c) Z direction.....	109
Figure 7.16 Resultant RMSE for coordinates on the racket face plane increases with camera pose estimate rotation error in the (a) X, (b) Y and (c) Z direction.	110
Figure 8.1 Chapter 8 calculated the total error resulting from calibration parameters and silhouette extraction on the ability of a view fitting method to estimate 3D racket position.	113
Figure 8.2 Approximate angles relative to the racket stringbed normal (0°) for the 21 slave camera poses.....	114
Figure 8.3 Schematic showing (a) a front view and (b) a top view of the master camera pose (red), the fully calibrated set of 21 slave camera poses (blue) and the racket in a common reference frame.....	116
Figure 8.4 Projected epipolar tangent lines on the image plane of camera pose 8 in Figure 8.3 (a) derived from the all camera poses. The data has some error therefore the lines are not exactly tangential.	117
Figure 8.5 Racket views corresponding to (a) camera pose 1 and (b) camera pose 13 from Figure 8.3. The green dots are exact projections of 3D points obtained from the racket scan and the red dots are projections of 3D coordinates reconstructed using pose estimates.	120
Figure 8.6 RMSE (mm) for reconstruction in the X, Y and Z directions of 3D points on the racket frame surface using camera pose estimates obtained from (a) LFCS	

parameters and synthetic silhouettes, (b) synthetic poses and rendered silhouettes and (c) the LFCS.	121
Figure 8.7 Camera pose estimate translation error in the X, Y and Z directions using rendered silhouettes of an irregular shaped plane with the same height and width as a standard size racket.....	125
Figure 9.1 The sixteen synthetic unknown views. (a) shows a front view and (b) shows a top view.	128
Figure 9.2 For each of the 16 unknown views in Figure 9.1 camera pose orientation was increased from 0 to 90° in 10 increments, forming ten sets of unknown views. The blue, red and green rectangles show the image plane at orientations of 0, 50 and 90° respectively.....	129
Figure 9.3 Racket silhouette images captured from view point 5 in Figure 9.1 at camera pose orientations of (a) 0, (b) 50 and (c) 90°.	129
Figure 9.4 Camera distances tested and their relative positions on the court.	130
Figure 9.5 View from of the camera pose circled red in Figure 9.4. The red dot indicates where the camera view was centred i.e. midway up the racket.	131
Figure 9.6 Simulated racket position during a serve around ball impact. The racket was rotated about an axis 762 mm from the butt in 2° increments. (a) shows a side view and (b) a front view of the simulated racket positions in Blender.....	132
Figure 9.7 Resultant reconstruction error on the racket face plane for unknown view orientations between 0 and 90° obtained using fully calibrated sets consisting of camera poses oriented upright and randomly.	133
Figure 9.8 Resultant RMSE for unknown views orientated at 20° obtained using fully calibrated sets containing camera poses orientated (a) upright and (b) randomly.	134
Figure 9.9 Resultant RMSE for unknown views orientated at 60° obtained using fully calibrated sets containing camera poses orientated (a) upright and (b) randomly.	135
Figure 9.10 Mean RMSE \pm standard deviation (mm) for reconstruction of coordinates in the Y, Z and R directions on the racket face plane from camera distances ranging between 2 to 14 m.	135
Figure 9.11 RMSE for reconstruction on the racket face plane in the (a) Y, (b) Z and (c) R directions for all the camera distances tested.	136
Figure 9.12 Camera pose estimates obtained during the simulated serve showing (a) a 3D view and (b) a side view.	137
Figure 9.13 Reconstruction error obtained using a fully calibrated set orientated upright in the (a) Y, (b) Z and (c) R directions and reconstruction error obtained using a fully calibrated set orientated randomly in the (d) Y, (e) Z and (f) R directions, during a simulated serve.....	138

Figure 9.14 Silhouette views associated with racket's pose at (a) -40° , (b) 0° and (c) 30°	142
Figure 9.15 The silhouette associated with ball impact when the racket is at 0° should be manually selected to be the first analysis frame. The method then works backwards or forwards to estimate racket position before and after impact respectively.	143
Figure 10.1 Errors that contributed toward 3D racket position estimation accuracy using the method.....	145

Nomenclature

ANOVA	Analysis of variance
C	Camera centre
CODA	Cartesian Optoelectronic Dynamic Anthropometer
COM	Centre of mass
(cx,cy)	Principle point
DLT	Direct Linear Transform
E	Essential matrix
e	Epipole
ETE	Epipolar tangency error
F	Fundamental matrix
FIFA	Fédération Internationale de Football Association
fsp	Frames per second
(fx,fy)	Focal length
ITF	International Tennis Federation
K	Calibration matrix
LED	Light Emitting Diode
LFCS	Laboratory-based fully calibrated set
MAC	Motion Analysis Corporation
MMC	Markerless motion capture
MOI	Moment of inertia
P	3D point in homogeneous coordinates $[x\ y\ z\ 1]^T$
p	2D normalised image point in homogeneous coordinates $[x\ y\ 1]^T$
PAT	Player Analysis Technology
Q	Quaternion vector
R	Rotation matrix
RMSE	Root-Mean Square Error
RMS ETE	Root-Mean Square Epipolar Tangency Error
t	Translation vector
$[t]_x$	Anti-symmetric matrix of a vector t

1 Introduction

The following chapters describe a 3.5 year study into the development of a novel markerless method to estimate 3D racket position.

1.1 Motivation for the research

Tennis technology has developed, which has had major influences on how the game is played. The reduction of racket mass and moment of inertia (MOI) in modern frames enable players to swing the racket faster (Mitchell, Jones and King 2000, Haake et al. 2007). Thus, in the hands of elite players, lighter rackets allow for higher rebound ball speed and accuracy (Mitchell, Jones and King 2000, Allen, Choppin and Knudson 2015). Furthermore, rebound ball speed has been found to be positively related to racket stiffness and inversely related to string tension (Allen, Haake and Goodwill 2011). However, the effect of racket properties on rebound ball dynamics is dependent on the impact conditions, speed, angle, spin and interactions with the player and stroke biomechanics (Allen, Choppin and Knudson 2015).

The International Tennis Federation (ITF) govern the game of tennis and implement rules in order to allow for its development but at the same time preserve the nature of the sport. A concern of the ITF is to maintain public and commercial interest, in order to prevent the demise of the sport due to lack of financial support. Current developments in computing power, motion analysis software, electronics and cameras, combined with a decrease in their cost has facilitated the development of systems that make it possible to collect valuable information about player performance (ITF 2014).

In 2014, the ITF created a new rule approving player analysis technology (PAT) during competitive tennis. These PAT products allow for large amounts of information to be gathered in real time that previously would have taken days or weeks or was simply unavailable. Providing players, coaches and manufacturers with an opportunity to obtain valuable feedback about performance during play. This could be feedback obtained by measuring something related to the player's performance when using a particular racket. Which in turn could be used to assist a player's verbal feedback and inform the development of bespoke racket designs.

Previously, tennis racket motion in practice play has been measured using cameras to track markers attached to the frame (Choppin 2008). Typically, markers are manually selected on the image plane but this can be time consuming and inaccurate. Furthermore, players are often reluctant to have modifications made to their rackets i.e. attachment of markers or sensors to the frame, because these usually increase the

moment of inertia of a racket, which can affect players swing mechanics. Brody (2000) found that elite players can distinguish differences in racket swing weight moment of inertia as small as 2.5%. However, the influence on racket MOI is dependent on the type of marker used. Mitchell, Jones and King (2000) attached LED active markers and power modules to the frame which increased racket MOI by 7%. In contrast, Choppin et al. (2007a) attached five low mass tape markers to the frame which increased racket MOI by less than 1%. Nevertheless, marker attachment can distract players and is not practical during real-play conditions.

To eliminate drawbacks associated with marker-based techniques, markerless motion capture (MMC) systems have been applied in tennis in order to measure stroke kinematics (Sheets et al. 2011). These MMC systems rely on obtaining a participant's visual hull (Laurentini 1994) obtained from multiple silhouette views. However, this requires multiple high-speed cameras, which are expensive and impractical to be setup in close proximity to the player. Nevertheless, markerless systems should continue to be developed and tested for sports applications (Sheets et al. 2011). In tennis, the ability to measure performance without the need to modify the playing environment would provide unprecedented access to information regarding ball-racket interactions.

Camera systems that measure tennis ball speed (Hawk-Eye), as well as spin (SpinDoctor Kelley 2011) without interfering with the playing environment already exist. In tennis, Hawk-Eye is used extensively by the media, providing match statistics and ball trajectory visualisations during line calling decisions. A non-intrusive method capable of measuring racket velocities and angles could be used alongside Hawk-Eye to enhance media broadcasts. It could also be used to further knowledge of ball and racket movements during match play and player tests in order to inform laboratory-based test method's (Allen, Choppin and Knudson 2015). Further work is therefore needed to investigate methods capable of measuring racket movements that do not require invasive and costly modification to the playing environment. This project will develop and validate a novel markerless method to estimate 3D racket position. Furthermore, it will provide recommendations to assist future development of the method for application in real-play conditions.

2 Literature review

2.1 Introduction

This chapter provides a review of existing literature related to the aim and objectives identified in section 1.2. Specifically, this chapter identifies the gap in the literature which the methods developed in this thesis are intended to fill.

2.1.1 History of tennis

Sports played in Europe in the twelfth century contributed to the evolution of the game of tennis (Clerici 1976). *Jeu de Paume* (meaning 'game of palm') was one such game, played in courtyards with a ball hit by hand (Haake et al. 2007) (Figure 2.1).

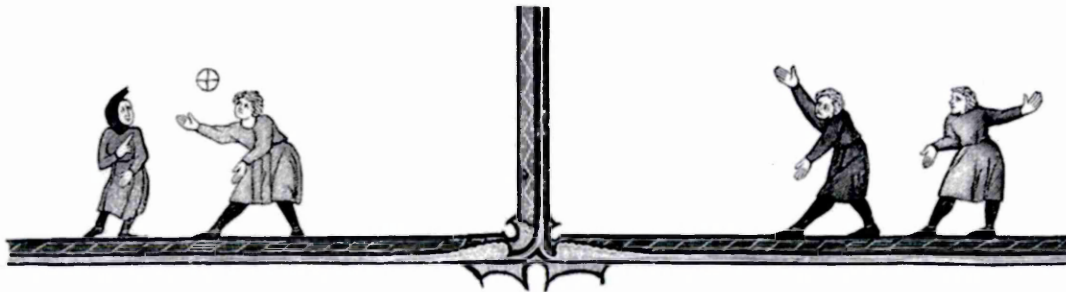


Figure 2.1 *Jeu de paume* (taken from Haake et al. 2007).

The game later evolved into *Royal Tennis* or *Real Tennis*, which is acknowledged as the forerunner for modern tennis (ITF 2015). In real tennis, the court is asymmetric with a net and is enclosed by walls. A cork based ball which can be played off the walls is used and rackets are shaped asymmetrically to allow players to hit balls close to the floor and corners of the court.

Lawn tennis evolved from real tennis which was 'invented' by Major Wingfield in 1874 (Clerici 1979). The game was more accessible because it could be played on any size or shape lawn and only required rackets, a net and ball (Figure 2.2). Lawn tennis was introduced to the members of the All England Croquet Club in 1875. Following the games popularity, the club changed its name to the All England Lawn Tennis and Croquet Club in 1877. The first competitive lawn tennis tournament was held at Wimbledon that same year, where the familiar rectangular court and scoring system was formally introduced.

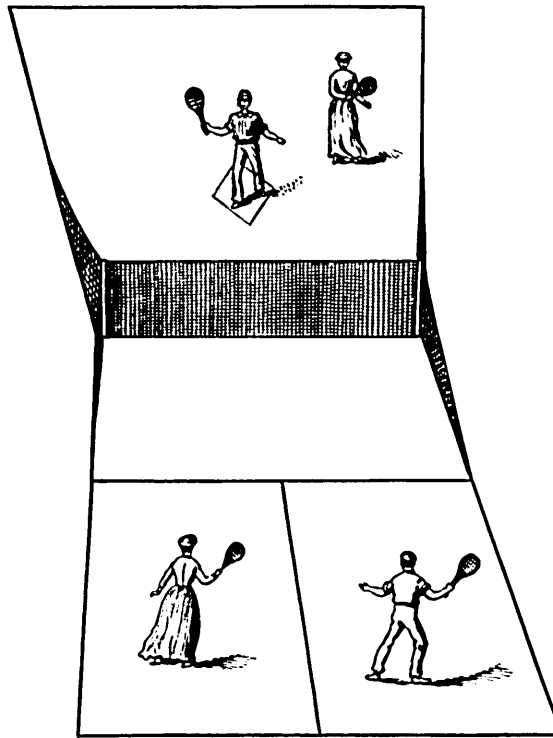


Figure 2.2 Wingfield's lawn tennis (taken from Haake et al. 2007).

The popularity of lawn tennis continued to increase such that the Lawn Tennis Association (LTA) was formed in 1888 and the International Tennis Federation (ITF) in 1913. Today, tennis is played recreationally by over 75 million people worldwide (Pluim et al. 2007), supporting a large manufacturing industry (ITF 2015). Professional tennis is popular among spectators and receives monetary investment in sponsorship, prize money and technology (ITF 2015).

2.1.2 Science and tennis

Lord Rayleigh's account of the irregular flight of a tennis ball (Rayleigh 1877) is the earliest example of a scientific investigation into the game of tennis. Published in 1877, this illustrates that the game of tennis has motivated scientific thinking for almost as long as the game has been played. Since then, the development of research into the science of tennis has continued to progress; measuring and understanding the physical principles that govern the game have become important for players, coaches, governing bodies and manufacturers alike.

Second to player ability, competitive tennis today is largely driven by technological innovation; if a manufacturer can produce superior equipment then there are financial gains to be made. No piece of equipment used in tennis has received more attention and been at the forefront of technological innovation than the racket (Coe 2000). The larger head size and higher stiffness of modern rackets have allowed players to hit

shots faster and with greater accuracy (Kotze, Mitchell and Rothberg 2000), effectively increasing the speed of the game (Haake et al. 2007). To the extent it was believed the server was gaining a significant advantage over their opponent, identified by an increase in the number of tie breaks (Brody 1990, Haake et al. 2007).

Miller (2006) quantified the 20 fastest serves recorded at Grand Slam events between 2002 and 2005 (Figure 2.3). Serve speeds increased year-on-year, highlighting the impact of player ability and racket development on serve speed (Miller 2006). Haake et al. (2007) produced a historical account of technological developments of the racket on serve speed. Simulations by Haake et al. (2007) suggest that due to the developments of rackets, serve speeds increased by 17.5% between the 1870s and 2007, with about 25% of the increase occurring since the 1970s. Haake et al. (2007) also state the 21 rule changes regarding the racket implemented by the ITF since 1978. Highlighting that previous rule changes by the ITF have been reactionary in nature. Therefore, it is important that thorough research is undertaken to assess how technological advances influence tennis performance.

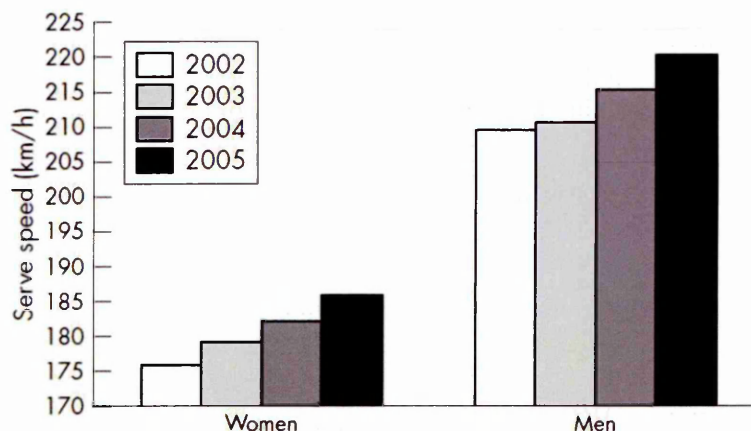


Figure 2.3 Average speed of 20 fastest serves at Grand Slam events between 2002 and 2005 (taken from Miller 2006).

2.1.3 Tennis racket performance

Many studies into racket performance fail to take into account how the racket will be used by the player. For instance, early physical testing used impacts normal to the racket face (Cross 1999, Dignall and Haake 2000, Goodwill 2002), which are not representative of actual tennis strokes. In fact, player testing conducted by Choppin, Goodwill and Haake (2011) showed impact angles deviated from the racket face plane normal by up to 33°. Few studies have attempted to quantify the effect of racket parameters on typical tennis shots using realistic impact conditions. Nevertheless, research into tennis biomechanics has received considerable attention over the last 25

years (Elliott, Marsh and Blanksby 1986, Elliott, Marshall and Noffal 1995, Elliott 2006, Sheets et al. 2011). Providing an understanding of the movements required from a player to generate powerful shots.

Previous studies have used cameras in order to measure tennis related movements (Elliott, Marsh and Blanksby 1986, Elliott, Marshall and Noffal 1995, Elliott 2006, Sheets et al. 2011, Choppin, Goodwill and Haake 2011). Stereo calibration and high-speed camera techniques have been applied in tennis (Choppin, Goodwill and Haake 2011), football (Driscoll et al. 2014) and the hammer throw (Kelley 2014), providing a means of investigating object motion and impact location in 3D. Allen, Haake and Goodwill (2009) used this technique to obtain data for oblique spinning impacts on a freely suspended racket, which was used to validate a finite element model.

Allen, Haake and Goodwill (2011) demonstrated that ball rebound speed increased by 37% with racket mass, in the range of 279-418g and by 31% as the centre of mass (COM) moved from 29.9 to 39.6 cm from the butt. Indicating that ball velocity increases with racket moment of inertia (MOI) about an axis 0.1 m from the butt, agreeing with Cross and Bower (2006), Cross and Nathan (2009), Cross (2010). Allen, Haake and Goodwill (2011) also found that rebound topspin increased by 23% with racket mass and 21% with COM position. However, the findings of Allen, Haake and Goodwill (2011) are limited to the impacts simulated; there were a number of necessary assumptions in the finite element model. These included geometrical and material simplifications which are difficult to quantify.

In tennis, MOI is a measure of a racket's resistance to angular acceleration about an axis 0.1 m from the butt (Cross and Bower 2006) and is one of the limiting factors for maximum swing speed. It has been found that the swing-speed a player can generate decreases with racket MOI (Cross and Bower 2006, Mitchell, Jones and King 2000, Schorah, Choppin and James 2015). Therefore, it is clear that the MOI of a racket interacts with a player in a manner in which the simulations conducted by Allen, Haake and Goodwill (2011) could not account for. A reduction of racket mass and corresponding reduction of moment of inertia may not result in decreased ball velocities because players are capable of swinging lighter rackets faster (Mitchell, Jones and King 2000, Whiteside et al. 2014).

Cross and Bower (2006) and Schorah, Choppin and James (2015) tested the hypothesis of an inverse relationship between swing-speed and MOI of an implement. A series of restricted swinging motion trials were performed using eight identical rods

with a common mass but variable MOI. It was found that for all participants, swing-speed decreased with respect to MOI according to a power relationship. However, data collected by Schorah, Choppin and James (2015) was limited to a specific motion and many complex movements take place during a tennis stroke. Therefore, measurement of racket velocity in real-play conditions could be used to help determine the effect of racket properties (particularly inertial) on performance more effectively. Such information could be used to inform future MOI studies whose aim is to predict a player's swing speed for racket fitting purposes.

2.2 Player testing and motion analysis

To further understand the effect of sports equipment on player performance, it is imperative that the test conditions are representative of play. Player testing is considered to be a suitable method for determining ball and racket movements during real tennis-play conditions. The need for accurate player data has led to the development of many different methods for obtaining such information, which will be described below.

Motion analysis has become an extensive branch of research in tennis; since the game is based on movement of the racket, the ball and the players. Motion analysis techniques can be split into two groups:

1. Intrusive techniques: motion analysis where the player or environment is altered in some way.
2. Non-intrusive techniques: motion analysis that does not require modification of the player or environment.

2.2.1 Intrusive techniques

Tracking object movement in sport often involves attaching some kind of marker to the player or their equipment. The markers themselves are classified as active or passive. An active marker transmits a signal as a way of identifying its location. A passive marker is tracked by picking up reflections from the markers surface.

2.2.2 Active markers

Active markers transmit a signal in order to be identified. Mitchell, Jones and King (2000) used a cartesian optoelectronic dynamic anthropometer (CODA) active marker system to compare the effect of racket inertial properties on serve speed. Two power module receivers each wired to two infrared LED active markers were attached around the racket frame (Figure 2.4). The CODA system can make real-time 3D position

measurements of the LED markers to within ± 0.1 mm in the horizontal and vertical directions and to within ± 0.6 mm of depth. However, attachment of the CODA system caused a total increase in mass of approximately 17 % (44 g) and corresponding increase in MOI of approximately 0.0024 kg m^2 to the racket. In general, players achieved faster head speeds with lighter rackets and exhibited an approximately 10% drop in rebound ball velocity for a 12% increase in racket MOI. Brody (2000) found that elite players can distinguish differences in racket MOI (which is equivalent to the 'swing weight' of a racket) as small as 2.5%. Furthermore, the CODA system alone cannot provide information about ball movements.

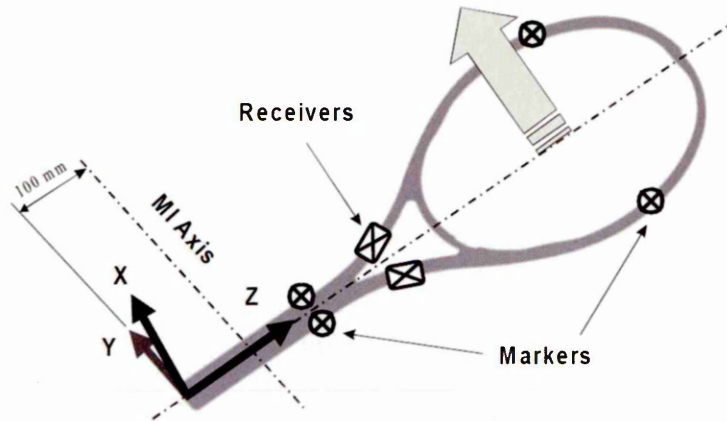


Figure 2.4 CODA active marker system on the racket (taken from Mitchell, Jones and King 2000).

Sensors that measure motion can also be considered as active markers. Motion tracking systems from Polhemus (*Polhemus*, 2012) allow real time 3D tracking with six degrees-of-freedom by tracking the position and orientation of each sensor. The wireless system produced by Xsens (*Xsens*, 2012) is a full-body motion capture suit incorporating sensors which can be tracked in six degrees-of-freedom. It can be used outdoors with a wireless range of up to 150 m. The system also includes extra sensors to attach to sports equipment. Figure 2.5 shows one of the wireless motion tracker sensors produced by Xsens and a suit fitted with multiple sensors for biomechanical analysis of a baseball pitch.

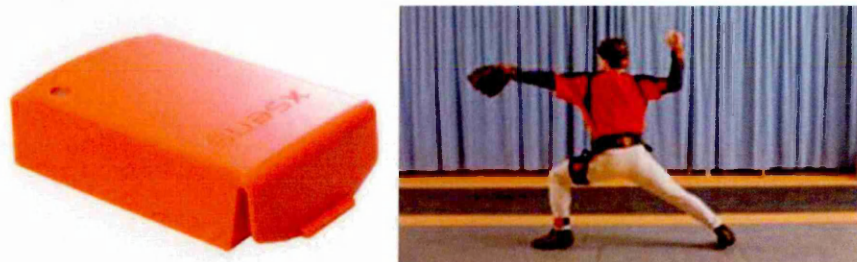


Figure 2.5 The wireless 3D motion tracker produced by Xsens, showing a single sensor and multiple sensors being used to analyse the biomechanics of a baseball pitch.

The combination of bulky sensors, transmitting equipment and batteries of the Xsens are not practical to measure racket movements; attachment to the frame would increase mass and MOI affecting player's swing mechanics. However, current developments in electronics and motion analysis software, coupled with a decrease in their cost has facilitated the development of bespoke sensor-based systems for tennis. Recently, a number of off-the-shelf sensor based products that measure racket and ball movements have been approved by the ITF as player analysis technology (PAT), and can be used during competitive tennis.

The Babolat Play Pure Drive (Babolat 2014) includes electronic sensors embedded in the racket handle to measure its orientation, acceleration and vibration (Figure 2.6). Data collected by the racket are sent to an auxiliary device e.g. a Smartphone or personal computer, via a wireless Bluetooth or wired connection. The Babolat Play Pure Drive (Babolat 2014) has a strung mass of 317.5 g, which is within 3 g of the same racket model without the embedded sensors. An independent report by tennis-technology (2015) showed that the Babolat Play sensors and plastic butt cap have a combined mass of 30 g. Babolat (2015) claim to measure racket power, type and number of strokes, impact location and ball spin using the sensors. However, there has been no published work on the validation of the racket and ball movements measured using the Babolat Play Pure Drive technology.



Figure 2.6 The Babolat Play Pure Drive (Babolat 2014) with embedded sensors in the handle and Bluetooth and wired connections in the butt.

The SONY Smart Tennis Sensor 'pod' (SONY 2015) and the ZEPP Tennis 'pod' (ZEPP 2015) shown in Figure 2.7 (a) and (b) respectively, attach to the butt of the racket. Both these PAT products have a mass of 8 g, with an additional 13 g for the Flex Mount required for attachment of the ZEPP Tennis 'pod' (Figure 2.7 (b)). These devices provide coaching information, including ball impact location on the stringbed, racket trajectory, racket swing speed, ball velocity (estimation) and spin, which is available on an auxiliary device. Figure 2.7 (c) shows an Artengo Personal coach 'pod' (mass 24g)

containing sensors which attach to the racket throat. Data regarding racket orientation, acceleration and ball impact location is sent to a purpose-built watch.

These PAT products are useful for coaching purposes and can provide valuable information about performance during strokes. However, players are prohibited from access to the performance data (the auxiliary device) during competitive play (ITF 2014). Furthermore, each of these PAT products increases racket mass and MOI which may make elite players reluctant to use them during competitive play. Finally, it is unlikely ball movements are measured accurately using these sensor-based PAT products; the sensors would either need to be fitted inside the ball or cameras used to track the ball. For these reasons sensor based systems will not be considered further.

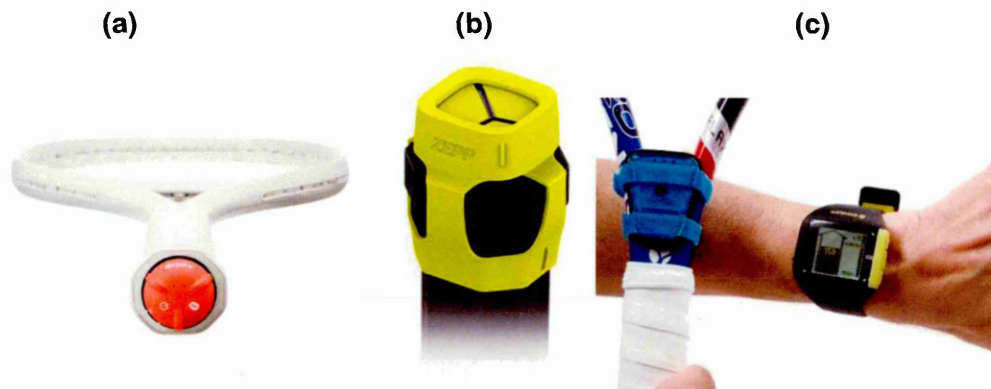


Figure 2.7 Approved PAT products: (a) SONY Smart Tennis Sensor 'pod' (adapted from SONY 2015), (b) ZEPP Tennis 'pod' (adapted from ZEPP 2015) and (c) Artengo Personal coach 'pod' (adapted from ARTENGO 2015).

2.2.3 Passive markers

Passive markers are tracked by picking up reflections from their surface and have been used in videogrammetric techniques to define specific points on a participant or striking implement to measure movements. Elliott, Marsh and Blanksby (1986) tracked movement of tennis players joint centres during a serve. Players' upper and lower limb joints were landmarked with tape and tracked using two high speed cameras which were calibrated using the 3D Direct Linear Transform (DLT) method (Abdel-Aziz and Karara 1971). Calibration involved filming an object consisting of a physical mesh of control points in the capture volume in which the service action was taking place. Elliott, Marsh and Blanksby (1986), Elliott, Marsh and Overheu (1989a) and Elliott, Marsh and Overheu (1989b) used two cameras running at 200 and 300 fps to analyse the serve, forehand and backhand respectively. The slower camera was used to record the player and racket, whilst the faster camera focused on the ball.

Bahamonde and Knudson (2003) performed a similar on-court 3D analysis but concentrated on biomechanical kinematics of groundstroke shots and the effect of player stance. Reflective bands were placed around the wrist and elbow of the arm with which the player held the racket to aid tracking of joint centres. Knudson and Blackwell (2005) measured ball and racket velocities in non-competitive play by tracking white markers attached to the racket frame. Footage of forehand shots was collected using a high speed camera which limited this study to a 2D analysis.

Choppin et al. (2007a) developed a methodology to measure 3D ball and racket movements using two synchronized high speed cameras operating at 1000 fps. The camera calibration toolbox for MATLAB was used to calibrate a control volume using a checkerboard (Strobl et al. 2007). The calibration allowed for reconstruction of 3D coordinates from the 2D pixel coordinates from the image planes of both cameras. Image-based 3D measurement systems have also been used to measure shoe-surface interactions in football (Driscoll et al. 2014). Choppin et al. (2007b) used the methodology to analyse ball and racket movements from footage recorded at the 2006 Wimbledon qualifying tournament. The high speed cameras were positioned at each side of the net facing the baseline, where the 2 x 2 x 2 m calibrated control volume was located (Figure 2.8).

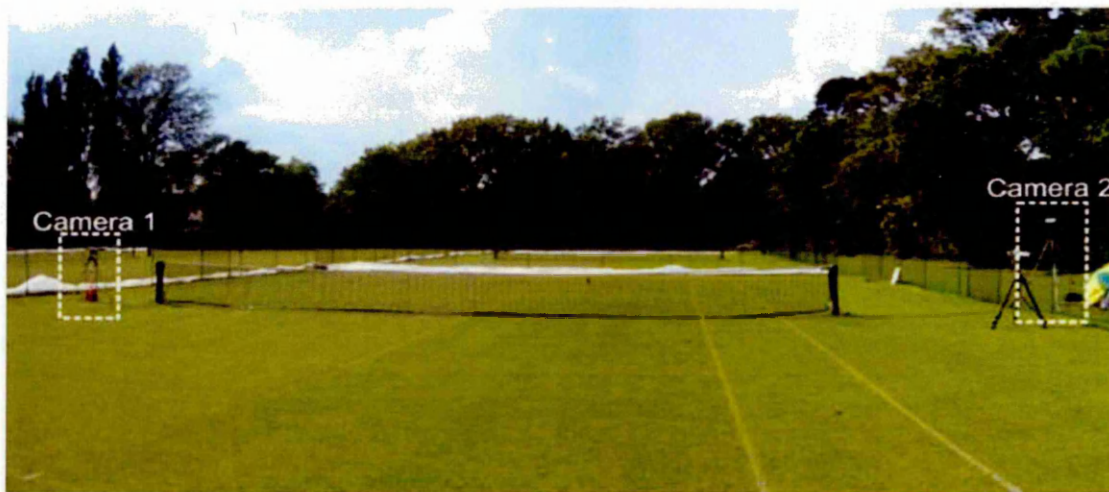


Figure 2.8 Stereo camera setup used at the 2006 Wimbledon qualifying tournament (taken from Choppin et al. 2007a).

Players were asked to carry out a typical practice session consisting of structured rallies with five reflective tape markers attached to their rackets at specific points that could be tracked. The tape markers posed minimal intrusion to players because they were small in size and mass. An example of how the markers were positioned on a racket is shown in Figure 2.9.

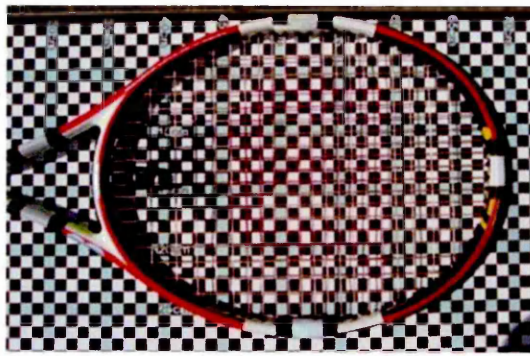


Figure 2.9 Photo showing the location of the 5 reflective tape markers on the racket frame (taken from Choppin et al. 2007a).

Images were recorded at 1000 frames per second which allowed racket and ball velocities, impact angle and position and ball velocity and spin rate to be measured. Manual digitisation was used to identify the tape markers in order to track racket and ball movements which can be time consuming and inaccurate. This method of analysis is subject to varying amounts of error, depending on the skill of the user and the quality and resolution of the images.

To date, the least intrusive marker-based technique is that developed by Choppin et al. (2007a). Choppin et al. (2011) reported mean racket velocity and angular errors of $\pm 0.5 \text{ m s}^{-1}$ and $\pm 1^\circ$ respectively. However, even with markers that inflict a minimal level of intrusion, elite tennis players are often sceptical about making racket modifications. Factors as small as cosmetic alterations to the racket frame maybe enough to distract them. Therefore, data from intrusive motion analysis will never truly represent match-play tennis, since it is confined to practice conditions. Notably, Goodwill et al. (2007) measured considerably lower spin rates in practice conditions compared with those in competitive match-play.

Choppin et al. (2007a) used markers that reflect natural light from the sun. However, some motion analysis techniques require additional light sources that the makers can reflect. The twelve camera Eagle motion capture system by the Motion Analysis Corporation (MAC 2012) uses red light, spherical retroflective markers and sensors with a red light filter. The spherical markers reflect red light towards the sensors in all directions. Thus, several combined sensors and external red light sources are needed. Figure 2.10 shows one of the cameras with the red light sensor attached. The Raptor series (2010) of motion capture systems can be used outdoors as well as indoors which was not possible with previous generations of camera systems from the Motion Analysis Corporation (Figure 2.10).



Figure 2.10 Example of a camera from the Motion Analysis Corporation with red light sensor and the Raptor system in action outdoors being used to track a golf shot.

Gillet, Leroy and Germaine (2003) used this type of passive marker system to study the balance of tennis players. However, the cameras require specific calibration and the volume over which they are calibrated must not include any other object that reflects red light. A minimum of two cameras must be used which are linked to a computer at all times. The spherical markers are lightweight but protrude from the surface of the object being tracked and could potentially be knocked off. Furthermore, marker identification often requires user input which is prone to mistakes for markers positioned close together.

2.2.4 Non-intrusive techniques

In 1980, Cyclops became the first non-intrusive line calling system to be used in competitive tennis (Pallis 2004). For serves that landed close to the service line, Cyclops was validated to determine if the ball fell in or out of the service lines. Cyclops achieved this by constantly monitoring five infrared beams aligned with the service lines. If one of the beams on the outside of the lines was broken then the serve would be automatically called out.

In 2006, Cyclops was replaced by the camera-based line-calling system, Hawk-Eye. Hawk-Eye, was patented in 2001 (Sherry and Hawkins 2001) and improved upon Cyclops by accurately identifying the location of the ball with respect to the court lines. In 2004, the ITF reported a maximum uncertainty value of ± 0.9 mm following an initial validation of the system. The ITF used high-speed videogrammetry to render Hawk-Eye suitable to officiate match play tennis (Capel-Davis and Miller 2007). Hawk-Eye is now used at almost every elite tennis tournament (ITF, 2014). Hawk-Eye uses at least four and often up to ten cameras for tennis. Two-dimensional ball position in the image sequences of each camera is determined using image processing techniques. Hawk-Eye uses the lines and physical dimensions of the court to calibrate the cameras so that triangulation can be used to obtain the 3D position of the ball relative to the court

lines (Hawkeye 2015). Ball trajectory analysis can also account for ball deformation, rolling and slip during impact (Cross 2002). In tennis, Hawk-Eye is used extensively by the media, providing match statistics and ball trajectory visualisations during line calling decisions (Figure 2.11).



Figure 2.11 An example of a line call review from Hawk-Eye (2015)

Hawk-Eye has also been applied to other sports including snooker, cricket and most recently football (Hawk-Eye, 2015). In 2012, Hawk-Eye's goal line technology gained authorisation from FIFA to be installed in football stadiums around the world. However, Hawk-Eye cannot measure ball spin and the cost of multiple high-speed cameras restricts its use. Kelley et al. (2010) validated a system that allows tennis ball spin rates and velocity to be measured using footage from a high-speed video camera. Calibration was performed using an image of the ball captured at a known distance from the camera. Spin rates were calculated based on the change in position of the logo on the surface of the ball. Kelley (2011) used the analysis software to semi-automatically measure spin rates at the 2008 and 2009 Wimbledon qualifying tournament. There was little difference in spin rates between years and spin rates for second serves were higher than spin rates for first serves.

Two-dimensional non-invasive motion tracking systems have predominantly been developed to track ball and player movement. Pingali, Jean and Carlborn (1998) and Yan, Christmas and Kittler (2005) describe methods for tracking ball and tennis player position from broadcast television footage. Both methods use temporal differencing techniques in order to track player and ball position. Pingali, Jean and Carlborn (1998) used positional data to provide a statistical analysis of matches and Yan, Christmas and Kittler (2005) developed a system designed for robustness rather than accuracy.

Prozone is a multi-camera player tracking system used in football (Di Salvo et al. 2006). It uses an eight-camera system to cover the entire football pitch which is initially calibrated using a linear 4-point transformation which is later refined using a 50 point algorithm to reduced errors induced by lens distortion. Prozone identifies and tracks

players based on shirt colour and movement thresholds. Di Slavo et al. (2006) validated Prozone analysing a series of shuttle runs performed by players at different speeds. Di Slavo et al. (2006) reported high player tracking accuracy; mean velocity error was 0.127 m s^{-1} , concluding that Prozone allowed real-time tracking of multiple players without the need for specialised equipment such as transmitters or colour coded shirts. Today Prozone is used by over 300 clubs and organisations across the world to inform and enhance coaching (Prozone 2015).

In tennis, the SAGIT computer vision system (Perš et al. 2002) has been used to automatically track the movements of both players, albeit with operator supervision (Martínez-Gallego et al. 2013). The SAGIT system uses footage from two cameras (384 x 288 pixels) mounted on the ceiling above an indoor court with each covering one half of the court (Figure 2.12). Following background subtraction, the SAGIT system tracks players based on colour and shape matching. Player location was converted from pixel to real world coordinates by first accounting for lens distortion followed by scaling and translating coordinates between the court plane and camera sensor (Perš et al. 2002).

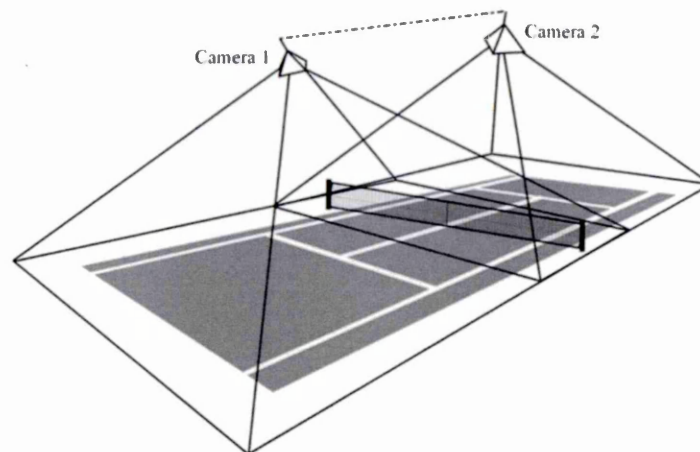


Figure 2.12 Camera locations used for analysis by SAGIT (taken from Martínez-Gallego et al. 2013).

TennisSense is another example of a multi-camera system for player and ball tracking in tennis (Conaire et al. 2009). It utilises nine synchronised cameras with pan, tilt and zoom capabilities arranged around the court. Four cameras located at each end of the court (8 in total) were used for coaching purposes and a single overhead camera performed all the analyses.

Dunn (2014) used a single camera in a stadium setting to track player position and identify foot-surface contacts during match play tennis. Dunn et al. (2014) developed a

semi-automatic method for identifying foot-surface contacts using a range of image processing techniques. Player location relative to the court lines was converted from pixel to real world coordinates using an elevated camera-plane model. Dunn et al. (2014) reported a 91.3% success rate of images being analysed by the foot surface contact algorithm.

Notably, non-intrusive motion analysis methods for tracking tennis racket movements have received little attention. This is likely due to the complex nature in which the racket is wielded by a player during a shot. However, an area of research in markerless motion capture (MMC) has demonstrated potential for measuring movements of sports implements.

2.3 Markerless motion capture

Markerless motion capture (MMC) can be considered as an area of non-invasive motion capture. Predominantly, MMC methods have used visual hulls to perform biomechanical analyses of human movement. A detailed account of the evolution of MMC methods can be found in Mündermann, Corazza and Andriacchi (2006).

2.3.1 Visual hulls

The term visual hull was coined by Laurentini (1994) and refers to the largest object that is consistent with a finite set of available silhouette views (Lazebnik et al. 2001 and Matusik et al. 2001). The visual hull concept is illustrated in Figure 2.13. Figure 2.13 (a) shows two cameras used to generate silhouette views of a toy duck. In Figure 2.13 (b) the camera centres are represented by small spheres, the image planes are positioned in front of the camera centres and the silhouettes are shown non-inverted.

The visual cones corresponding to each silhouette are shown in Figure 2.13 (c). The visual cone is the volume of space that the actual object cannot lie outside, given the observed silhouettes (Forbes 2007). Intersection of the visual cones is the object's visual hull (Figure 2.13(d)). With only two silhouette views, the visual hull is a poor approximation of the object's shape. However, if more silhouette views are added, more information about the volume of space that the object doesn't occupy is added and the visual hull becomes a better approximation of the actual object (Figure 2.13 (e)). In order to determine the visual hull corresponding to a set of silhouettes, the cameras that captured the images must be fully calibrated (Forbes, Voigt and Bodika 2003, Forbes 2007). A camera is fully calibrated when its internal parameters i.e. focal length and principal point and external parameters i.e. rotation and translation relative to a reference frame (pose) are known. Forbes (2007) uses the term *fully calibrated*

silhouette set to describe a collection of camera poses in a common reference frame, with the internal parameters known.

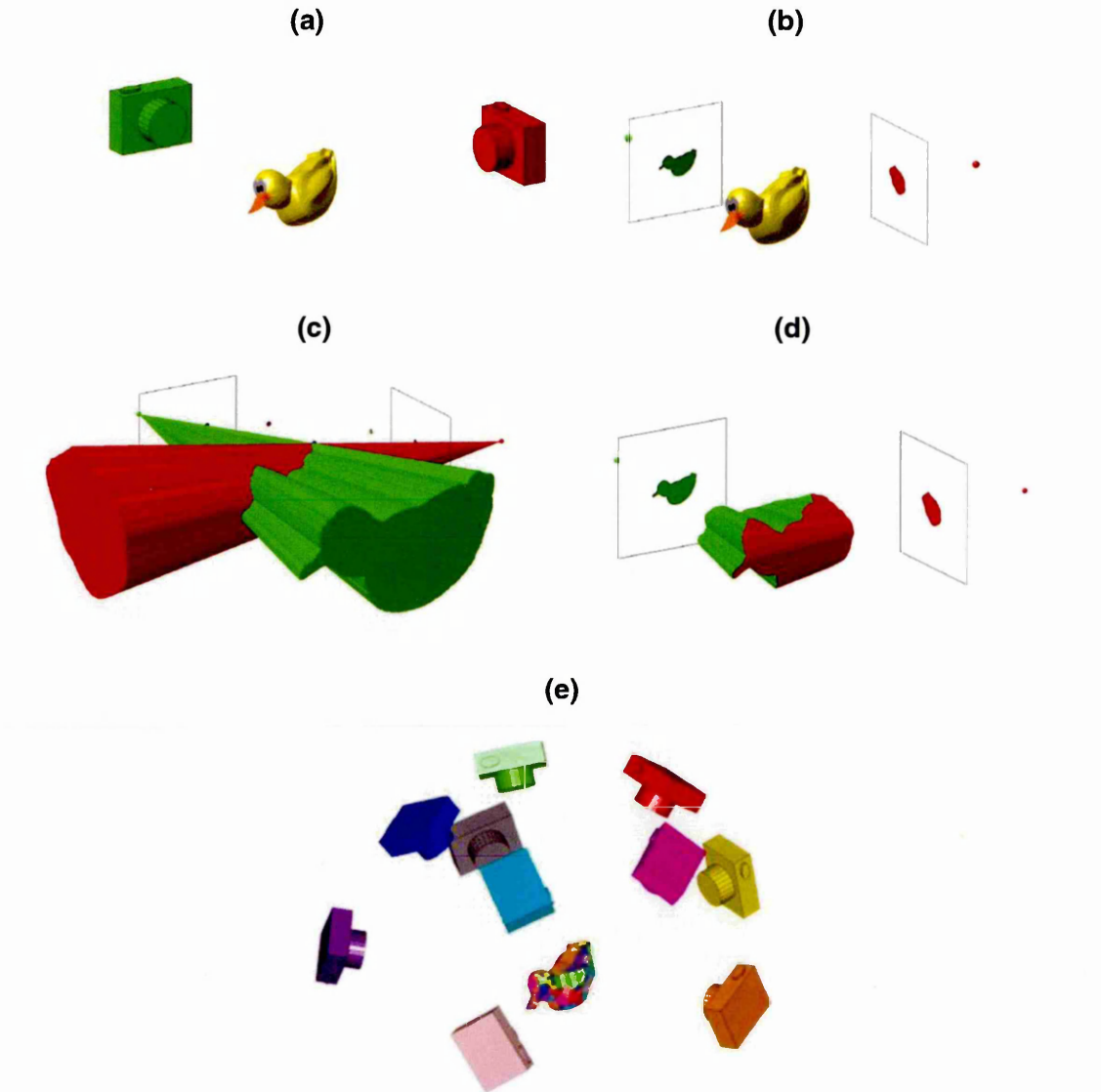


Figure 2.13 The visual hull concept: (a) a duck viewed by two cameras, (b) two silhouette views of a duck, (c) visual cones corresponding to the two silhouettes, (d) the visual hull corresponding to the two silhouettes and (e) the visual hull corresponding to ten silhouettes (taken from Forbes 2007).

2.3.2 Intrinsic and extrinsic camera parameters

The process of an intrinsic camera calibration provides a model of the camera's geometry. The most basic camera model is that of a pinhole camera (Hartley and Zisserman 2003). For a pinhole camera, light is assumed to enter from the scene as a single ray and no lenses are required to focus rays onto the image sensor. The distance from the pinhole aperture to the image plane is a camera's focal length (Figure 2.14). In Figure 2.14 the pinhole aperture is reinterpreted as the centre of projection

because every incoming ray heads towards this point. The point at which the optical axis intersects the image plane is referred to as the principal point (Figure 2.14). Thus for a perpendicular pinhole model, only a scale factor is needed to transform real world to image coordinates:

$$s \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} \quad (2.1)$$

where uv and XYZ are image plane and real word coordinates respectively, s is a scale factor and 1 represents the projection of the image plane coordinate to infinity.

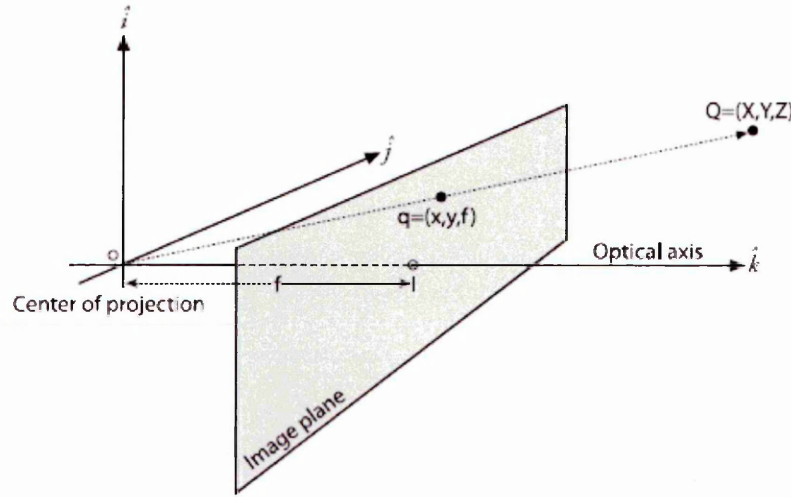


Figure 2.14 Pinhole camera geometry illustrating focal length and principal point which govern the projection of real world point Q in the image plane as q (taken from Bradski and Kaehler 2008).

Most cameras require a lens to allow for a range of camera viewing angles and focal lengths for when the distance at which the image of an object is changed (Bradski and Kaehler 2008). Projection of a 3D point in the physical world onto the image plane is given by:

$$s \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} f_x & \varsigma & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} \quad (2.2)$$

where f_x and f_y are focal lengths in the u and v image axes respectively (expressed in pixel units), ς represents pixel skew (assumed to be zero since the angle between the x and y pixel axes is often very close to 90° , justifying the assumption of rectangular pixels (Bouquet 2010)), c_x and c_y model horizontal and vertical displacement from the centre of the image plane respectively.

Most commercially produced lenses are spherically shaped and introduce radial and tangential distortions to an image (Bradski and Kaehler 2008). Radial distortion is zero at the optical centre of the lens and increases toward the periphery due to its spherical shape. Tangential lens distortion is due to manufacturing defects i.e. non-parallel alignment of the lens and the sensor. The effect of radial and tangential distortion on an image can be calculated. Bouguet (2010) uses Heikkila and Silven's (1997) intrinsic camera model which includes both radial and tangential lens distortion. Collectively, the camera model parameters f_x , f_y , ς , c_x , c_y and lens distortion term kc , a 5 element vector, are the intrinsic camera parameters. Thus given the calculation of the intrinsic parameters, any uv pixel coordinate in the image plane can be normalised to the camera coordinate system via the following steps.

Subtract principal point and divide by focal length:

$$[x_n, y_n]^T = \left[\frac{u-u_0}{f_x}, \frac{v-v_0}{f_y} \right]^T \quad (2.3)$$

Remove pixel skew:

$$[x_n, y_n]^T = [x_n - \varsigma * y_n]^T \quad (2.4)$$

Compensate for lens distortion:

$$r = x_n^2 + y_n^2 \quad (2.5)$$

$$K_{radial} = 1 + (kc_1 * r) + (kc_2 * r^2) + (kc_5 * r^3) \quad (2.6)$$

$$\Delta x = (2kc_3 * x_n * y_n) + kc_4(r + 2x_n^2) \quad (2.7)$$

$$\Delta y = kc_3(r + 2y_n^2) + 2kc_4 * x_n * y_n \quad (2.8)$$

$$[x_n, y_n]^T = \left[\frac{x_n - \Delta x}{K_{radial}}, \frac{y_n - \Delta y}{K_{radial}} \right]^T \quad (2.9)$$

where x_n and y_n are normalised horizontal and vertical image coordinates respectively, $kc_{1,2,5}$ are radial distortion coefficients and $kc_{3,4}$ are tangential distortion coefficients. Bouguet (2010) uses a loop of 20 iterations to converge to the normalised image coordinate $[x_n, y_n]^T$. Intrinsic model optimisation is often run over a subset of the

distortion coefficients. Once the intrinsic camera parameters are known, any 2D coordinates in the image plane can be re-projected into object space by applying the rotation matrix R and translation vector t (collectively referred to as the extrinsic parameters) (Figure 2.15), subject to an error with respect to a ground truth value. The error depends on the quality of the calibration and the method used (Zhang 1999, Forbes 2007, Choppin 2008).

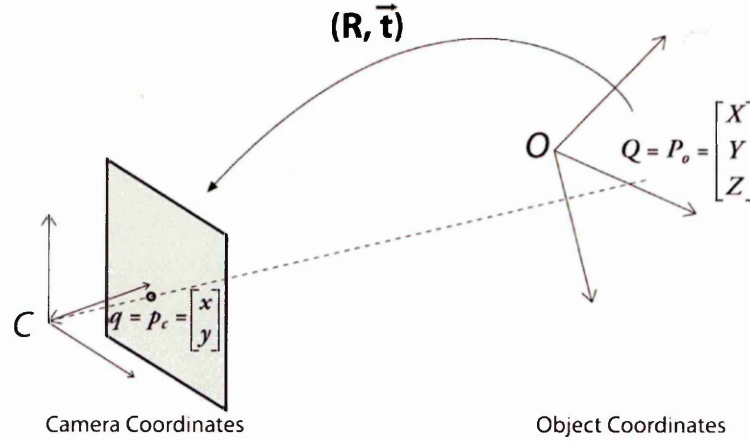


Figure 2.15 The normalised 2D image coordinate P_c (camera coordinate) is related to the 3D (object) coordinate P_o by applying the rotation matrix R and translation vector t (taken from Bradski and Kaehler 2008).

The calibration procedure generally involves recording a series of control points which are used to calculate the required parameters. The control points can be provided by an accurately manufactured 3D calibration object (Abel-Aziz and Karara 1971), a 2D planar grid (Zhang 1999), or 1D laser pointer (Svoboda, Martinec and Pajdla 2005). The calibration parameters are often calculated using the linear least squares approach outlined in Abdi (2003), or for more complex non-linear cases using the Levenburg-Marquardt algorithm (Moré 1977). Both minimise an error function in order to optimise the position of P_o in Figure 2.15. Details of the latest calibration procedures for MMC applications are discussed in sections 2.3.7 and 2.3.8.

2.3.3 Silhouette extraction

Visual hull construction also depends on accurate silhouette extraction (Forbes, Voigt and Bodika 2003). However, there are different methods for extracting silhouette boundaries (which are also referred to as contours). A cubic B-spline snake is one such method and has been used in a number of studies (Cipolla and Blake 1990, Cipolla and Blake 1992, and Wong and Cipolla (2004). Cubic B-spline snakes provide a compact representation for silhouettes of various complexity and can achieve a higher resolution than the original image (Wong and Cipolla 2004). The process of extracting a silhouette using a B-spline snake is illustrated in Figure 2.16. Firstly, a B-spline snake

is manually initialised close to the target silhouette (Figure 2.16 (a)). Points are then sampled along each spline segment and a search for the silhouette edge along the direction normal to the local tangent at each sample point is performed (Figure 2.16 (b)). The sample points on the B-spline snake are then updated using linear least-squares methods, so that they attach to the silhouette edge that was found (Figure 2.16 (c)).

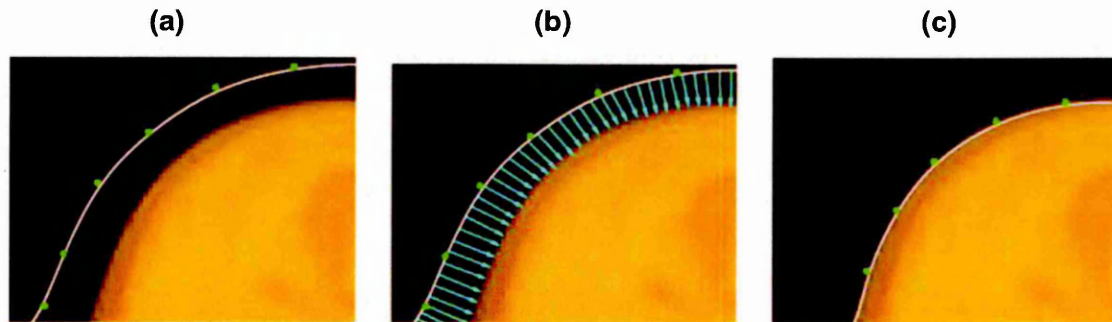


Figure 2.16 (a) A B-spline snake is initialised close to the target silhouette, (b) points are sampled from each spline segment and a search for the silhouette edge along the direction normal to the local tangent at each sample point is performed. (c) The points are updated to attach to the silhouette edge that was found (adapted from Wong 2001).

The snake was originally introduced by Kass, Witkin and Terzopoulos (1988) and is described as an energy minimising spline guided by internal and external forces. The external force requires a user-supplied input function to provide some higher level information from the image e.g. the sample points (Figure 2.16 (a)). The internal forces are determined by the shape (elasticity) of the contour and are defined in terms of first and second order derivatives. However, B-Spline snakes have a slow convergence speed because of large numbers of sample points to optimise and difficulty in setting the weighting factors that comprise the internal forces (Brigger, Hoeg and Unser 2000). Marsousi et al. (2010) used B-Spline snakes to detect the boundaries in left ventricular echo-cardiographic images with a mean computational time of 1.24 ± 0.28 seconds.

Furthermore, B-spline snakes are required to identify boundaries in noisy images, which cannot be detected by simple low level features such as grey thresholding or edges (Stammberger et al. 1999). However, silhouette images for visual hull applications are often captured in controlled environments by using backlights to ensure background pixels are substantially brighter than foreground pixels (Mündermann, Corazza and Andriacchi 2006, Forbes 2007). This simplifies the boundary extraction problem by allowing methods to be used which rely on these low level image processing functions alone.

Polygonal silhouette boundaries allow simpler and more efficient methods to be developed (Matusik, Buehler, and McMillan 2001, Lazebnik 2002). These silhouettes are extracted using threshold-based segmentation in order to separate foreground from background pixels. Forbes (2007) employed Otsu's (1979) method to determine appropriate threshold values. Each pixel in the original image is compared to the chosen threshold value, if the pixel is above the threshold value the corresponding pixel value in the binary image is 1, otherwise it is 0. Forbes (2007) provides an algorithm for sub-pixel resolution polygonal silhouette boundary extraction. Broadly, it proceeds by finding a pixel inside the silhouette, travelling down the vertical image axis to locate the silhouette boundary, traversing the silhouette to determine its pixel resolution boundary and finally traversing the silhouette once more to determine its sub-pixel resolution boundary. Mean computational time of Forbes' (2007) algorithm to detect sub-pixel resolution boundaries for 72 silhouette images was 0.01 ± 0.01 seconds.

Figure 2.17 shows an example of sub-pixel resolution polygonal silhouette boundary extraction. The pixel resolution boundary is shown in green and runs across pixel edges. Sub-pixel silhouette boundary resolution is achieved using linear interpolation based on pixel intensity value. Each vertex of the sub-pixel boundary lies on a pixel-length line segment associated with each edge of the original (green) boundary.

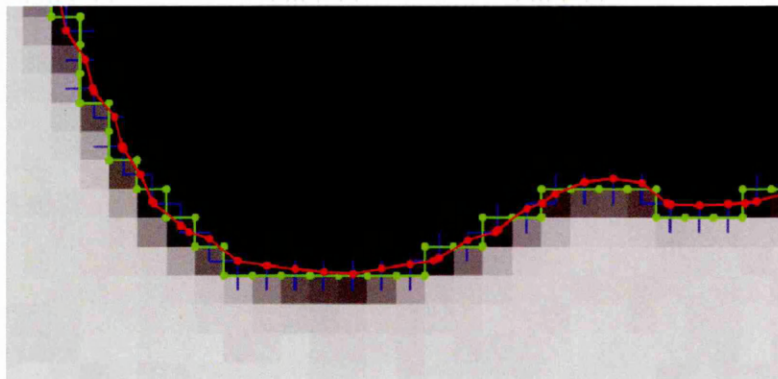


Figure 2.17 An example of sub-pixel resolution silhouette boundary extraction. The pixel and sub-pixel resolution boundaries are shown in green and red respectively. Each vertex of the sub-pixel boundary lies on a pixel-length segment (shown in blue) associated with each edge of the pixel resolution boundary (taken from Forbes 2007).

The resultant sub-pixel boundary is the same as the marching squares boundary which is the 2D equivalent of the marching cubes algorithm (Lorensen and Cline 1987). The marching square algorithm considers pixel-sized squares centred at each pixel corner. The corners of the squares which lie on pixel centres are classified into foreground and background using a threshold. Thereafter a look-up table of square classifications

determine how the boundary enters and exits each square. The sides of the squares are equivalent to the pixel-length line segments used by Forbes (2007).

2.3.4 Quantifying silhouette quality

Silhouette quality quantification is often omitted or considered by subjective assessment (Liu, Zheng and Xiong 2009). However, some modern works have focused on silhouette quality with respect to the impact upon the performance of the intended application. This is especially important for applications in outdoor environments, where complex backgrounds and varying illumination make the silhouette extraction task more difficult. Liu, Zheng and Xiong (2009) present a method based on 1D foreground-sum signal processing to assess the quality of a silhouette sequence in order to improve gait recognition algorithms. The silhouette quality quantification (SQQ) tool is applied to a silhouette quality weighting (SQW) framework which emphasises good silhouettes more than bad ones via a weight function. Liu, Zheng and Xiong (2009) showed that recognition performance in outdoor environments can be improved by using the proposed SQQ and SQW methods.

Marsousi et al. (2010) quantified the accuracy of silhouette boundary detection from left ventricular echo-cardiographic images using Dice's coefficient equation. The Dice's coefficient equation measures the overlap between automatically and manually obtained boundaries. The manually obtained boundaries were deemed as gold standard and extracted by experts. However, despite the practical use of Dice's coefficient, specialists were more concerned about the area and volume of the left ventricle of the obtained boundary.

Ng et al. (2015) developed an automated wrinkle detection method that uses Hessian Line Tracking (HLT). As a final step of the HLT method, post processing is required to reduce noisy pixels from the binary wrinkle map. This involves performing a number of morphological operations including median, directional and area filtering. Wrinkles could then be detected from the binary image and the centre line of each wrinkle was used for validation. Ng et al. (2015) validated the reliability of the proposed automatic wrinkle detection method against manual annotation results using the Jaccard Similarity Index (JSI). The JSI was used to quantify the accuracy of inter-reliability between the manual coders, which on average was 94%. Thereafter, the inter-reliability results were deemed as ground truth and any overlap between the estimated and ground truth line larger or equal to 40% was considered as a correct detection.

However, the aforementioned silhouette quality quantification examples are designed to be applied to silhouettes extracted from noisy backgrounds i.e. when there is no clear difference between background and foreground pixels. As previously mentioned, most visual hull applications take place in controlled environments using fully calibrated cameras and backgrounds that facilitate silhouette extraction. Therefore, the subjective assessment of silhouette quality is adequate (Forbes 2007). This could involve superimposing an extracted silhouette boundary on top of the original image to visually observe whether the boundaries matchup (Forbes 2007). However, of more importance is the measure of consistency across a set of silhouettes associated with a camera's pose. Such a measure can also be used to quantify silhouette quality since consistent silhouette sets require high quality silhouettes (Wong 2001).

2.3.5 Measuring silhouette consistency

A consistent silhouette set is one that has been produced using exact calibration parameters and silhouettes (Forbes, Voigt and Bodika (2003) , Forbes 2007). In reality however, there will always be error associated with calibration parameters and segmented silhouette pixel boundaries. Therefore, it is useful to formulate a measure of the degree of inconsistency of a silhouette set (Forbes 2007). The epipolar tangency constraint provides an efficient method for measuring silhouette set consistency (Wong 2001). Other measures of silhouette consistency also exist; *silhouette coherence* (Hernández 2004, Hernández, Schmitt and Cipolla 2007) and the *silhouette calibration ratio* (Boyer 2006) that use more information contained in the silhouettes therefore are more computationally expensive (Forbes 2007).

2.3.6 The epipolar tangency constraint

The epipolar tangency constraint is a geometrical constraint that applies to a pair of calibrated cameras and associated silhouettes. Figure 2.18 illustrates the epipolar tangency constraint for a pair of fully calibrated silhouettes of a toy duck as used by Forbes, Voigt and Bodika (2003) and Forbes (2007).

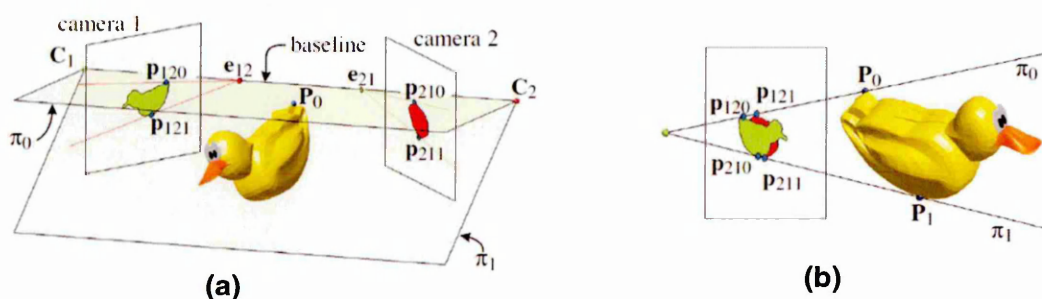


Figure 2.18 The epipolar tangency constraint for a pair of fully calibrated silhouettes: (a) shows a frontal view and (b) a side view (taken from Forbes 2007).

The line that joins the camera centres C_1 and C_2 is called the baseline. Projection of each camera centre into the opposite camera's image plane is called an epipole. In Figure 2.18, epipole e_{12} is the projection of C_2 in the image plane of Camera 1, and epipole e_{21} is the projection of C_1 in the image plane of Camera 2. Note that the epipoles do not necessarily lie in the visible region of the image plane; theoretically the image plane extends to infinity in both directions.

The planes π_0 and π_1 converge at the baseline and are tangent to the 3D duck. As long as the baseline does not pass through the object then there will be two such outer tangent planes for any 3D object. The 3D points P_0 and P_1 indicate where the planes touch the surface of the duck, and are referred to as frontier points. Since the planes pass through both the camera centres and touch the surface of the duck, the frontier points must project onto the boundaries of the silhouettes in both image planes.

A projection of a frontier point is called an epipolar tangency point. In Figure 2.18, the tangency points p_{120} and p_{210} , are projections of the frontier point P_0 , and the tangency points p_{121} and p_{211} , are projections of the frontier point P_1 , in the image planes of camera 1 and camera 2 respectively. The notation method p_{ijk} was used by Forbes, Voigt and Bodika (2003) so that i indicates the image plane number on which the point is projected, j is the number of the opposite image plane and k indicates to which frontier point i and j corresponds.

The relative rotation R and translation t between two camera views i and j (i.e. camera 1 and camera 2 in Figure 2.18) can be used to derive a relation between a 3D coordinate and its observed pixel points p_i and p_j . The corresponding location p_i of the normalised homogeneous image coordinate p_i is given by

$$p_j = R * (p_i - t). \quad (2.10)$$

The pose parameters R and t can be used to compute a 3x3 essential matrix

$$E_{ji} = [t]_x * R. \quad (2.11)$$

The essential matrix (Longuet-Higgins 1981) encapsulates the extrinsic geometry of views i and j , relating normalised image coordinates. The skew symmetric matrix is computed from the X, Y and Z components of the translation vector t using

$$[t]_x = \begin{bmatrix} 0 & -t_z & t_y \\ t_z & 0 & -t_x \\ -t_y & t_x & 0 \end{bmatrix}. \quad (2.12)$$

If the intrinsic camera parameters are known, the fundamental matrix can be computed from the essential matrix

$$F_{ji} = K_i^{-T} * E_{ji} * K_j^{-1}. \quad (2.13)$$

The 3x3 K matrices stores a cameras intrinsic calibration parameters. Figure 2.19 shows outer epipolar tangent lines for the duck example from Figure 2.18. If there is no error in the camera parameters or segmentation of silhouette boundaries then the viewing ray of a tangency point must project exactly onto the corresponding ray in the opposite image. Thus a line between an epipole e_{ij} and tangency point p_{ijk} is the same as the line $F_{ji} * p_{ijk}$, yielding the epipolar constraint such that

$$p_i^T * F_{ji} * p_j = 0. \quad (2.14)$$

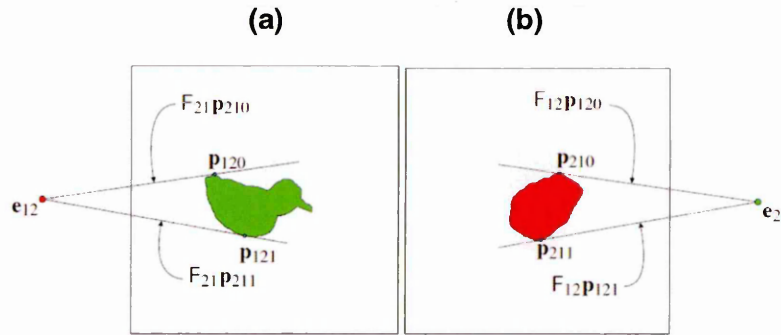


Figure 2.19 The epipolar tangency constraint for a pair of duck silhouettes. Projection of a tangency point in (a) the left image is constrained to lie on the epipolar tangent line in (b) the right image, vice versa (taken from Forbes 2007).

2.3.7 A measure of inconsistency based on epipolar tangents

Real camera calibration parameters and segmented silhouette pixel boundaries contain error. Therefore, the epipolar tangent lines don't project exactly through the tangency point in the opposite image resulting in inconsistent silhouette pairs (Figure 2.20).

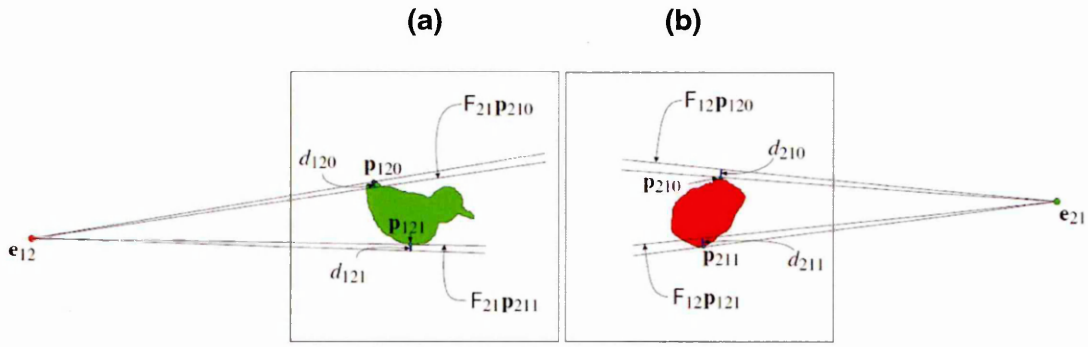


Figure 2.20 Error associated with real camera parameters and segmented silhouette boundaries resulting in the epipolar tangency lines not projecting exactly through the tangency points in (a) the left and (b) the right images. The reprojection error is used to measure the degree of inconsistency for a silhouette pair (taken from Forbes 2007).

Reprojection errors can be computed as a measure of the degree of inconsistency between a pair of silhouettes with an associated pose. These reprojection errors are the perpendicular distance d between a tangency point and a projected epipolar line, and can be computed using the fundamental matrix

$$d_{ijk} = \frac{p_{jik}^T * F_{ij} * p_{ijk}}{\sqrt{(F_{ij} * p_{ijk})_1^2 + (F_{ij} * p_{ijk})_2^2}} \quad (2.15)$$

where $(F_{ij} * p_{ijk})_1^2$ and $(F_{ij} * p_{ijk})_2^2$ are first and second element of the vector $(F_{ij} * p_{ijk})^2$. The perpendicular distance d in equation 2.15 is related to the Sampson approximation (Hartley and Zisserman 2003). Hartley and Zisserman (2003) recommend using the Sampson approximation to give a good first-order approximation to the perpendicular error distance.

Except when the baseline passes through the 3D object, the two outer epipolar tangency points are always available in any pair of views resulting in four reprojection errors. For a fully calibrated set of silhouettes, the measure of inconsistency is based on the root mean square (RMS) value of all the reprojection errors and is referred to as the epipolar tangency error (ETE) across a set (Wong 2001, Forbes 2007).

The ETE can be formulated as an objective function that can be minimised (Price and Morrison 2007). The cost of the objective function can be minimised using iterative optimisation methods: the conjugate gradient method (Wong and Cipolla 2004) and the Levenberg-Marquardt method (Forbes, Voigt and Bodika 2003, Price and Morrison 2007) have previously been used. The Levenberg-Marquardt method implicitly computes the sum of squares of the components of a function $F(x)$ (MATHEMATICS,

2015). Since the sum of squared reprojection error distances (ETE) are minimised, the Levenberg-Marquardt method is often the preferred choice algorithm (Forbes, Voigt and Bodika 2003, Price and Morrison 2007, Sheets et al. 2011).

The following sections will review previous uses of visual hulls, include details of calibration procedures and provide examples of how ETE distance minimisation is implemented for a variety of applications in biomechanics and sport.

2.3.8 Visual hulls for biomechanical applications

In the last ten years, noted work using visual hulls for biomechanical analysis of human movement has been carried out in laboratory environments (Mündermann, Corazza and Andriacchi 2006, Corazza et al. 2006, Corazza et al. 2010). Mündermann, Corazza and Andriacchi (2006) discuss both the need and advantages of markerless motion capture (MMC) for 3D human movement biomechanical applications. Skin movement artefact is identified as the main limiting factor of marker-based approaches provoking the development of MMC methods (Mündermann, Corazza and Andriacchi 2006). However, it is the validation of markerless methods against marker-based data that has eluded most studies (Mündermann, Corazza and Andriacchi 2006). Mündermann et al. (2005) demonstrated that a MMC system consisting of 8 cameras provided less than 1° difference for sagittal and frontal plane knee joint angles during walking trials compared with a marker-based method.

Corazza et al. (2010) tracked walking trials and gymnastic flips using variable numbers of cameras. Calibration was performed offline to calculate the intrinsic and extrinsic parameters and a common global reference frame. Offline calibration methods usually make use of specific patterns or objects to determine parameters (i.e. someone must be physically present in the capture volume with the specialised pattern or object so that calibration data can be collected (Sinha and Pollefeys 2010)). The tracking approach employed a Levenberg-Marquardt optimisation scheme to minimise the sum of squared error distances between corresponding points on laser scanned models (markers were also scanned to identify joint centre locations) and participants' visual hulls. For the walking trials, average deviations between joint centres calculated with marker-based and markerless systems was 15 ± 10 mm. While a qualitative comparison through visual inspection showed good agreement between the methods for the gymnastic flip sequence. Furthermore, Corazza et al. (2010) concluded that visual hull based approaches for tracking human movements are not suitable unless configurations with no less than eight cameras are used.

Sheets et al. (2011) used eight synchronized and calibrated cameras (640 x 480 pixels) (Figure 2.21) operating at 200 Hz to create visual hulls of a tennis player in order to track upper limb movements during a serve. The high-speed cameras were calibrated offline and placed so that reconstruction of the arm surface when raised above the player's head prior to ball contact was of the highest possible quality.

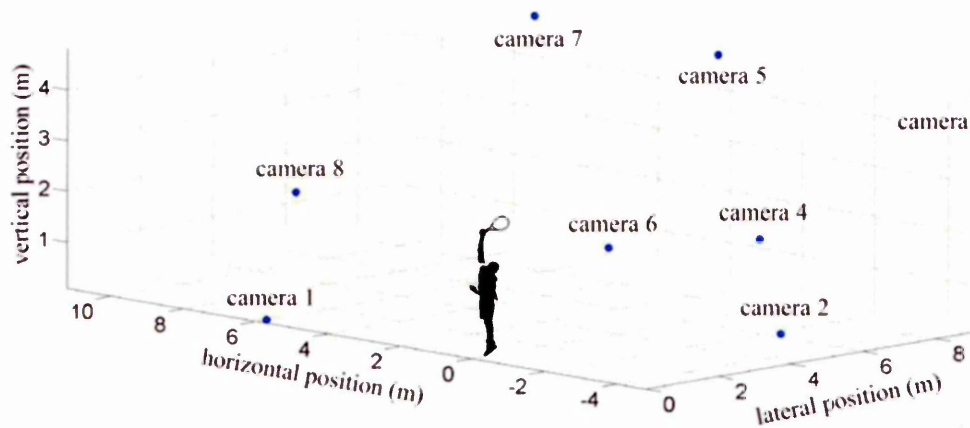


Figure 2.21 Positions of the eight synchronised cameras surrounding the player (taken from Sheets et al. 2011).

Figure 2.22 (a) shows frames captured from camera number four during a service trial. Image segmentation formed eight silhouette views of the player which were projected back into the capture volume using the intrinsic and extrinsic parameters of each camera determined during calibration. Volume intersection techniques were then used to reconstruct a visual hull of the player and racket (Figure 2.22 (b)). Subject-specific segment models made using a 3D laser scanner were used to identify fifteen joint centre positions (Figure 2.22 (c)). In order to measure a player's joint centre position during a serve, the subject-specific segment models were optimally fitted into the visual hull for each frame using exponential mapping (Figure 2.22 (d)).

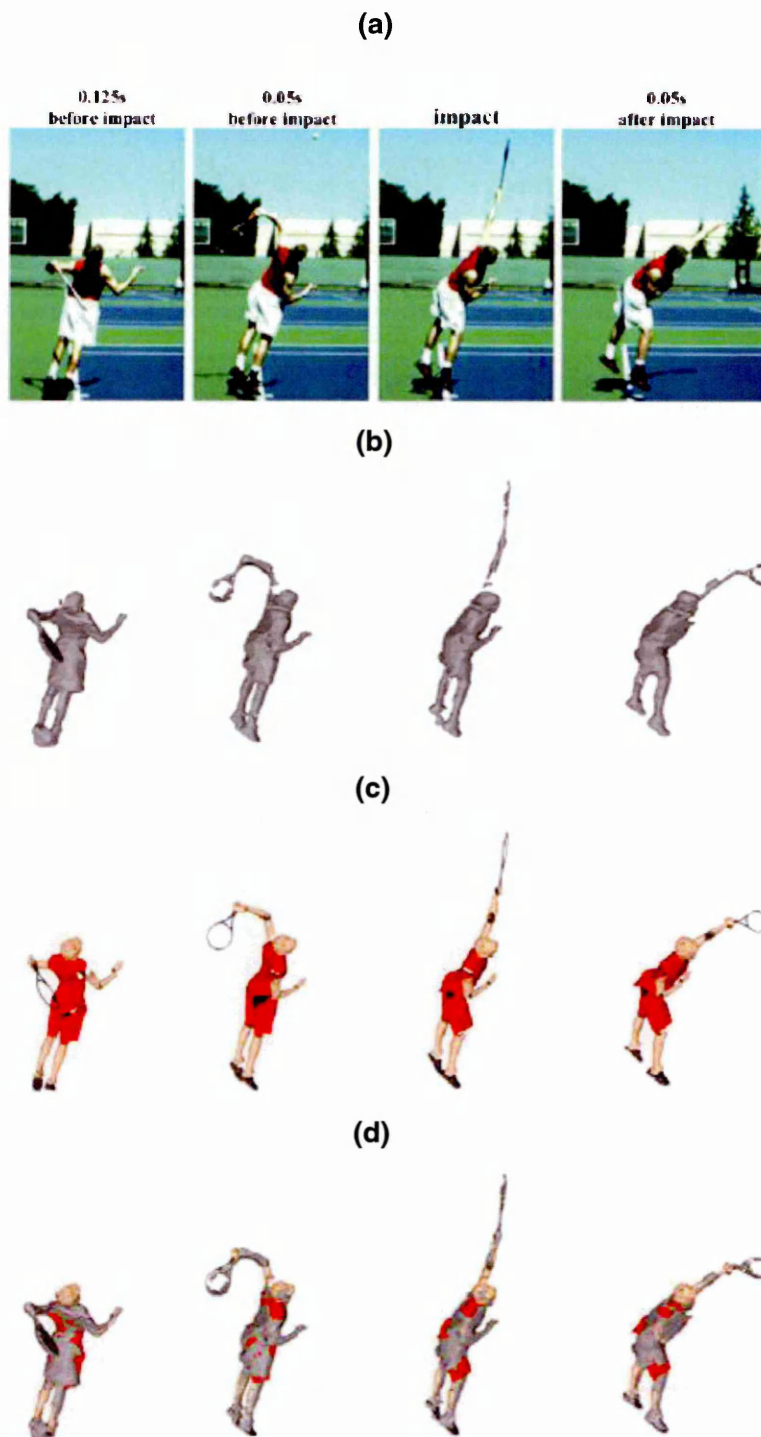


Figure 2.22 (a) Footage from camera four of a representative service trial, (b) corresponding visual hulls, (c) tracked subject-specific segment models and (d) correspondence between visual hull and model surfaces (taken from Sheets et al. 2011).

During tracking, the Levenberg-Marquardt optimisation routine was used to minimise reprojection error distances between the surface of the visual hull (constructed using the eight cameras) and the subject-specific model. Optimisations were initialised by repositioning the subject-specific model (from a reference pose) to match the first visual hull with one arm and the racket above the head. To reduce chances of the optimisation converging to a local minimum, it was repeated eighty times with

perturbations to the initial candidate pose. Once the optimisation had converged the solution was used as a candidate pose for the next frame. Position of the racket centre of volume and the upper limb joint centres was measured with respect to the court reference frame.

Sheets et al. (2011) measured racket velocities and upper limb joint centre trajectories during flat, kick (topspin) and slice (sidespin) serves. Peak racket velocity was significantly slower during a kick serve (24.6 m s^{-1}) than that of a flat or slice serve (26.1 m s^{-1}) with an average uncertainty of $\pm 0.54 \text{ m s}^{-1}$. These are lower than average peak racket head velocity of 34.8 m s^{-1} reported by Elliott, Marsh and Blanksby (1986). Sheets et al. (2011) contend the difference in peak racket velocities between this study and others (Chow et al. 2003, Reid, Elliott and Alderson 2007) is mainly due to the point tracked on the racket. Sheets et al. (2011) used the centre of volume of the racket's visual hull, which is closer to the handle than the tip of the racket tracked by Elliott, Marsh and Blanksby (1986), Chow et al. (2003) and Reid, Elliott and Alderson (2007).

The model-based markerless tracking method presented by Sheets et al. (2011) requires eight calibrated high-speed cameras, which come at a cost and are difficult to set up rendering its use restricted. In addition, such strategic placement of the eight cameras in relatively close proximity to the court and player is not practical for real-play conditions. This MMC system likely requires help from the participant or other assistants with the set up, calibration and high-speed camera operation.

Pansiot (2009) used visual hulls in favour of monocular frameworks for tennis stroke recognition. Calibration was conducted using the 2D planar method (Zhang 1999) which was implemented using Bouguet's (2010) camera calibration toolbox for MATLAB. Pansiot (2009) found views of visual hulls performed better than 2D segmented binary views in allowing strokes to be recognised. However, only four cameras were used and some silhouettes were poorly segmented therefore the overall quality of the visual hulls was low. Pansiot (2009) improved stroke recognition rate using spherical harmonic decomposition by effectively reducing postural information obtained from visual hulls into a lower dimension descriptor. However, the extreme compactness of the spherical harmonics meant that subtle changes in player posture or motion went undetected.

Ceseracciu et al. (2011) modified methods developed by Corazza et al. (2010) by using subject-specific segment models to obtain kinematic information from visual hulls of

swimmers. The calibration process consisted of two image sequences. The first involved filming a series checkerboard images on dry land which were passed to Bouguet's (2010) camera calibration toolbox for MATLAB in order to iterate an accurate set of parameters from initial assumptions. The second sequence acquired images of a 12 control point calibration grid measuring 2 x 1.1 x 1.1 m to obtain the relative pose of 6 synchronized underwater cameras. Following correction of the calibration parameters for underwater application, Ceseracciu et al. (2011) minimised the squared error distances between paired vertices of the segment models and the visual hulls using the Levenberg-Marquardt method. To test the accuracy of the proposed markerless system joint centre trajectories were compared with those reconstructed using manual marker digitisation. Over the five participants tested, resultant root mean squared distances (RMSD) between the markerless and manually tracked trajectories were 177 ± 83 , 140 ± 32 and 74 ± 28 mm for the shoulder, elbow and wrist joints respectively. Calibration of only six cameras limited the accuracy of this study.

Studies that have used visual hulls for biomechanical analysis of human movement in sport rely on multiple calibrated high-speed camera systems which are expensive and inefficient to set up in real play conditions. Furthermore, a minimum eight cameras are required to build visual hulls of adequate accuracy. Such systems are therefore not user friendly for players, coaches and manufacturers; requiring highly trained individuals in calibration and high-speed camera operation. However, another area of work uses the fully calibrated silhouette set alone, to estimate the 3D position associated with single uncalibrated views of a 3D object. Since an explicit visual hull is not required, these *single view fitting* methods could alleviate the problems related to operating multiple high-speed cameras in the field.

2.3.9 *Single view fitting*

Previous work has focussed on maximising the quality of visual hulls (Matusik et al. 2000, Wong 2001). This has allowed visual hulls to be used for measurement of size and shape properties of 3D objects (Forbes, Voigt and Bodika 2003, Forbes 2007). Forbes, Voigt and Bodika (2003) developed a method to merge fully calibrated silhouette sets, increasing the number of associated views and in turn visual hull quality. Calibration parameters were acquired using a self-calibration technique which used silhouette images of different size ball bearings. The parameters were calculated by maximising the silhouette consistency between the ball silhouettes and absolute scale was enforced because the size of the balls was known. In order to merge two sets, the Levenberg-Marquardt method was used to adjust the relative pose parameters between the sets to minimise the reprojection error distances for each

silhouette pairing. Forbes (2007) merged ten, six-view silhouette sets of stones and found that the merged sets provided greater volume estimation accuracy compared to the original six-view sets.

Price and Morrison (2007) adopted methods developed by Forbes, Voigt and Bodika (2003) and merged views of a particle to form a 60-view fully calibrated silhouette set. Using the set, Price and Morrison (2007) developed a single view fitting optimisation framework. This involved using the Levenberg-Marquardt method to adjust the relative pose between a silhouette whose true pose is unknown and each silhouette in the fully calibrated 60-view set, in order to minimise ETE distances. Figure 2.23 is a simplified illustration of the method showing a 6-view fully calibrated set and the unknown viewpoint associated with the particle's pose.

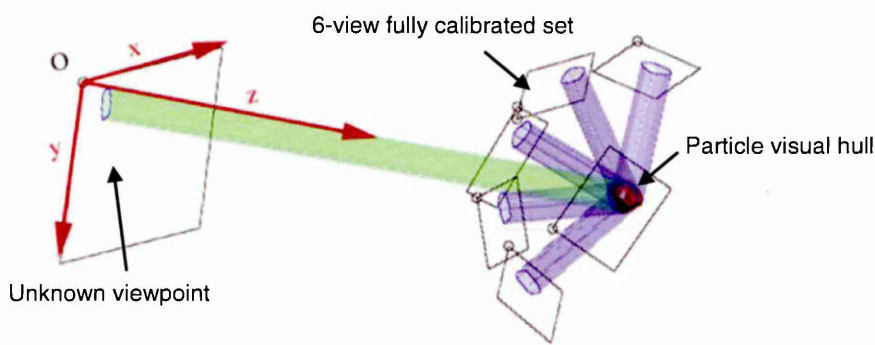


Figure 2.23 A 6-view fully calibrated silhouette set associated with a six-camera setup, the resulting visual hull model of the particle created by intersecting all the visual cones. The 'unknown' viewpoint associated with the particles pose is shown with respect to the fully calibrated set (adapted from Price and Morrison 2007).

Price and Morrison (2007) found that initial pose estimate accuracy increased with the irregularity of a particle's shape and the number of views in the fully calibrated silhouette set. This was because for regular shaped particle's, more than one associated pose could be equally consistent. Once the initial pose had been estimated, Price and Morrison (2007) applied the method to estimate the trajectory of non-contacting simulated particle trajectories. This involved using a rigid body motion model to predict the pose of the particle during simulated high-speed video frames in which its motion remain unhindered. It was found that low ETE values (a threshold of 1.5 pixels was used) corresponded to lower angular and velocity error for particle trajectories. Price and Morrison (2007) warn however, that the accuracy of the method is sensitive to the initial candidate relative camera pose, especially for regular shaped particles. In their study a maximum of 100 candidate relative poses were used and orientations were expressed as quaternions and repeatedly sampled from a normal distribution.

The conference paper by Price and Morrison (2007) was relatively brief, considering the complexity of the method. There was no information regarding camera calibration and silhouette extraction, which are the main contributors to error in silhouette-based pose optimisation methods (Forbes 2007). Furthermore, the estimated trajectories were limited to trials using synthetic silhouettes. Price and Morrison (2007) relied on using low ETE values i.e. below the 1.5 pixel threshold, as a performance metric. However, it is difficult to quantify such values which have no physical meaning. Finally, the influence of symmetrically shaped objects (such as a racket) on the accuracy of the method was not investigated.

In a similar study, Wong and Cipolla (2004) created an initial visual hull using a camera and uncalibrated images captured under circular motion. Calibration parameters were estimated using the correspondences obtained from the epipolar tangents and the constraint of circular motion. To improve the initial visual hull, Wong and Cipolla (2004) minimised the ETE in order to estimate the pose associated with arbitrary silhouette views. In Figure 2.24 views one to eleven were those assumed to be constrained under circular motion forming an initial visual hull which was refined by addition of views twelve to fifteen. Wong and Cipolla (2004) used a camera to capture all the views in Figure 2.24, highlighting that single view fitting methods do not necessarily require multi-camera systems to create fully calibrated silhouette sets.

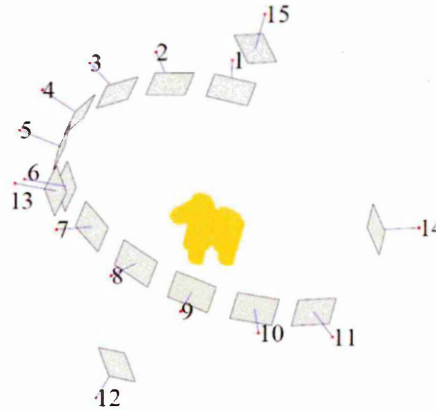


Figure 2.24 Arbitrary camera pose numbers 12 to 15 estimated from the circular fully calibrated silhouette views 1 to 11 (taken from Wong and Cipolla 2004).

Wong and Cipolla (2004) did not provide ground truth poses obtained using an alternative and acknowledged calibration technique in order to validate the accuracy of the camera poses in Figure 2.24. Wong (2001) demonstrated lower ETE values associated with camera poses obtained using the direct linear transformation (DLT) method (Abel-Aziz and Karara 1971) compared with the circular motion algorithm developed by (Wong and Cipolla 2004). This suggests Wong and Cipolla's (2004) calibration method provides less accurate parameters than conventional techniques.

The success of the method developed by Wong and Cipolla (2004) was judged on the aesthetic appearance of the visual hull alone.

Azevedo et al. (2008) and Tavares, Azevedo and Vaz (2008) used a camera and two separate checkerboard image sequences to create a fully calibrated set in order to reconstruct visual hulls. Camera intrinsic parameters were obtained from the first sequence, which involved capturing images of a checkerboard in different orientations (sample images 1 to 3 Figure 2.25). In the second sequence, the object to be reconstructed was placed on the checkerboard, keeping the camera untouched the sequence was acquired by turning the checkerboard (and object) until a full rotation (12 images) had been performed (sample images 4 to 6 Figure 2.25).

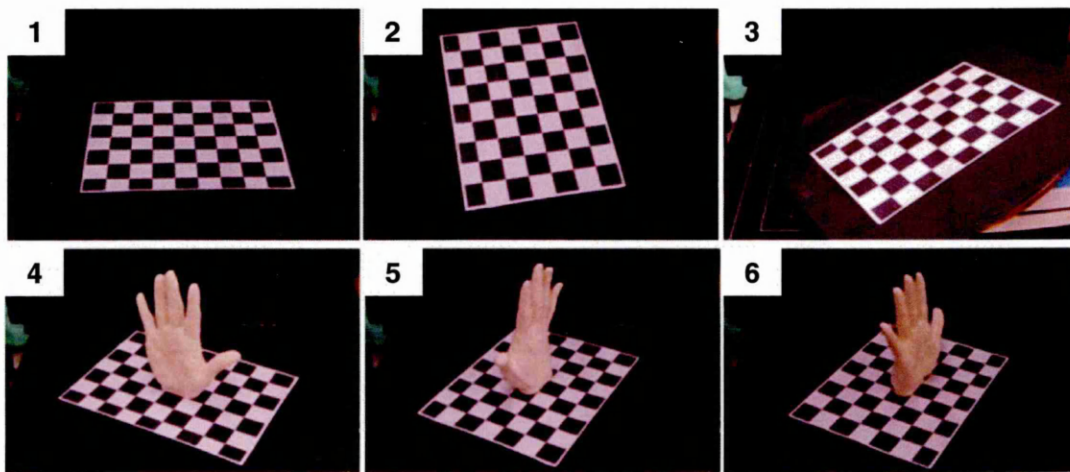


Figure 2.25 Images 1 to 3 are examples from the first checkerboard sequence used to obtain the intrinsic parameters. Images 4 to 6 are examples from the second sequence and were used to determine camera pose parameters (adapted from Tavares, Azevedo and Vaz (2008)).

Tavares, Azevedo and Vaz (2008) created a circular configuration of camera poses from which the associated silhouette views were extracted. To overcome the checkerboard intersections being occluded in the second image sequence, Tavares, Azevedo and Vaz (2008) manually selected four reference points in each image so that visible pattern points could be ordered. Tavares, Azevedo and Vaz (2008) used Bouguet's (2010) camera calibration toolbox for MATLAB and noted that calibration accuracy decreased with the number of visible pattern points. Azevedo et al. (2008) and Tavares et al. (2008) only constructed visual hulls for non-symmetrical objects and did not use view fitting methods to refine the hulls, as in Wong and Cipolla (2004) and Wong (2001).

A view fitting method could be applied to estimate the 3D position of a racket. Similar to the objects used by Wong and Cipolla (2004) and Price and Morrison (2007) a racket can be assumed to be rigid. However, the methods demonstrated by Wong and Cipolla (2004) and Price and Morrison (2007) would need to be adapted and developed for use with a racket. For instance, a racket has a symmetrical shape with two axes of reflective symmetry (Srinivasan 2004). Therefore, the view fitting method would need to be developed to accurately estimate the pose associated with a racket that has reflective symmetry from some view points. Price and Morrison (2007) found that there was more ambiguity for pose estimation associated with regular shaped objects because more than one possible pose can be equally consistent. Furthermore, accurate camera calibration parameters (intrinsic and extrinsic) would be required to calculate the relative pose of a camera with respect to a racket. Such a method would not require multiple high-speed cameras nor an explicit visual hull, since it is the fully calibrated silhouette set which would be used to estimate the relative pose of uncalibrated views associated with a racket's 3D position.

2.3.10 Conclusion

Racket velocity and angle affect the trajectory of the ball and can be measured using different methods close to impact to help quantify performance. Marker-based and markerless method were considered not only in terms of their ability to track racket movement, but also practicality to be implemented into real-play conditions, cost and set-up time. Sensor systems were deemed unsuitable due to a corresponding increase in racket mass and MOI when attached to the frame, which could negatively influence players' swing mechanics. Furthermore, the possibility of obtaining accurate information about ball movement is uncertain with sensor systems.

Camera systems are advantageous because they are less invasive compared with sensor systems. To date, the least invasive and most user friendly passive marker-based system for tracking racket movements is that developed by Choppin, Goodwill and Haake (2011). However, the method requires manual digitisation of markers which can be time consuming and inaccurate. Existing non-invasive multi-camera systems such as Hawk-Eye come at a high cost and do not track racket movements. However, recent development of markerless systems that use visual hulls to track movements for biomechanical analysis present a non-invasive option for racket tracking. The method presented by Sheets et al. (2011), although at the forefront of racket tracking, requires a minimum of eight high-speed cameras. Furthermore, the user must be highly skilled in calibration and high-speed camera operation rendering the system impractical for players, coaches or manufacturers to use in real tennis-play conditions.

Single view fitting methods (Wong 2001 and Price and Morrison 2007) do not require multiple camera set ups; a camera can be used to capture different images of an object. Estimation of racket position using single view fitting methods has not been explored. Ultimately, the method would only require the single 'unknown' viewpoint associated with a racket's pose to be captured in real play conditions. The fully calibrated set could be obtained in the laboratory ahead of the data collection. Limiting the number of cameras required and as a result the cost and setup time. The success of the method would depend on accurate camera calibration and silhouette extraction. Therefore, an important first step would be the development of a method to estimate the relative pose of a camera with respect to a racket in order to create a fully calibrated set.

2.4 Chapter findings

Newly developed, motion analysis tools should provide measurements that are of practical use to the end-user, i.e. International Tennis Federation (ITF), players, coaches, manufacturers and broadcasters. A literature review has identified the following problem statement:

'Previous studies into racket performance fail to take into account how the racket will be used by the player. This is because an affordable, user friendly and non-intrusive method to estimate 3D racket position using a camera during real-play conditions does not exist'.

Most existing methods developed to measure racket movements rely on tracking some type of marker attached to the frame. However, it is impractical to apply markers to players rackets during real play conditions. Attachment of markers to the racket frame can distract players and potentially alter their swing mechanics (Sheets et al. 2011). Furthermore, manual digitisation of markers in the image is time consuming and can be inaccurate. However, due to the inherent difficulty in tracking complex racket movements, most existing non-invasive tracking methods concentrate on providing information solely about player and ball movements i.e. Hawk-Eye. Nevertheless, recent developments in MMC methods that use visual hulls have potential to advance current racket tracking methods.

An eight-camera MMC method that tracks subjects' visual hulls has previously been applied in tennis (Sheets et al. 2011). On average, Sheets et al. (2011) measured peak racket speeds to within ± 0.54 m/s during tennis serves by applying the MMC method

developed by Corazza et al. (2010). The MMC method developed by Corazza et al. (2011) produced an average deviation of 15 ± 10 mm with respect to marker-based results. However, calibration and placement of multiple high-speed cameras is not practical in real tennis-play conditions.

View fitting methods provide potential for tracking racket movements without the need for multiple high-speed cameras to be set-up on the court. The method would rely on a silhouette of a racket captured with a camera whose relative pose is unknown. A candidate relative pose could be used to measure the inconsistency between the unknown silhouette and a set of racket silhouettes captured with a fully calibrated camera (known intrinsic and extrinsic parameters). By adjusting parameters of the candidate relative pose to minimise the inconsistency (ETE), estimation of the true 3D position of the racket could be made.

A view fitting method would not require a visual hull of a racket; only the fully calibrated silhouette set is needed, which could be created in a controlled environment. Thus only the camera (whose relative pose is unknown) would be required to be positioned close to the court. Previously, view fitting methods have been applied to rigid objects under controlled laboratory conditions; such a method has never been applied in sport. Camera calibration (intrinsic and extrinsic) and silhouette extraction have been identified to be the main limiting factors of silhouette-based pose optimisation methods.

Most MMC studies with a biomechanical focus use offline techniques in order to calibrate multiple camera systems. Offline calibration techniques usually make use of specific patterns or objects to determine parameters. The 2D-planar method devised by Zhang (1999) is often favoured for its accuracy, flexibility and ease of use for visual hull applications. Self-calibration techniques have also been employed which use the silhouette themselves to estimate camera parameters. However, these techniques are often limited to the application for which they are developed. Silhouette quality is determined by image resolution and the technique used to segment the boundaries. Polygonal segmented boundaries are favoured because high order B-spline snakes are more computationally expensive and thus allow more efficient methods to be developed (Matusik, Buehler, and McMillan 2001, Lazebnik 2002).

Figure 2.26 illustrates the main sources of error that should be considered for the development of a view fitting method for estimation of 3D racket position. The total error in Figure 2.26 is the combination of all the errors; systematic error arising from the calibration and error in the segmentation of silhouette boundaries.

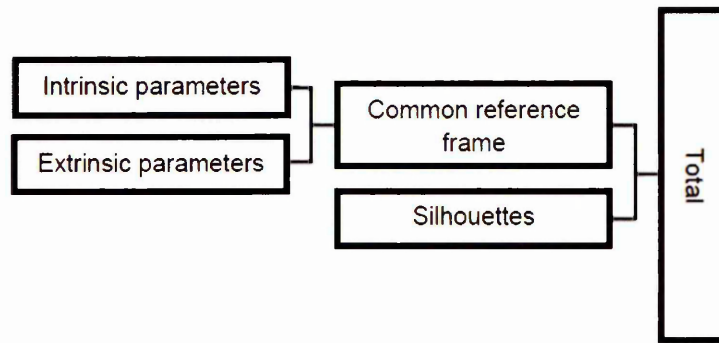


Figure 2.26 Camera calibration (intrinsic and extrinsic), determination of a common local reference frame and silhouette extraction are the main contributors to error in view fitting methods.

Based on Figure 2.26 a number of research questions can be identified in order that a view fitting method to estimate 3D racket position can be developed. The questions are as follows:

1. How can the relative pose of a camera with respect to a racket be accurately obtained?
2. What is the most appropriate method to obtain high quality racket silhouettes?
3. What is the most appropriate configuration for a fully calibrated set to be used for a view fitting method to estimate 3D racket position?
4. Can racket position be estimated in 3D without using markers and with a single camera?
5. Does the markerless method have potential to estimate 3D racket position in real play conditions?

2.4.1 Conclusion

Future tennis research should focus on furthering knowledge of racket movements during match play (Allen, Choppin and Knudson 2015). The literature review highlighted that development of a non-invasive racket tracking method was necessary. Such a method could be used alongside existing ball trajectory tracking systems i.e. Hawk-Eye and SpinDoctor (Kelley 2011) and player tracking tools (Dunn 2014). This would assist in furthering current understanding of ball-racket interactions during real play conditions. Separately, marker-based and markerless methods for tracking racket movements in real play conditions have been described. Currently, the application of MMC methods to tennis, particularly racket tracking, is limited. Development of a markerless view fitting method to track racket movements in real play conditions based on the approaches presented could allow for the influence of equipment on player

performance to be quantified. An appropriate accuracy criterion for the method is to estimate 3D racket position to within 15 ± 10 mm. This was the level of accuracy obtained for the MMC method developed by Corazza et al. (2010), which Sheets et al. (2011) applied to measure tennis serve kinematics. The first stage in the development of a view fitting method would be to devise an appropriate calibration method to determine the relative pose of a camera with respect to a racket.

2.5 Aim and objectives

The aim of this project is to develop and validate a novel markerless method to estimate 3D racket position.

Objectives:

1. To obtain accurate calibration parameters (intrinsic and extrinsic) for a camera using the 2D planar method.
2. To obtain the relative pose of a camera with respect to a racket using the calibration parameters obtained in objective 1.
3. To determine an appropriate silhouette image resolution to be used for the method by measuring the inconsistency of a fully calibrated silhouette set associated with a 3D racket model.
4. To use synthetic views of a 3D racket model to determine the optimum configuration for a fully calibrated set associated with a racket.
5. To create a fully calibrated set from real data and to validate 3D racket pose estimates against known racket positions.
6. To use synthetic views of a 3D racket model to assess the method's ability to estimate 3D racket position during simulated play conditions.
7. To make recommendations for future development of the method for application in real-play conditions.

3 Intrinsic camera calibration

3.1 Introduction

The literature review identified camera calibration as an important precursor of the development of a markerless method to track racket movements. The first step in the calibration process is to obtain the intrinsic parameters (Figure 3.1). Figure 3.1 shows the place of the intrinsic parameters amongst the main sources of error that contribute to the accuracy of a view fitting method.

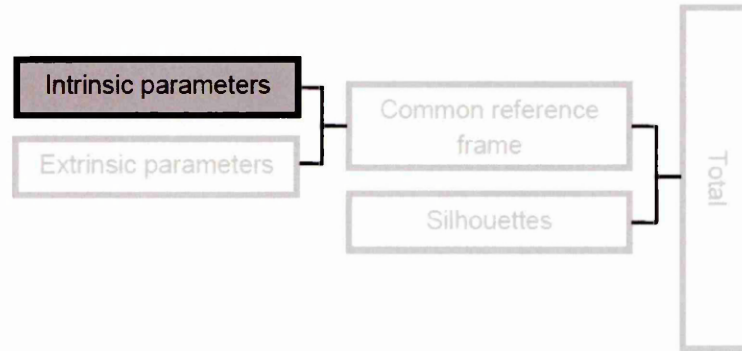


Figure 3.1 Chapter 3 calculated camera intrinsic parameters and quantified the associated error.

The intrinsic parameters account for the effect of the camera's lens and the imperfections in the camera's construction on the images captured. A view fitting method relies on a fully calibrated set, comprising of multiple camera poses, with the intrinsic parameters known. Therefore, it is necessary to determine the conditions for a repeatable and accurate intrinsic camera calibration. The literature showed that the 2D planar calibration method devised by Zhang (1999) has been used previously in 3D racket position estimation studies (Choppin et al. 2011, Allen, Haake and Goodwill 2011, Sheets et al. 2011). Zhang's (1999) method was selected due to the accuracy, flexibility and ease of use it provides. However, in order for this technique to be effective for the purpose of this project, a number of methodological concerns need to be addressed to ensure its accuracy.

3.2 Aim and objectives

The aim of this chapter is to use the 2D planar calibration method to obtain repeatable and accurate intrinsic parameters for a camera. This intrinsic camera model will be used throughout this project for development of a view fitting method to estimate 3D racket position.

Objectives:

1. To consider the practical steps that influence the quality of a calibration.
2. To perform a series of intrinsic calibrations using the steps from objective 1.
3. To assess the repeatability of the model generated in objective 2.
4. To assess the accuracy of the model to reconstruct 3D points.

3.3 Calibration procedure

The popular and convenient implementation of Zhang's (1999) algorithm provided in the form of the camera calibration toolbox for MATLAB (Bouguet 2010) was used. It relies on images of a planar checkerboard pattern of known geometry to provide 2D image coordinates. The image coordinates are used to compute planar homographies between the camera and checkerboard (Kelley et al. 2010, Choppin et al 2011, Dunn et al. 2014).

3.3.1 Checkerboard manufacture

The repeatability and accuracy of the 2D planar method relies on the accuracy of the checkerboard used i.e. its regularity and precise knowledge of the square size ensures minimisation of errors. A checkerboard calibration uses the physical size of each square to scale the calibration volume hence, if a square is measured incorrectly there will be a systematic error in every result (Choppin 2008). Therefore, the checkerboard used in this project was manufactured using:

- A solid section of wood, which had a flat surface and measured 18 mm thick. This was necessary to ensure the checkerboard would not bend when held in different orientations.
- A printed checkerboard pattern on high quality matt paper which was attached to the surface of the wood section using spray adhesive. Care was taken to spray the adhesive evenly so no air pockets appeared under the checkerboard surface.

3.3.2 Camera settings

Before a calibration it is important that the cameras are set up properly so that the images are compatible with Bouguet's (2010) toolbox for MATLAB. Lens aperture, focus, shutter speed and lighting all influence the final image.

- **Aperture:** controls the amount of light entering the camera and the depth of field. A smaller camera aperture is preferable, since this increases the depth of field resulting in the images staying in focus over a greater range of checkerboard movement.

- **Focus:** the camera should be focussed at the centre of the control volume to ensure the checkerboard intersections appear clear in the image.
- **Shutter speed:** since checkerboard movement is generally slow, a low shutter speed can be used to allow more light on to the checkerboard. The camera should be mounted on a tripod to avoid the images blurring.
- **Lighting:** if performed indoors artificial light should be used to allow for lower camera aperture. However, the optimum amount of artificial light depends on the environmental surroundings; if not enough light illuminates the checkerboard, the intersections will be unidentifiable. However, if too much lighting is used the checkerboard intersections can 'bloom', effectively erasing the intersections, rendering the images useless.

3.4 Recording calibration images

With the camera set up optimally, images should be captured at the highest available resolution. Choppin (2008) assessed the accuracy of Zhang's (1999) 2D planar calibration method and found that when converted into millimetres, calibration accuracy decreased with checkerboard resolution. Zhang (1999) showed that as the number of checkerboard images increase calibration error reduces. Zhang (1999) recommends that a minimum of 15 calibration images are used to provide the best results. However, to obtain sufficiently accurate intrinsics, the checkerboard images need to cover the majority of the field of view of the camera (Kelley 2014). Therefore, this may require recording in excess of 15 images to ensure full image coverage.

Zhang (1999) also investigated the effect of the angle between the checkerboard surface and the image plane on 2D point reprojection accuracy. The failure rate was high when the planes were almost parallel and at angles above 60°. Therefore, a selection of checkerboard angles between 20 to 50° relative to the camera's image plane is most appropriate (Zhang 1999). It is vital that the checkerboard occupies as much of the image as possible during calibration. The more pixels it is composed of, the easier it will be for the software to resolve the intersections providing a more accurate calibration. However, if the checkerboard is held too close to the camera, the required range of orientations would not be possible. Furthermore, it would not be possible to obtain control points close to the image periphery due to the size of the squares.

Checkerboard images are passed to Bouguet's (2010) camera calibration toolbox for MATLAB. For each checkerboard image, 2D image coordinates are extracted from the intersections between the white and black squares (Figure 3.2 (a) and (b)). The

calibration software semi-automatically detects the intersections by interpolating along the high contrast surface of the checkerboard within a predefined search region. It is able to determine the location of the intersection to within a 1000th of a pixel and account for lens distortion.

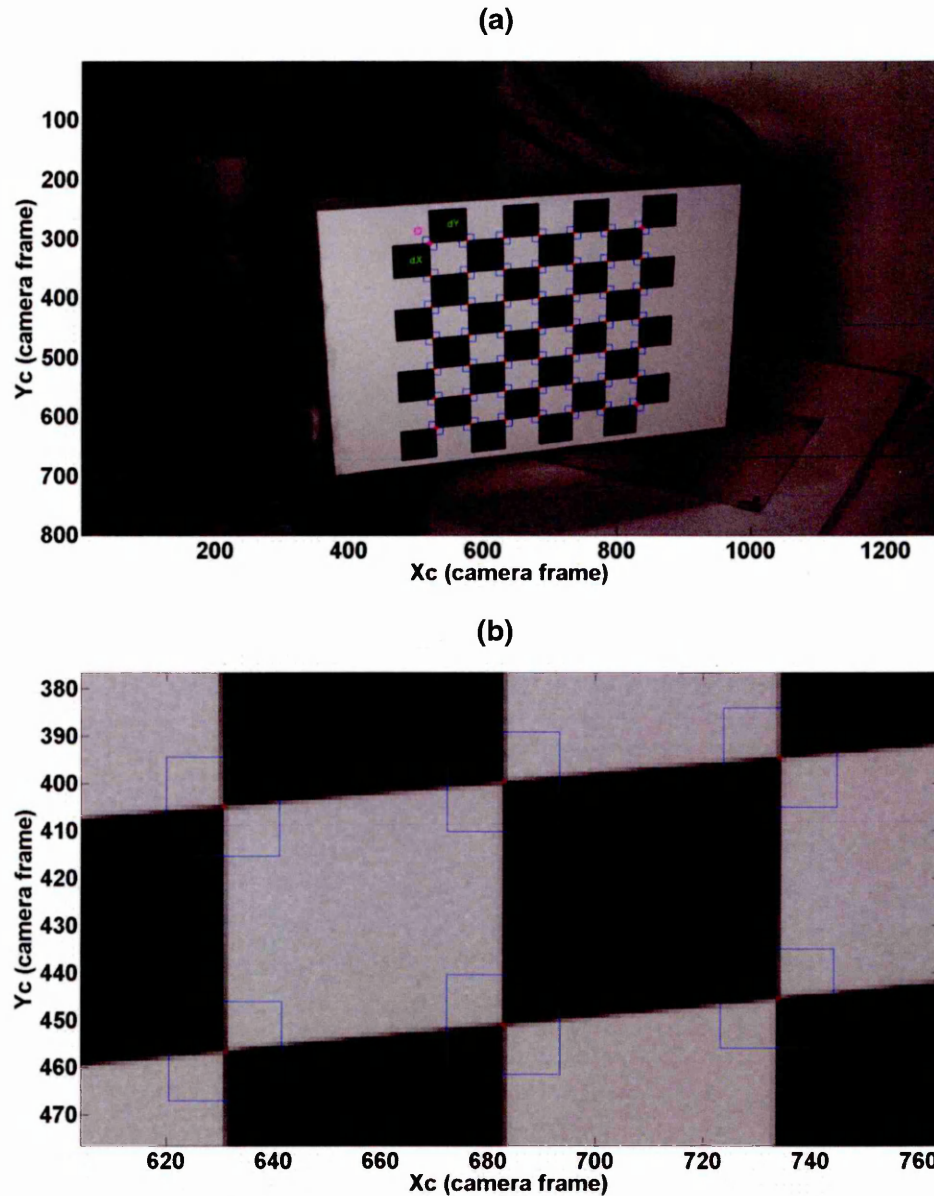


Figure 3.2 A checkerboard image (a) after semi-automatic intersection detection and (b) at a higher zoom level showing the intersections as red points within the search region.

Once all the intersections have been detected for a set of calibration images, the intrinsic parameters can be generated. The parameters can be used to re-project coordinates back onto checkerboard image intersections with knowledge of the physical size of each square. The pixel discrepancy between the reprojected coordinates and those originally detected by the software indicates the quality of the calibration. The parameters can be improved by running the calibration over a sub-set of the distortion coefficients i.e. by providing more realistic initial conditions. In some

cases, increasing the size of the search regions (blue squares in Figure 3.2 (a) and (b)) can improve intersection detection however, this could also increase the chance of incorrect detection.

3.5 Summary

By identifying potential sources of error in the calibration procedure, steps can be taken to ensure they are minimised.

- Camera set up is important to ensure the checkerboard intersections are clear in the image. This ensures the control points on each image are detected by the software.
- Previous work has shown a range of checkerboard orientations provide more accurate calibration results. In practice, the orientations are based on human judgement; a range between 20 and 50° are most appropriate.
- The quality of the calibration can be maximised by ensuring the checkerboard intersections are detected and by using the appropriate lens distortion model so the best parameters are generated. This relies on capturing high quality calibration images.

By taking into account the aforementioned considerations, the following section will perform a full error analysis for a series of intrinsic calibrations. This is necessary to identify an appropriate intrinsic model to be used throughout this project. The model will be used to create an accurate fully calibrated set for development of a view fitting method to estimate 3D racket position.

3.6 Intrinsic calibration error analysis

The practicalities of a 2D planar calibration that could induce error have been identified. This section investigates the repeatability of the intrinsic parameters and assesses their accuracy by using the parameters to reconstruct the geometry of the calibration plane.

3.7 Method

The camera used for this study was a black and white high-speed Vision Research Phantom Miro M110 camera. The F-stop value (aperture) used was F22 and the shutter speed was set to 1/8th of a second. Images were captured at a resolution of 1280 x 800 pixels, which was the maximum resolution available. A checkerboard of 8 x 8 squares each measuring 50 x 50 mm was used, which was not too large or heavy to be moved around. It was positioned approximately 2 m from the camera during the calibration, to allow for a range of orientations (between 20 and 50° (Zhang 1999)) and

to obtain good image coverage. Ten separate intrinsic calibrations were performed, each comprising of approximately 60 checkerboard images to ensure good image coverage.

The calibrations were run over different sub-sets of distortion coefficients to obtain the best parameters. The two distortion models used were a complete model (radial + tangential) and a 4th order radial distortion model without a tangential component. The repeatability across the ten calibrations using both distortion models is presented below.

Accuracy was assessed by reconstructing the coordinates on the calibration plane (Dunn et al. 2012), using the parameters. This involved placing a checkerboard flat on the ground and reconstructing its geometry using the known intersection points and calculated intrinsic parameters from each of the ten calibrations. Mean reconstruction error \pm standard deviation across the ten calibrations using both distortion models is presented.

3.8 Results

3.8.1 Image coverage

Figure 3.3 shows the image coverage achieved with approximately 60 checkerboard images for one of the ten calibrations. Each red dot corresponds to a checkerboard intersection. Checkerboard squares of this size allowed control points to be obtained close to the image periphery. These experiments were also used to check whether full image coverage (and accurate intrinsic parameters) was possible around a racket i.e. the checkerboard was angled around the racket (Figure 3.3).

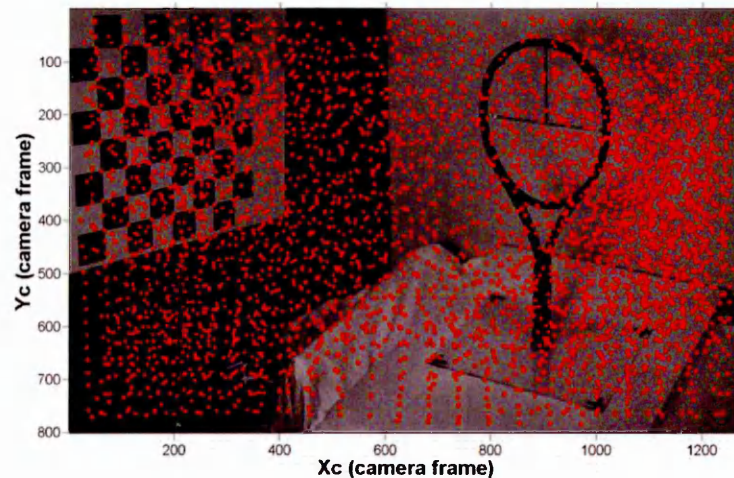


Figure 3.3 Image showing the coverage achieved; each red dot represents a checkerboard intersection.

3.8.2 Lens distortion

Full image coverage was necessary to accurately model lens distortion which is generally highest toward the image periphery. Figure 3.4 (a) and (b) show radial and tangential lens distortions respectively, corresponding to the input passed to the toolbox illustrated in Figure 3.3. Due to the spherical shape of the lens, maximum pixel displacement was 2.5 pixels, located toward the image periphery (Figure 3.4 (a)). In Figure 3.4 (b), maximum pixel displacement due to tangential lens distortion was 0.7 pixels.

Zhang (1999) justified a 4th order radial distortion model which has no tangential component by the fact that most modern lenses are manufactured with no imperfections. Since the tangential distortion measured was low, it was removed from the intrinsic model used in this study. Furthermore, Bouguet (2010) argues that it is not necessary to push the distortion model beyond the 4th order radial for standard non-wide angled lenses.

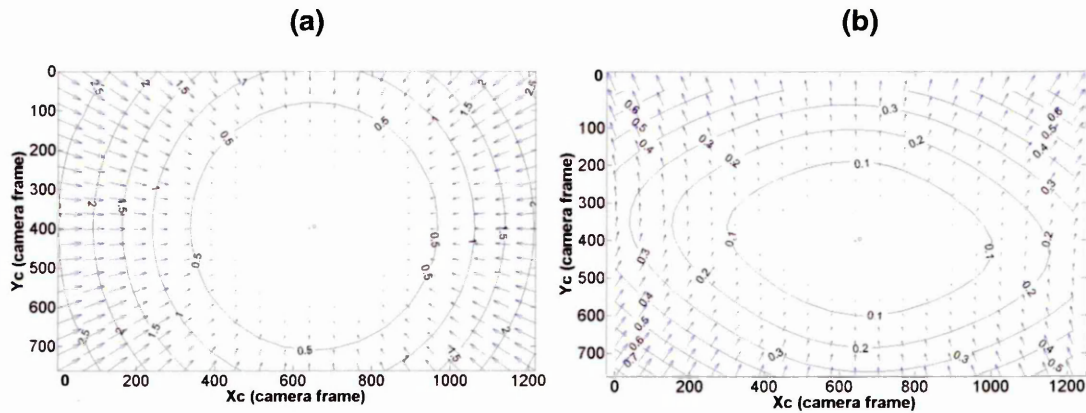


Figure 3.4 Illustration of the (a) radial and (b) tangential lens distortion; the numbers indicate the pixel displacement which is highest towards the image periphery.

3.8.3 The intrinsic model

Figure 3.5 shows an example of how well the 4th order radial distortion model (red circle) fitted the original pixel coordinate detected by the software (green crosshair).

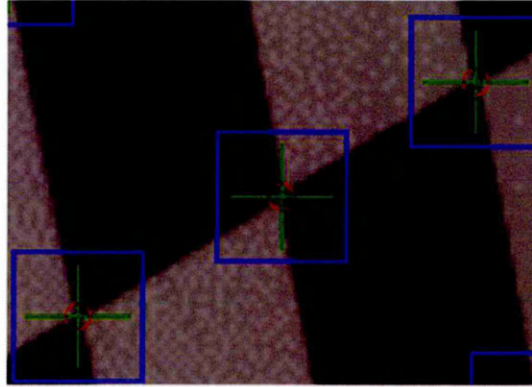


Figure 3.5 Image showing the fit of the intrinsic model (red circle) with respect to the point detected by the software (green crosshair).

The pixel discrepancy between all reprojected intersection points and points detected by the software for one of the ten calibrations is shown in Figure 3.6. Pixel points were located to within ± 0.04 of pixel in the X and Y image axes.

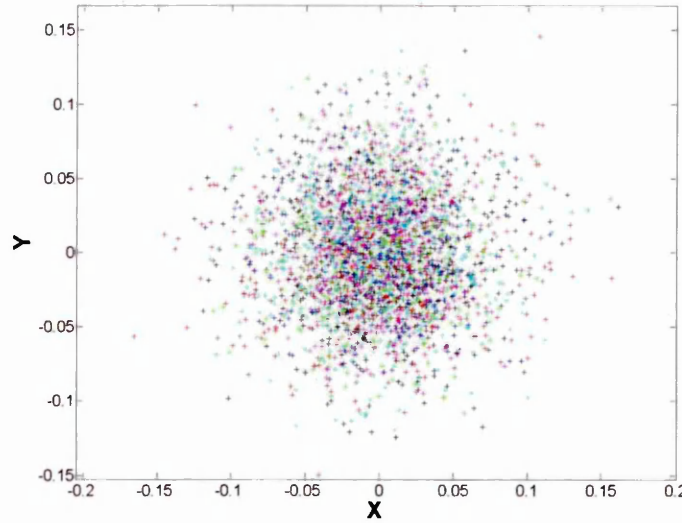


Figure 3.6 Pixel discrepancy between all reprojected intersection points and points detected by the software for one of the calibrations.

3.8.4 Repeatability

Bouguet's (2010) camera calibration toolbox for MATLAB uses two different focal lengths; f_x and f_y in the u and v image axes respectively (expressed in pixel units). The reason for two different focal length values is that individual pixels are not guaranteed to be exactly square (Bradski and Kaehler, 2008). However, values for f_x and f_y are

usually very similar. Figure 3.7 (a) shows f_x focal length values estimated using the complete (radial + tangential) and the 4th order radial distortion model without a tangential component respectively, over the ten intrinsic calibrations. The mean values for f_x were 1779.02 ± 1.47 and 1779.05 ± 1.31 pixels (indicated by the circle in Figure 3.7 (a)) for distortion models with and without a tangential component respectively. Removal of the tangential component decreased the range in f_x from 4.92 to 4.39 pixels across the 10 calibrations.

Figure 3.7 (b) shows f_y focal length values estimated using the complete and the 4th order radial distortion model without a tangential component respectively, over the ten intrinsic calibrations. Mean f_y increased from 1776.88 ± 1.55 to 1777.10 ± 1.26 pixels when the tangential component was removed. Furthermore, removal of the tangential component decreased the range in f_x from 5.48 to 4.41 pixels across the 10 calibrations.

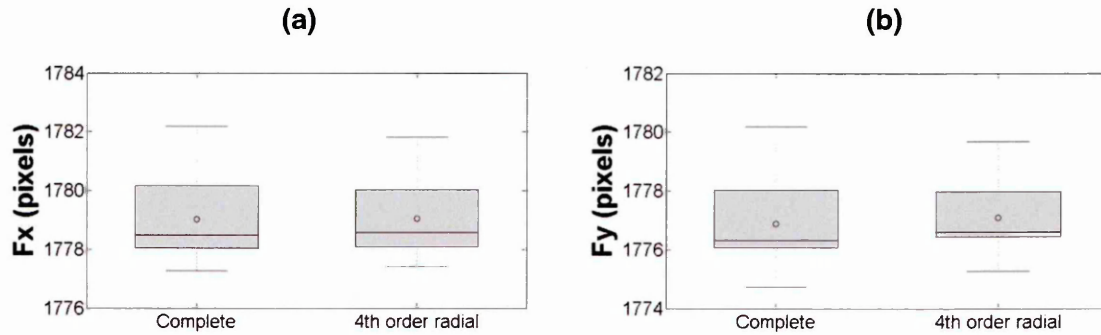


Figure 3.7 The distribution of focal length values for (a) f_x and (b) f_y across the ten intrinsic calibrations, showing the mean (circle), median, upper and lower quartiles and upper and lower deciles.

The two parameters of the principal point c_x and c_y describe the displacement in the u and v image axes respectively away from where the optical axis intersects the centre of the image plane (Bradski and Kaehler 2008). Figure 3.8 (a) shows values for the principal point in the u image axis, estimated using the complete and the 4th order radial distortion model without a tangential component respectively. The mean value for the principal point in the u image axis was 651.08 ± 1.52 and 651.74 ± 0.50 for when the tangential component was included then removed respectively. The range across the ten calibrations decreased by 3 pixels when the tangential component was removed from the model.

Figure 3.8 (b) shows values for the principal point in the v image axis, estimated using the complete and the 4th order radial distortion model with without a tangential component respectively. The mean value for the principal point in the v image axis was

393.99 \pm 3.58 and 400.33 \pm 0.58 for when the tangential component was included and removed respectively. Range in c_y across the ten calibrations also decreased when the tangential component was removed from the model.

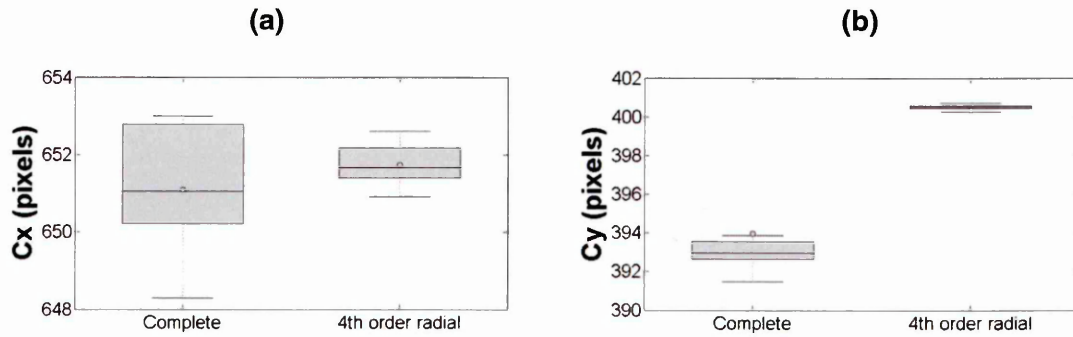


Figure 3.8 The distribution of principal point values for (a) c_x and (b) c_y across the ten intrinsic calibrations, showing the mean (circle), median, upper and lower quartiles and upper and lower deciles.

3.8.5 Calibration error

Pixel projection error (as shown in Figure 3.6) across the ten calibrations is shown in Figure 3.9. Mean projection error was 0.041 ± 0.002 pixels in the u and v image axes. Furthermore, there was not a significant difference ($p > .05$) in projection errors between the two distortion models. This suggests inclusion of the tangential component in the distortion model has no impact on the accuracy of the calibration.

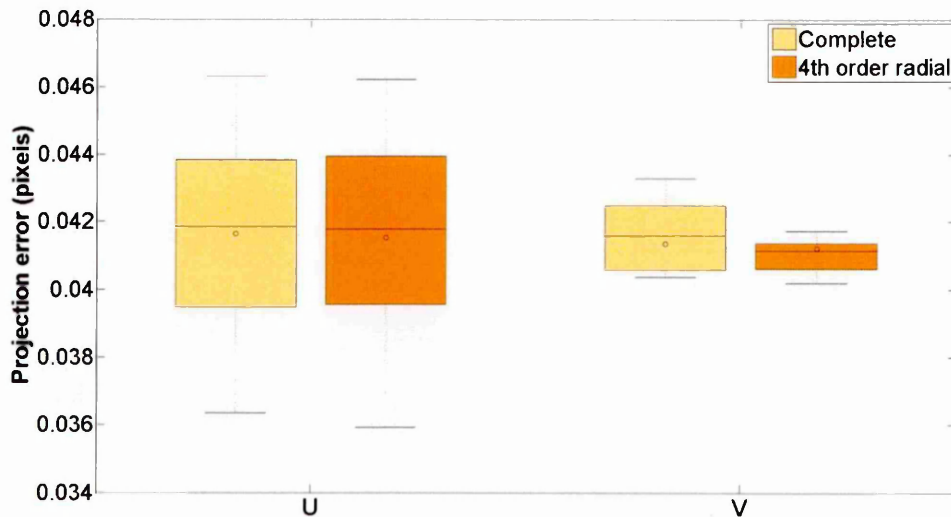


Figure 3.9 Pixel projection error in the u and v direction for each of the 10 intrinsic calibrations, showing the mean (circle), median, upper and lower quartiles and upper and lower deciles.

Figure 3.10 shows mean reconstruction error \pm standard deviation across the ten calibrations for the 49 coordinates (red dots) on the calibration plane (checkerboard) in the X, Y and resultant (R) directions. The reconstructed checkerboard position was not used in any of the intrinsic calibrations and reconstruction was performed once using parameters from each of the ten calibrations. The relative position of the camera with respect to the checkerboard origin (red dot in the top the left corner) was 1951, -34 and 1085 mm in the X, Y and Z directions respectively. Mean reconstruction error was marginally lower in each direction when the intrinsic parameters were estimated using a 4th order radial distortion model without a tangential component. Standard deviation values were also lower when this model was used. Reprojection errors in the Y direction represented the largest component of error.

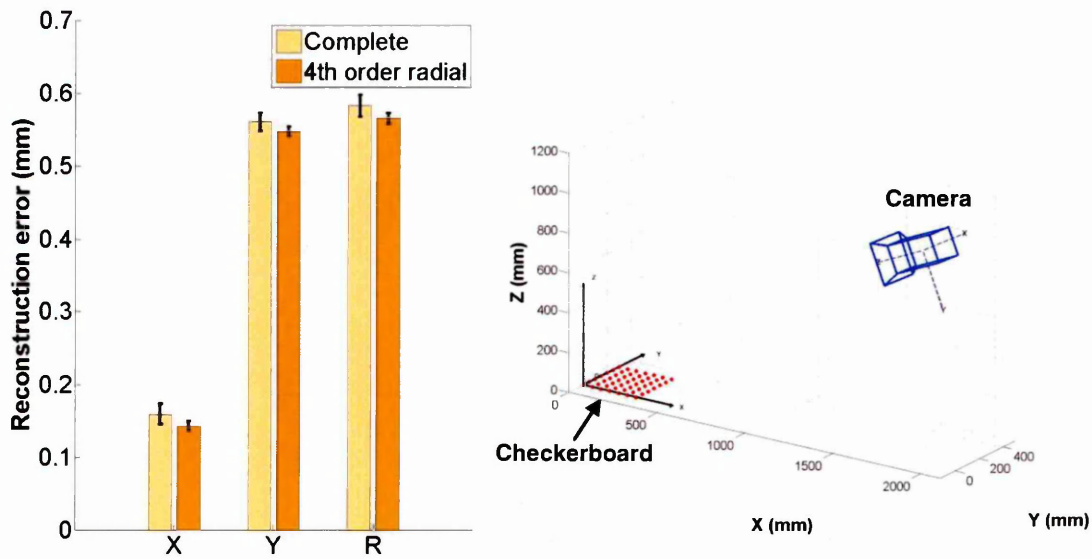


Figure 3.10 Mean reconstruction error \pm standard deviation error across the ten calibrations for coordinates on the calibration plane in the X, Y and resultant (R) directions.

3.9 Discussion

This chapter has highlighted the considerations which should be taken into account when trying to maximise the repeatability and accuracy of a camera's intrinsic parameters. These are:

- To ensure the checkerboard is manufactured to a high standard. The checkerboard must have a flat surface and the intersections accurately measured.
- To ensure maximum possible image coverage is achieved. This requires a checkerboard with the appropriate size squares that allow for control points to

be obtained close to the image periphery. Providing more repeatable and accurate intrinsic parameters (Kelley 2014).

- To ensure a range of checkerboard orientations are captured. Based on findings by Zhang (1999) orientations between 20 and 50° provide the best results. A square size of 50 x 50 mm allowed the software to detect all the intersections for all the checkerboard orientations.
- To ensure camera set up is optimum for the environmental conditions used, so that the checkerboard intersections are clear in the image.

3.9.1 Repeatability

When taking into account the methodological considerations highlighted in this chapter; camera intrinsic parameters were found to be more repeatable when a 4th order radial distortion model without a tangential component was used. This was compared to using a complete distortion model (radial + tangential). A 4th order radial distortion model (without a tangential component) was used by Zhang (1999) and is justified by the fact that the lens used in this study was manufactured to a high standard and thus was aligned with the camera sensor. The tangential distortion associated with the lens used was illustrated in Figure 3.4 (b), maximum induced pixel displacement was only 0.7 pixels. Radial distortion was higher; maximum induced pixel displacement was 2.5 pixels in the image corners (Figure 3.4 (a)). Therefore, it is necessary to estimate the intrinsic parameters by including radial distortion in the model.

Figure 3.7 showed that there was a marginal difference in the repeatability of the focal length (f_x and f_y) using the two distortion models tested. Mean values for f_x were 1779.02 ± 1.47 and 1779.05 ± 1.31 using a complete and a 4th order radial distortion model (without a tangential component) respectively. While mean values for f_y were 1776.88 ± 1.55 and 1777.10 ± 1.26 using a complete and a 4th order radial distortion model (without a tangential component) respectively. Previously, intrinsic parameter repeatability has received limited attention. Forbes (2007) reported an effective focal length standard deviation value in excess of 50 pixels obtained using a self-calibration technique. This was much higher than the values reported in this study, highlighting the accuracy of the 2D planar method when taking into account the methodological considerations outlined in section 3.3.

For principal point (c_x and c_y) there was a marked improvement in the repeatability when a 4th order radial distortion model without a tangential component was used (Figure 3.8). Mean displacement of the principal point in the u image axis (c_x) was 651.08 ± 1.52 and 651.74 ± 0.50 using a complete and a 4th order radial distortion

model (without a tangential component) respectively. Corresponding to a 67% improvement in the repeatability of c_x based on the standard deviation across the ten calibrations. Mean position of the principal point in the v image axis (c_y) was displaced from 393.99 ± 3.58 to 400.33 ± 0.58 when a complete distortion model was substituted for a 4th order radial distortion model (without a tangential component). This corresponded to an 84% improvement in the repeatability of c_y based on the standard deviation across the ten calibrations.

3.9.2 Calibration error

Figure 3.9 showed there is marginal decrease in pixel error when the intrinsic parameters were estimated using a 4th order radial distortion model without a tangential component. This translated to the planar reconstruction of coordinates on the calibration grid; intrinsic parameters estimated using the 4th order radial distortion model without a tangential component produced lower reconstruction errors (Figure 3.10). Therefore, the most repeatable intrinsic model is also the most accurate. Higher resolution calibration images could potentially improve reconstruction accuracy further.

3.10 Conclusion

This chapter provided an assessment of the checkerboard calibration method; identifying potential sources of error and providing methodological guidelines in order to produce repeatable and accurate intrinsic models for a camera. Therefore, an appropriate intrinsic camera model has been identified to be used throughout this project. Its reliability is an important factor and will have implications on the creation of a fully calibrated set for the development of a view fitting method to estimate racket position in 3D. The findings of this chapter are specific to the camera used and therefore are essential for this project to progress. Chapter 4 will investigate the influence of the intrinsic parameters on the extrinsic parameters by introducing a second camera.

4 Calibration of a stereo system

4.1 Introduction

Chapter 3 demonstrated the checkerboard method devised by Zhang (1999) was suitable for achieving repeatable and accurate intrinsic parameters for a camera. This chapter uses considerations from chapter 3 to investigate the influence of the intrinsic parameters on the extrinsic parameters (Figure 4.1).

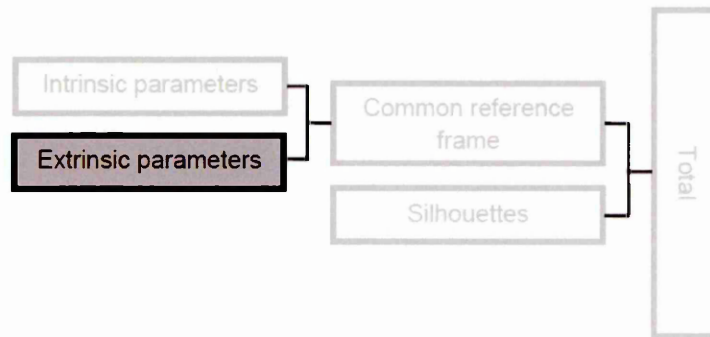


Figure 4.1 Chapter 4 calculated the extrinsic parameters for a stereo pair and quantified the associated error.

The relative pose (rotation and translation) of a stereo camera pair is calculated and acceptable errors one might experience are quantified. This was an important preliminary step for obtaining the relative pose of a camera with respect to a racket.

4.2 Aim and objectives

The aim of this chapter was to obtain the relative pose of a stereo camera pair. This was an essential step towards obtaining the relative pose of a camera with respect to a racket.

Objectives:

1. To perform a stereo calibration using the 2D planar method.
2. To assess the repeatability of the pose parameters.
3. To assess the accuracy of the parameters by reconstructing the geometry of the calibration plane.
4. To provide a set of accurate extrinsic parameters for calculation of the relative pose of a camera with respect to a racket in chapter 5.

4.3 Stereo Imaging

The objective of a stereo calibration is to estimate the intrinsic and extrinsic (pose) parameters for a pair of cameras. Using these parameters, 3D position of a point in a scene, which is identified and matched in two stereo images can be determined by the triangulation method (Weng et al. 1992) (Figure 4.2). The corresponding location p_r of 2D coordinate p_l in Figure 4.2 is given in equation 2.10 (page 25).

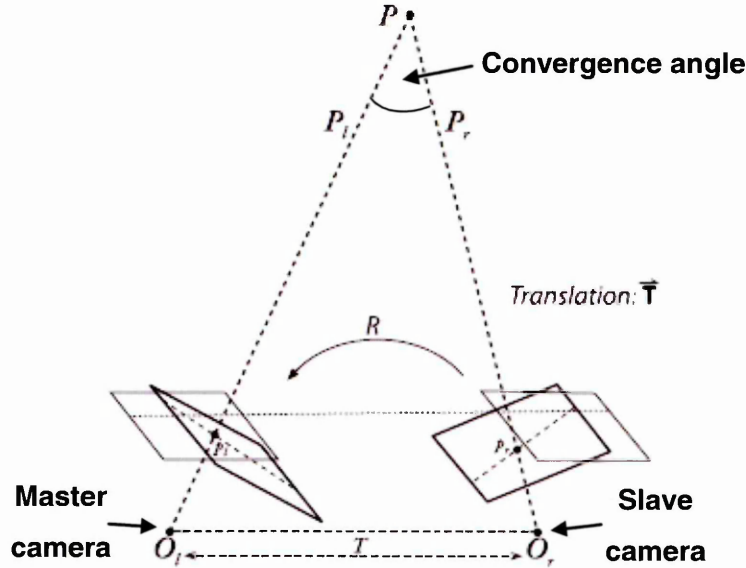


Figure 4.2 A stereo system. The 3D position of the point P can be determined by triangulation of the matched 2D coordinates p_l and p_r (adapted from Bradski and Kaehler 2008).

Bouguet's (2010) camera calibration toolbox for MATLAB relies on matching checkerboard images to estimate the relative pose of a slave camera with respect to a master camera (Figure 4.2). The intrinsic parameters for both cameras are also estimated from the matching images. Usually the intrinsic parameters are recomputed using the extrinsic information during a stereo calibration to improve the accuracy of the calibration; the calibration toolbox for MATLAB (Bouguet 2010) provides the user with this option. This chapter established whether or not intrinsic recomputation was necessary for the purpose of the methods developed and cameras used in this project.

4.4 Method

Two black and white high-speed Vision Research Phantom Miro M110 cameras were used for this study, with the same settings as outlined in chapter 3. The cameras were positioned with a convergence angle (Figure 4.2) of approximately 30° , to allow for the checkerboard to be in the view of both cameras and to achieve adequate image coverage.

The cameras were connected and synchronised so that an f-sync signal passed from the assigned master camera clock source to the slave camera. The slave camera was located on the right hand side with respect to the master camera (Figure 4.2). The Phantom camera control software (Phantom Help File, Vision Research, 2014) has a feature that allows multiple matched calibration images to be recorded. Calibration images were captured at a resolution of 1280 x 800 pixels, which was the maximum resolution available for both cameras.

Ten repeat stereo calibrations were performed using Bouguet's (2010) camera calibration toolbox for MATLAB. During the calibrations, both cameras remained untouched. The checkerboard was the same as that used in chapter 3; it had 8 squares in both directions each measuring 50 mm in the horizontal and vertical direction. Fifty images were taken for each calibration and the checkerboard was orientated in a range of different locations throughout the calibration volume. Effort was made to cover the entire field of view for both cameras.

Repeatability of the relative pose (rotation and translation) of the slave camera with respect to the master camera across the ten stereo calibrations was investigated. Calibration accuracy was assessed by using the parameters to reconstruct the geometry of the calibration grid (checkerboard). Three sets of calibration parameters were generated; those estimated using a complete distortion model (radial + tangential) without recomputed intrinsics, those estimated using a 4th order radial distortion model without the tangential component and without recomputed intrinsics and those estimated using a 4th order radial distortion model without tangential component but with recomputed intrinsics. For each set, ten reconstructions were performed.

Chapter 3 showed a 4th order radial distortion model without the tangential component provided more accurate results for the camera settings used. However, a complete distortion model is also included in this chapter for comparative purposes. It is important to assess the influence of the intrinsics on the relative pose of a stereo pair (chapter 3 used a single camera), for development of a view fitting method.

Finally, the sides of the squares (that make up the checkerboard) were measured from the reconstruction data. This was done by selecting adjacent reprojected checkerboard intersections from an image pair and calculating the difference. Reprojected distances were compared with the known horizontal and vertical size of each square. In this way the accuracy of the ten stereo calibrations and effectiveness of the point detection software is assessed.

4.5 Results

4.5.1 Image coverage

Figure 4.3 (a) and (b) show the typical image coverage achieved in one of the ten stereo calibrations in the field of view of the master and slave camera respectively. The majority of the field of view in both cameras was covered apart from the lower inside corners in both views. This was a result of a small area where the checkerboard was not visible in both camera views.

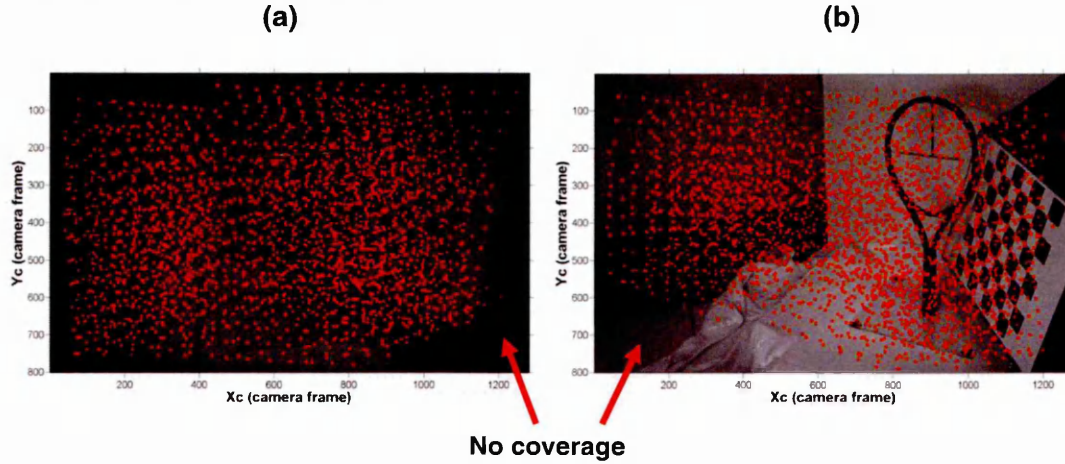


Figure 4.3 An example of the image coverage achieved for one of the ten stereo calibration repetitions in the field of view of (a) the master camera (b) the slave camera.

4.5.2 Intrinsic parameters

Based on findings from chapter 3, the intrinsic parameters for both cameras were estimated using a 4th order radial distortion model without a tangential component. Figure 4.4 (a) and (b) illustrate the radial distortion for the image coverage achieved in Figure 4.3 (a) and (b) respectively. In Figure 4.4 (a) and (b) maximum pixel displacement was 1.6 and 2.5 pixels respectively. These values were similar to that estimated for the single camera intrinsic calibration in chapter 3 (Figure 3.4 (a), page 47). Indicating the image coverage achieved in Figure 4.3 was sufficient. Table 4.1 presents the mean \pm standard deviation of the intrinsic parameters (f_x , f_y , c_x , c_y) for the master and slave camera over the ten calibrations. The standard deviation values of the intrinsic parameters for both cameras were low.

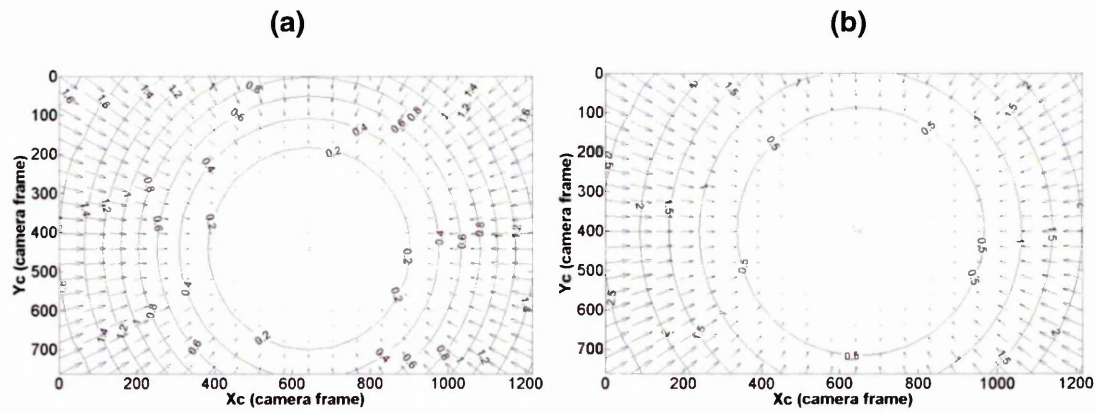


Figure 4.4 Typical pixel displacement due to radial distortion for (a) the master camera and (b) the slave camera for one of the ten stereo calibrations.

Table 4.1 Mean \pm standard deviation of the intrinsic parameters for the master and slave camera's over the ten stereo calibrations.

Intrinsics (pixels)	Master camera	Slave camera
f_x	1803.61 ± 0.95	1779.64 ± 1.65
f_y	1801.84 ± 1.14	1778.92 ± 1.48
c_x	642.35 ± 0.55	651.03 ± 0.39
c_y	440.31 ± 0.35	400.02 ± 0.56

4.5.3 Repeatability of the relative pose parameters

Figure 4.5 is a schematic of one of the ten completed stereo calibrations (including relative checkerboard positions), showing the relative position of the slave camera (right camera) with respect to the master camera (left camera). Note that the axes set of the master camera is centred at the origin and the reference frame has been changed for plotting purposes.

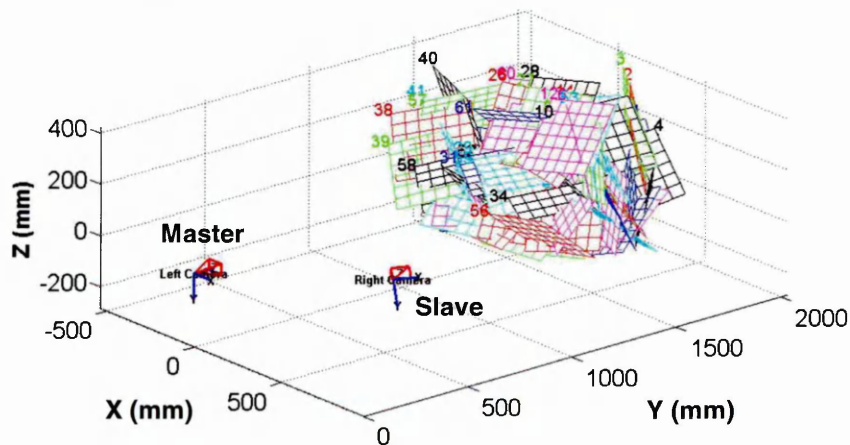


Figure 4.5 Schematic of one of the ten completed stereo calibrations , showing the relative position of the slave with respect to the master camera and the checkerboards

Figure 4.6 (a), (b) and (c) show mean translation in the X, Y and Z directions respectively for the position of the slave with respect to the master camera, over the ten stereo calibrations. Poses are shown for the three intrinsic distortion models. Other than in the X direction, translation had a lower interquartile range when the 4th order radial distortion model without intrinsic recomputation was used for the stereo calibration. Mean camera translation differed by 0.07, 0.37 and 0.26 % in the X, Y and Z direction respectively when the intrinsics were recomputed.

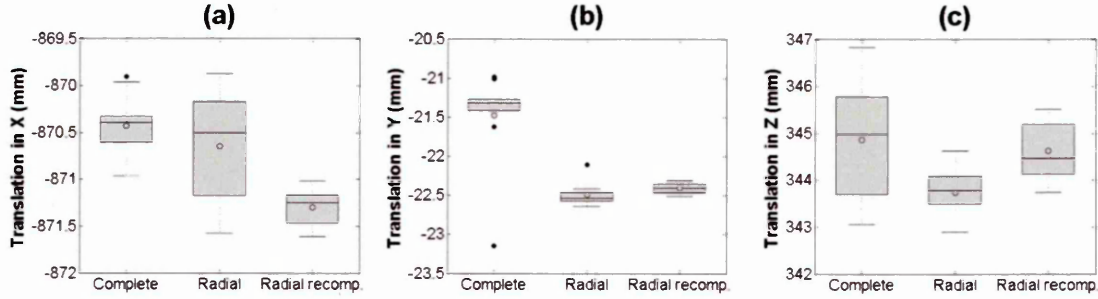


Figure 4.6 Translation in the (a) X, (b) Y and (c) Z direction of the slave camera with respect to master camera estimated using the three intrinsic models, showing the mean (circle), median, upper and lower quartiles and upper and lower deciles.

Figure 4.7 (a), (b) and (c) show mean rotation angle about the axis of X, Y and Z respectively for the orientation of the slave with respect to the master camera. The range in orientation across the 10 stereo calibrations was less when the 4th order radial distortion model with recomputed intrinsics was used. Mean rotation angle differed by 0.45% about the X axis and by less than 0.01% about the Y and Z axes when the intrinsics were recomputed. Removal of the tangential component of distortion significantly influenced camera rotation ($p < .05$) however, recomputation of the 4th order radial distortion model did not significantly influence camera rotation ($p > .05$).

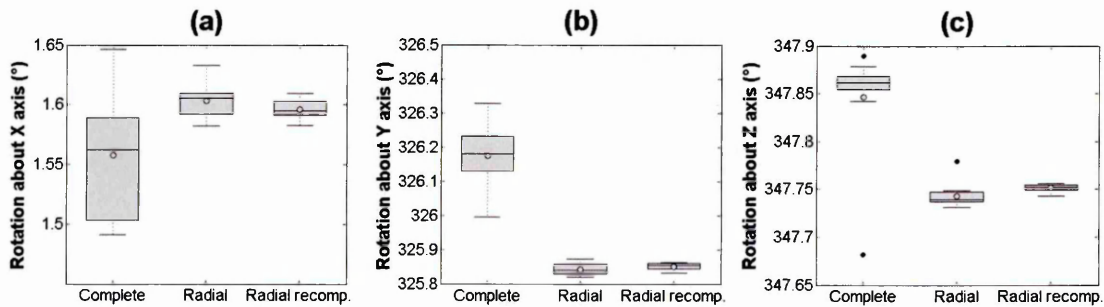


Figure 4.7 Rotation about the axis' of (a) X, (b) Y and (c) Z of the slave camera with respect to master camera calculated using the three intrinsic models, showing the mean (circle), median, upper and lower quartiles and upper and lower deciles.

4.5.4 Accuracy

Mean reconstruction error \pm standard deviation for the forty nine intersections of the checkerboard (red dots) in the X, Y and Z directions is shown in Figure 4.8. The reconstructions were performed using pose parameters from each of the ten stereo calibrations obtained using the three intrinsic models. Mean reconstruction error was marginally lower in each direction when a 4th order radial distortion model without recomputation was used for the calibrations. Reconstruction error in the Z direction represented the largest component of error; using the 4th order radial distortion model without recomputed intrinsics, it was 0.13 ± 0.07 mm. For the same intrinsic model, reconstruction error was less than 0.1 in the X and Y directions. The complete intrinsic model produced the largest reconstruction error of 0.15 ± 0.07 mm, in the Z direction. Figure 4.8 shows that recomputation of the intrinsics did not improve reconstruction accuracy. Lower standard deviation values across the ten calibrations are observed in the X and Y directions but not in the Z direction, when the intrinsics were not recomputed.

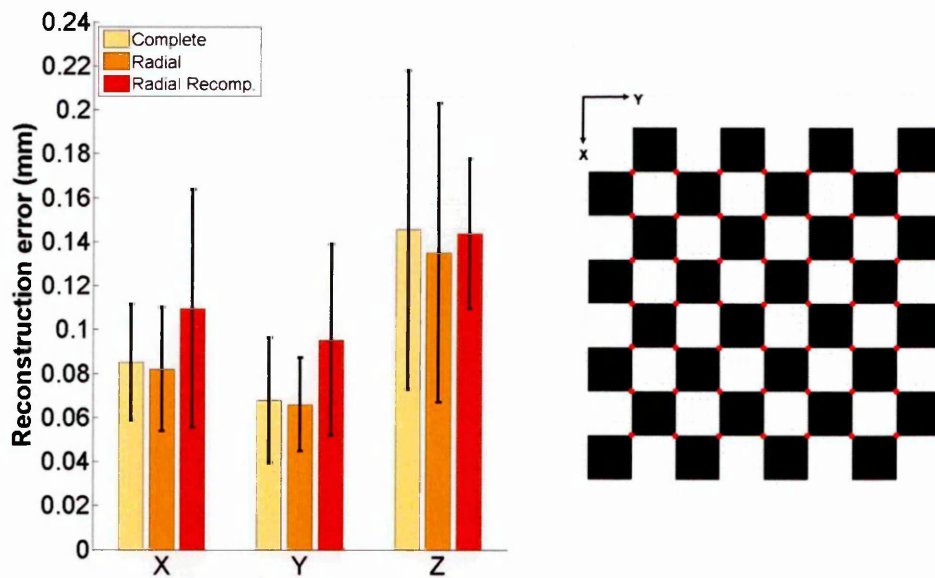


Figure 4.8 Mean reconstruction error \pm standard deviation for the geometry of the checkerboard in the X, Y and Z directions using a complete and 4th order radial distortion model.

Figure 4.9 shows the mean \pm standard deviation of the checkerboard square sizes over the ten stereo calibrations as calculated by the software. Square size was computed from the results in Figure 4.8 using the three intrinsic models. The mean is very close to the true value of 50 mm for all the intrinsic models. The complete model is the least consistent; it provides the least and most accurate result in the horizontal and vertical directions respectively. The 4th order radial distortion model provides very similar

results whether the intrinsics are recomputed or not. However, the standard deviation was lower when the intrinsics were recomputed. Intrinsic parameter recomputation did not significantly improve ($p > 0.05$) the accuracy of the checkerboard square size, as calculated by the software.

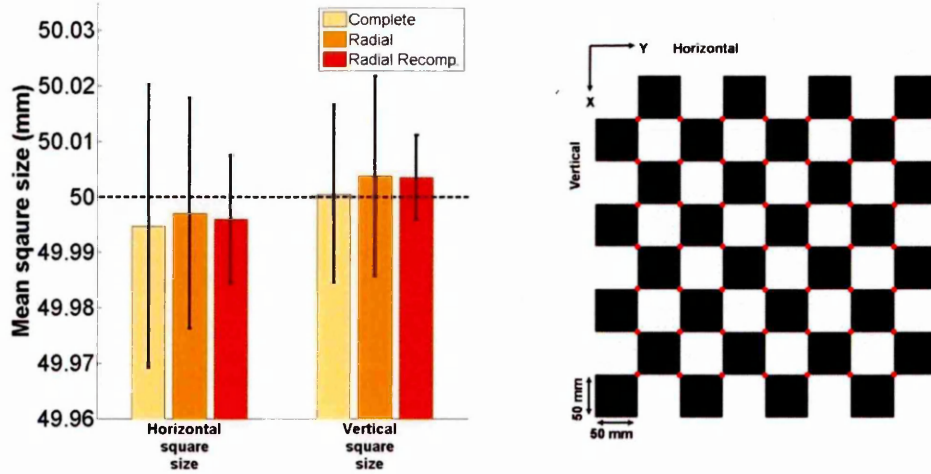


Figure 4.9 Mean \pm standard deviation checkerboard square size (horizontal and vertical), measured using the MATLAB camera calibration software for the three intrinsic models. The black dashed line indicates the known square size of 50 mm.

4.6 Discussion

The relative pose of a slave camera with respect to a master camera has been calculated. Stereo calibrations were performed using the 2D planar checkerboard method developed by Zhang (1999). This was a preliminary step for obtaining the pose of a camera relative to a racket in chapter 5.

4.6.1 Methodological considerations

In addition to the methodological considerations identified in chapter 3, the convergence angle of the cameras should be accounted for, in order to ensure the accuracy of a stereo calibration. Care should be taken to position the cameras so that maximum possible image coverage can be achieved in both views. This can be ensured by positioning the cameras so that the optical axes intersect in front of the checkerboard. This is based purely on human judgement and ensures a larger portion of both cameras field of view contain the checkerboard. The quality of a calibration improves with image coverage (Zhang 1999, Kelley 2014).

4.6.2 Repeatability

Table 4.1 presents the repeatability of the intrinsic parameters for both the master and slave camera. All the parameters, for both cameras, had low standard deviation values across the ten calibrations. In general, focal length (f_x and f_y) was estimated to within

± 1.5 pixels and principal point (c_x and c_y) to within ± 0.5 of a pixel over the ten stereo calibrations. Figure 4.3 showed the typical image coverage achieved for one of the ten stereo calibration repetitions. Obtaining coverage in the lower inside corners of both images was not possible since the checkerboard was not visible. However, the repeatability of the intrinsics in this chapter were in the range of that obtained in chapter 3 for a single camera, for which total image coverage was achieved. This suggests the image coverage achieved in Figure 4.3 was adequate to provide repeatable intrinsic parameters.

Figure 4.6 and Figure 4.7 showed that the relative pose (rotation and translation) of a slave camera with respect to a master camera was more repeatable if a 4th order radial distortion model without the tangential component was used for a stereo calibration. This was not surprising since chapter 3 showed that a 4th order radial distortion model without the tangential component was more repeatable compared to a complete distortion model, for 2D planar calibration. Furthermore, recomputation of the intrinsics for a 4th order radial model had a marginal influence on the relative pose of the cameras. This shows that for the camera and lens settings used:

- The repeatability of the intrinsics translates to the extrinsics for a stereo pair.
- Exclusion of the tangential component from the distortion model does provides a more repeatable stereo calibration.
- Intrinsic recomputation has a marginal influence on the repeatability of the intrinsics when a 4th order radial distortion model is used.

4.6.3 Accuracy

Figure 4.8 demonstrated that a stereo calibration performed using a 4th order radial distortion model without the tangential component and without recomputed intrinsics produces the most accurate reconstruction of the calibration plane. As mentioned previously, this finding is specific to the camera and lens settings used. However, these settings will be used throughout this project and are therefore essential in order to obtain the best results for a view fitting method to estimate 3D racket position.

Inclusion of the tangential component in the distortion model had a negative influence on the accuracy of the stereo system. Zhang (1999) justifies removal of the tangential component on the basis that most modern lenses are manufactured with no imperfections. Lens distortion factors have previously been applied in stereo calibrations applied in sport to reduce calibration error. Kelley (2014) simulated a 4th order radial distortion model without the tangential component in order to stereo

calibrate a volume within a hammer cage. Choppin et al. (2007a) and Driscoll et al. (2014) state that distortions were accounted for in order to calibrate stereo systems in a tennis and football environment respectively, but don't provide the exact sub-set of radial and tangential coefficients used.

Figure 4.8 also shows that using a 4th order radial distortion model without the tangential component and without the intrinsics recomputed is likely to produce lower reconstruction errors compared to if the intrinsics were recomputed. Figure 4.9 shows that estimated checkerboard square size was very similar whether the intrinsics were recomputed or not; suggesting recomputation does not improve the accuracy of the parameters for the lens and camera settings used. This is in contrast to Kelley (2014) who found intrinsic recomputation provided more accurate reconstructions using a standard stereo calibration method i.e. left and right checkerboard images captured simultaneously. However, the scenario simulated by Kelley (2014) was sub-optimal for a standard stereo calibration; the purpose of the work was to develop a non-standard method for calibration of an occluded hammer cage. The current study has shown that with optimal environmental conditions and by taking into account a number of methodological considerations (when using the 2D planar calibration method) intrinsic recomputation is not necessary to improve the accuracy of a stereo calibration.

4.7 Conclusion

For the proposed calibration method, a 4th order radial distortion model without the tangential component and without intrinsic recomputation should be used. The reconstruction using this intrinsic model was the most accurate and repeatable. The proposed calibration method is suitable to be adapted for creation of an accurate fully calibrated set, required for the development of a novel view fitting method to estimate 3D racket position. The next step in the development of a view fitting method is to use the information gathered in this chapter to obtain the accurate relative pose of a camera with respect to a racket in order to create a fully calibrated set.

5 Obtaining the pose of a camera relative to a racket

5.1 Introduction

A prerequisite for markerless motion capture (MMC) methods is to determine the internal and external camera parameters and a common local reference frame through a calibration procedure (Mündermann, Corazza and Andriacchi 2006). Methods to obtain accurate internal and external camera parameters were presented in chapter 3 and chapter 4 respectively. This chapter uses the parameters to create a common local reference frame; a method is developed to obtain the relative pose of a camera with respect to a racket (Figure 5.1). Previously, this has never been done for the purpose of a view fitting method. The error associated with the pose of a camera with respect to a racket is calculated.

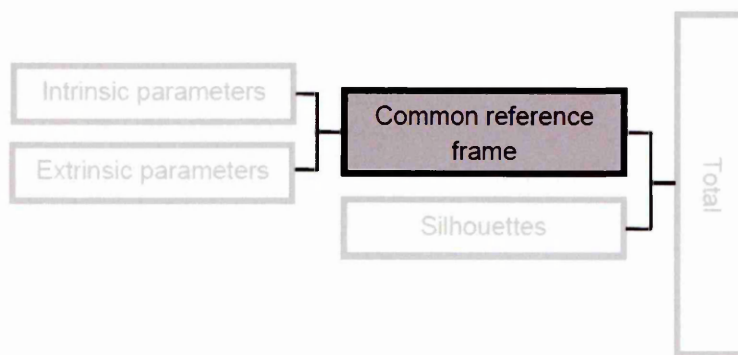


Figure 5.1 Chapter 5 placed a camera and a racket in common local reference frame and quantified the associated error of the cameras pose.

5.2 Aim and objectives

The aim of this chapter was to accurately obtain the relative pose of a camera with respect to a racket.

Objectives:

1. To construct a rig on which the racket can be mounted and to define a local reference frame.
2. To ensure the racket is accurately mounted on the rig.
3. To assess the variability of the relative pose due to control point error.
4. To assess the accuracy of the relative pose.

5.3 Method

The first step in obtaining the pose of a camera relative to a stationary tennis racket was to build a rig on which the racket could be mounted. The two main requirements of the rig were:

1. To contrast against the colour of the racket to aid silhouette extraction.
2. To provide a common local reference frame for the racket and camera.

The rig was constructed from wood, clear Perspex and a white sheet so that it contrasted highly with the racket (Prince Warrior 100L ESP) which was standard size (686 x 274 x 25 mm) and painted matte black. This racket was used throughout the project. A backlight was placed behind the rig to assist silhouette extraction. To prepare a site for the local reference frame, the racket was mounted upright on a clear Perspex board (33 x 40 x 4 mm) (Figure 5.2). Black paint was dropped into four grooves that had been machined into the board to form a *lower* calibration plane (P_1 to P_4 Figure 5.2).

A metal rod extended through a hole in the lower calibration plane and screwed into the racket butt to provide support. The racket was not strung because it is only the silhouette outline that is of interest therefore the empty stringbed provided space for a second, *upper* calibration plane to be defined. Two metal rods of 4 mm diameter were fixed perpendicular to one another through the grommet holes of the racket. The grommet holes had a diameter marginally larger than 4 mm. Three paint markers were applied to the rods at known locations so the racket face could be defined as an upper calibration plane (P_5 to P_7 Figure 5.2). In Figure 5.2 the vertical rod is off-centre because a racket does not have a grommet hole that is aligned with the long axis. Choppin et al. (2007a) applied markers to the frame itself in order to define the racket face as a plane however, this would impede silhouette extraction.

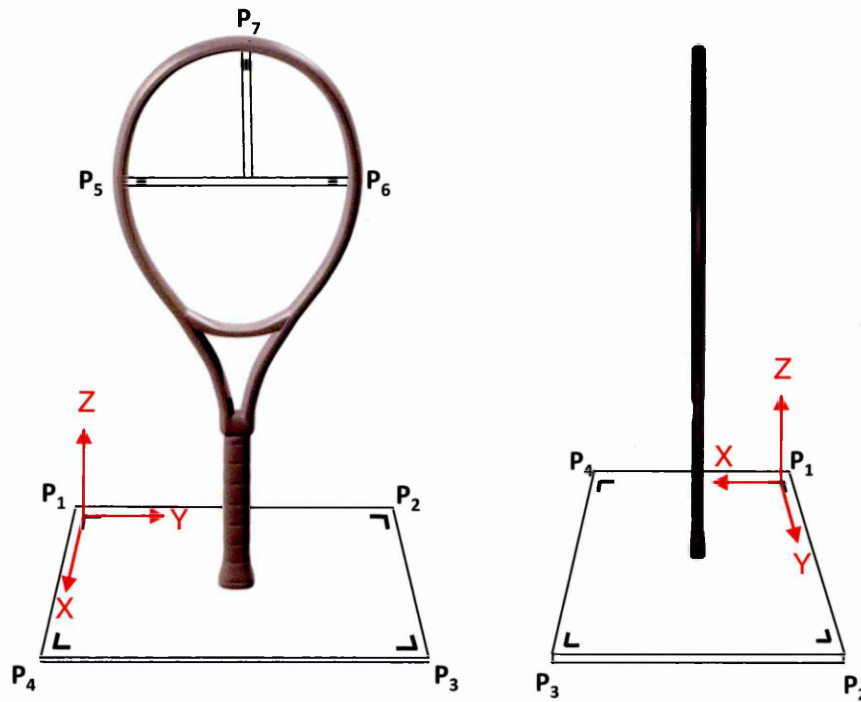


Figure 5.2 The lower calibration plane was defined by coordinates P_1 to P_4 and P_5 to P_7 defined the upper calibration plane which aligned with the racket face plane.

Bouguet's (2010) camera calibration toolbox for MATLAB also includes an option to use non-planar orthogonal calibration objects. Therefore, it was important that the upper plane (P_5 to P_7 Figure 5.2) was perpendicular to the lower plane (P_1 to P_4 Figure 5.2) to create an orthogonal 3D calibration object. This involved measuring the angle offset at the tip of the racket relative to the base of the handle. Consider two planes intersecting the racket long axis parallel and perpendicular to the racket face plane (Figure 5.3 (a)). The first angle to be measured was the offset of the parallel plane (XZ plane) with respect to the global Z-axis (Figure 5.3 (b)). The second angle to be measured was the offset of the perpendicular plane (YZ plane) with respect to the global Z-axis (Figure 5.3 (c)).

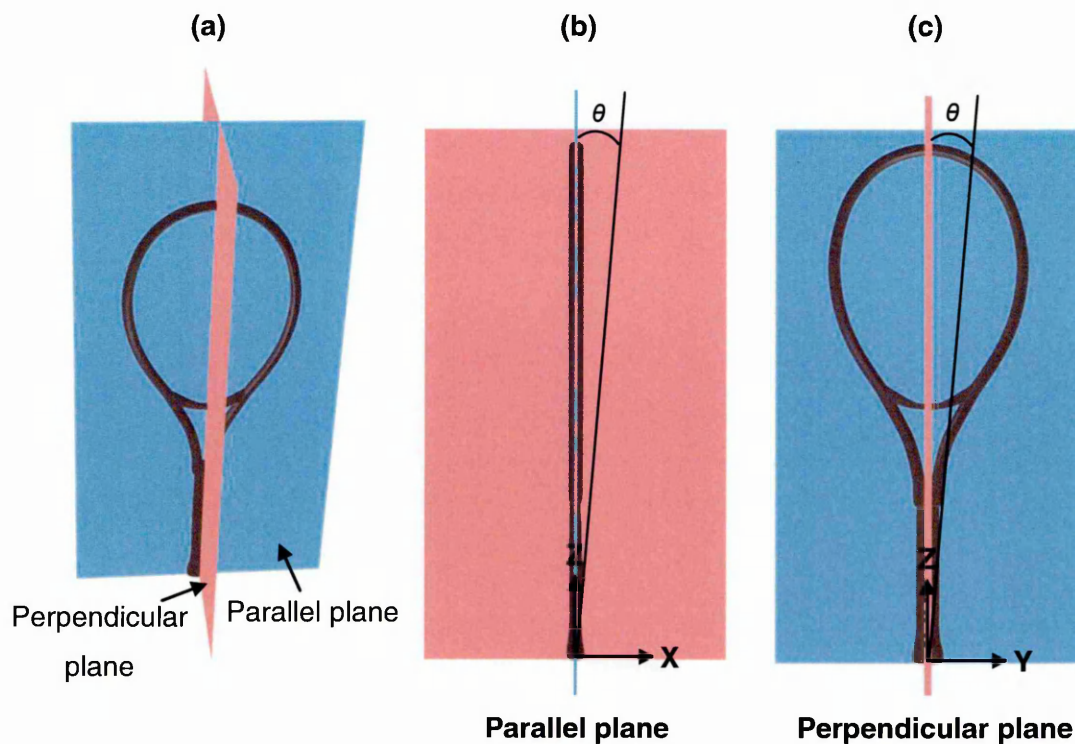


Figure 5.3 Illustration of the angles measured showing (a) the intersecting planes, the angle offsets of the planes (b) parallel and (c) perpendicular with the racket face plane.

Firstly, an inclinometer was used to manually measure the angle of the planes parallel and perpendicular to the racket face plane with respect to the global Z-axis (Figure 5.3). Racket position was modified until the inclinometer showed an angle of 0° . Modification involved placing small sections of material under the racket butt. This method was useful for physically positioning the racket however, the angles measured were unlikely to be of adequate accuracy. Since readings were taken from a limited number of positions around the racket frame and in some cases the inclinometer had to be physically supported.

To obtain more accurate measurements, a laser scan of the racket and the upper and lower calibration planes was performed (Figure 5.4). A colleague assisted in performing the scan and measuring the angles from the data.

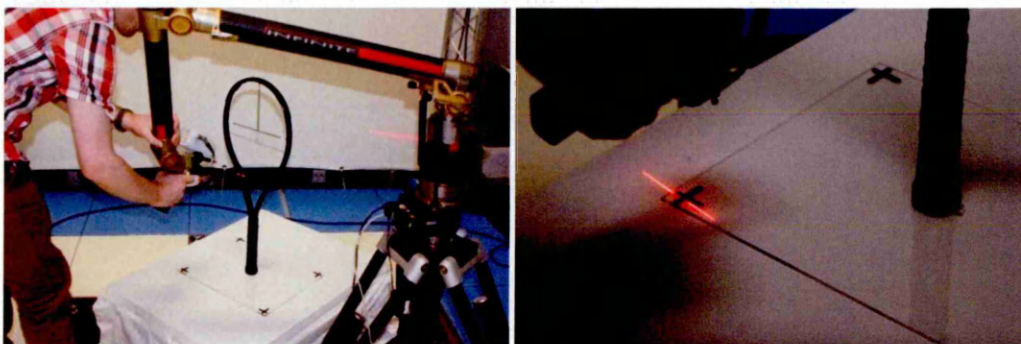


Figure 5.4 Laser scanning of the racket and calibration planes.

The scanner provides an accuracy to within ± 0.001 mm. Figure 5.5 shows the rendered scan. To determine measurements from the scan data, the global origin was centred at the racket butt and the lower calibration plane was made perpendicular to the global Z-axis using the software (Geomagic Studio 2014). The long axis of the scan was intersected by the planes perpendicular (Plane XZ) and parallel (Plane YZ) to the racket face plane (Figure 5.5). The offset of these two planes with respect to the global Z-axis was then measured as shown in Figure 5.3.

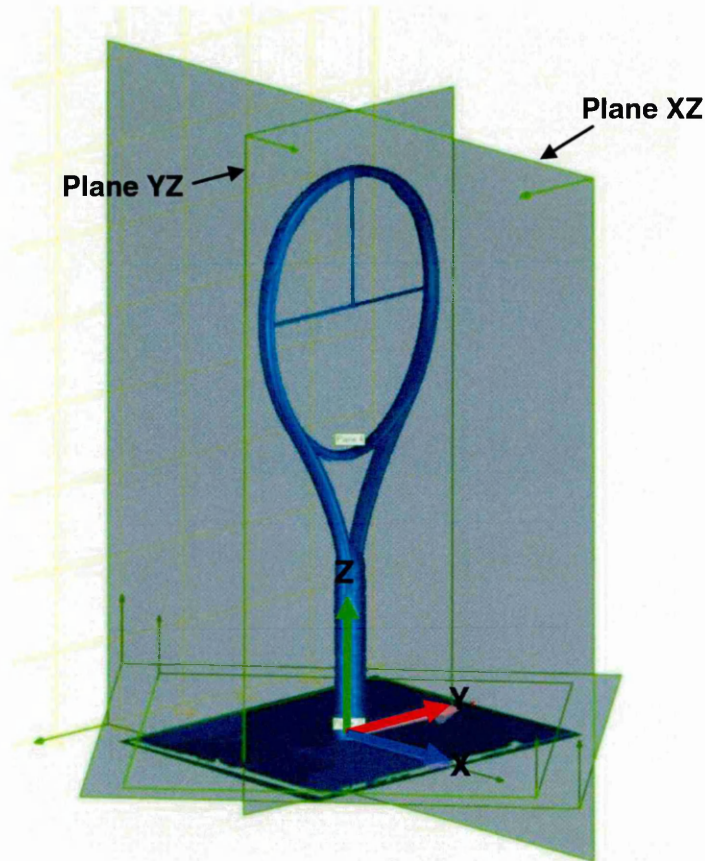


Figure 5.5 Rendered laser scan of the racket and calibration planes. The long axis of the racket scan was intersected by two virtual perpendicular planes used to measure the angles (Plane YZ and Plane XZ).

The offset of Plane YZ and Plane XZ (Figure 5.5) was 0.2 and 0.3° respectively. These angles were low enough to be deemed negligible, indicating the structure was orthogonal. It was not possible to measure the rotation of the planes about the global Z-axis however, the scan was used to measure the relative location of the control points. Mean location of the X coordinate of the control points P_5 to P_7 (Figure 5.2) in the Plane YZ was 0.3 ± 0.5 mm, confirming the racket face plane was aligned straight. Based on these measurements, this experimental setup was deemed appropriate to be used to obtain the relative pose of a camera with respect to a racket.

The relative position of the seven control points (P_1 to P_7 Figure 5.2), with respect to control point P_1 were also measured from the laser scan (Table 5.1). This involved manually selecting the control points on the surface of the scan, which was facilitated by the software (Geomagic Studio 2014), which allowed the user to zoom-in without the scan losing visual quality.

Table 5.1 Relative position of the seven control points as measured from the laser scan. The control point P_1 was centred at the origin creating a left-handed reference frame.

	P_1	P_2	P_3	P_4	P_5	P_6	P_7
X (mm)	0.00	0.00	292.62	291.85	147.85	146.73	147.61
Y (mm)	0.00	360.75	361.64	0.00	69.93	289.93	185.06
Z (mm)	0.00	0.00	0.00	0.00	518.94	518.94	661.39

5.4 Relative pose of the master camera to the racket

Given the relative pose of a slave camera with respect to a master camera obtained through stereo calibration in chapter 4 (R_{stereo} and T_{stereo} in Figure 5.6). The cameras were placed in a common local reference frame containing the racket. This was achieved using the orthogonal calibration object outlined in the previous section. i.e. a combination of the upper and lower calibration planes (P_1 to P_7 in Figure 5.6). The control point P_1 was selected to be centred at the local origin creating a left-handed reference frame. The distance of P_1 from the base of the racket handle was 233 mm.

The common local reference frame between the racket and the master camera (R_{local} and T_{local} in Figure 5.6) was obtained by manually digitising the seven control points (P_1 to P_7 Figure 5.6) in the image plane of the master camera (Figure 5.7). A MATLAB function was written that allowed the user to zoom-in so that the seven control points could be selected with greater accuracy. Each of the seven control points was digitised ten times and the mean taken. To avoid a learning effect, short breaks were taken between each of the ten digitisation stints and the points were digitised in the same order (P_1 to P_7).

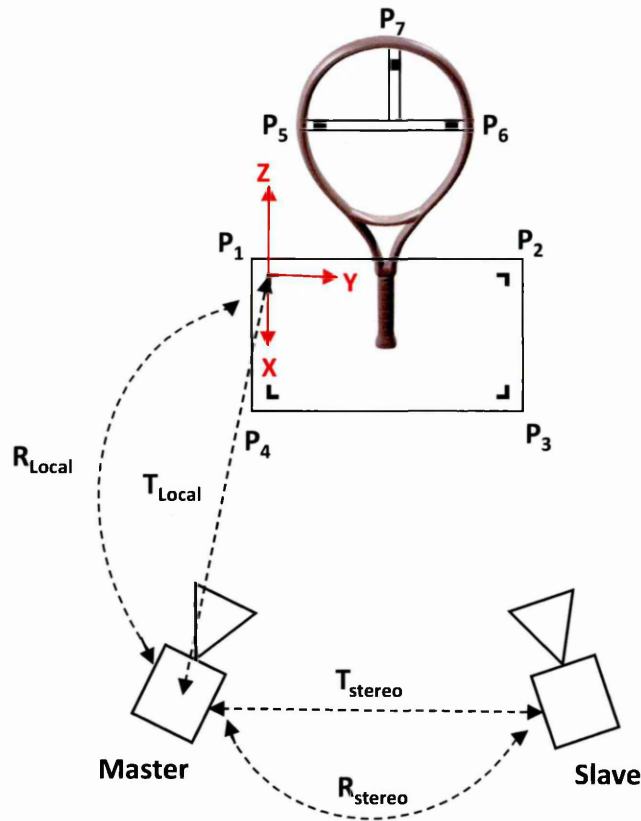


Figure 5.6 The cameras and the racket were defined in a common local reference frame using the seven orthogonal control points, P_1 to P_7 .

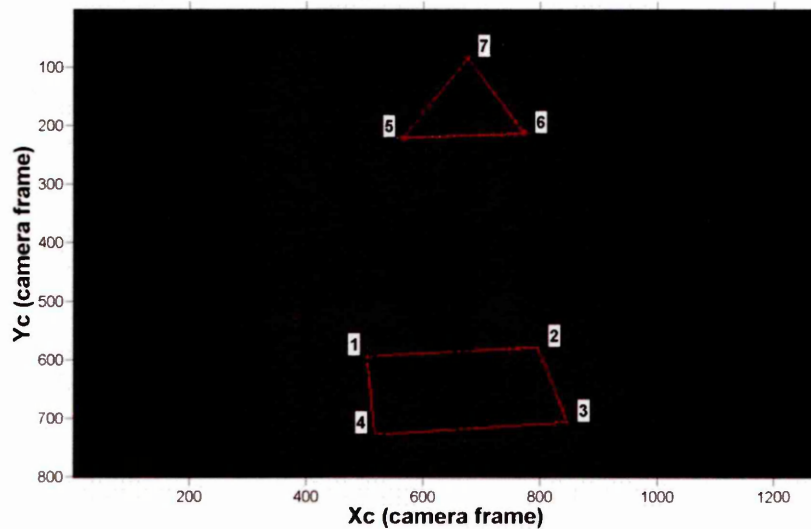


Figure 5.7 Seven manually digitised 2D image coordinates in the view of the master camera.

The seven digitised pixel coordinates and their corresponding world dimensions measured from the laser scan, were then passed to the MATLAB camera calibration toolbox (Bouquet 2010). The toolbox uses this information along with the cameras

intrinsic parameters to compute the relative pose of the camera with respect to the racket (R_{local} and T_{local} in Figure 5.6).

5.5 Accuracy

5.5.1 Control point error

This was the error in the relative pose of the master camera in the common reference frame caused by manual digitisation of the seven control points. The uncertainty in manual digitisation was determined by digitising the seven control points ten times in the image plane of the master camera (Figure 5.7). The mean standard deviation over the ten repeats in the u and v image axes is given in (Table 5.2). The overall error due to manual digitisation was concluded to be ± 0.1 pixels.

Table 5.2 Assessment of error in manual digitisation of the seven control points.

	U-coordinate	V-coordinate
Mean standard deviation	0.09	0.12

The influence of the ± 0.1 pixel uncertainty on the relative pose of the master camera was then quantified. This was done by projecting the 3D control points in to the 2D image plane of the master camera using the calibration parameters obtained in section 5.5. Projecting the points removed noise resulting from calibration parameter error. Random Gaussian noise that had a mean of zero and a standard deviation of 0.1 pixel was then added to the projected pixel coordinates. One thousand independent trials were made: each trial involved adding the random Gaussian noise to the projected pixel coordinates which were used to compute the relative pose of the master camera. The standard deviation for the relative pose (rotation and translation) of the master camera are displayed in table 5.3. The ± 0.1 pixel error in manual digitisation of the seven control points propagated to a mean standard deviation of ± 0.61 mm for translation and $\pm 0.02^\circ$ for rotation of the master camera.

Table 5.3 Assessment of standard deviation in camera pose resulting from manual digitisation.

Camera pose parameters	\pm standard deviation
X (mm)	0.56
Y (mm)	0.65
Z (mm)	0.63
X (degrees)	0.02
Y (degrees)	0.02
Z (degrees)	0.01

5.5.2 Relative camera pose error

The relative pose of the slave camera with respect to the master camera (R_{stereo} and T_{stereo} in Figure 5.6) was obtained in chapter 4 through a stereo calibration. In this section, the error associated with R_{local} and T_{local} was quantified against R_{stereo} and T_{stereo} in Figure 5.6, which was treated as the ground truth. This involved digitising the seven control points in the view of both the master and the slave camera. Mean standard deviation in control point digitisation across ten repeats was 0.08 and 0.07 of a pixel in the view of the master and slave camera respectively. The digitised control points and corresponding world dimensions were then passed to the calibration toolbox for MATLAB (Bouquet 2010) in order to obtain the relative pose with respect to the racket. To test the influence of the number of control points, three relative poses for each camera were calculated using:

- Seven control points (P_1 to P_7 in Figure 5.2).
- Lower plane (P_1 to P_4 in Figure 5.2).
- Upper plane (P_5 to P_7 in Figure 5.2).

Equation 2.10 (page 25) was then used to calculate the relative pose between the master and slave camera. Figure 5.8 shows how well these relative poses matched up with the relative ground truth pose. Camera pose accuracy increased with the number of control points. This indicates that using the seven control points in the form of a non-planar calibration object provided superior accuracy compared to using planar calibration objects comprised of fewer control points.

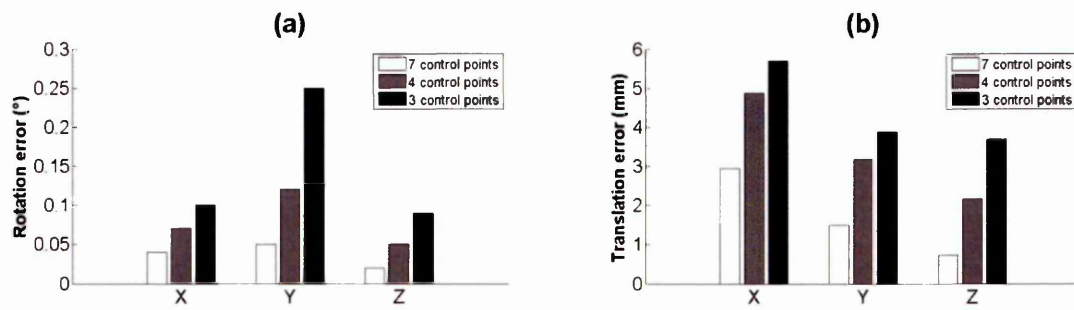


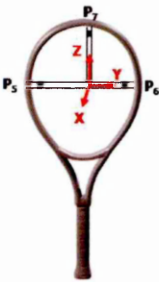
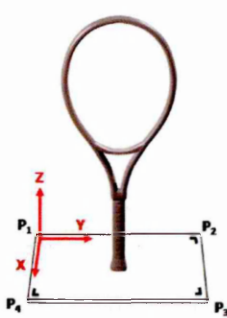
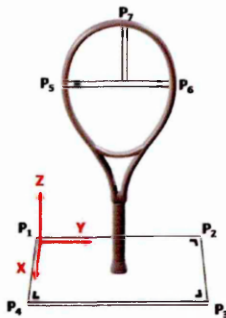
Figure 5.8 Bar graphs showing error in relative pose of the slave with respect to the master camera obtained through manual digitisation of 7, 4 and 3 control points.

5.5.3 Reconstruction error

The relative poses in Figure 5.8 obtained using 3, 5 and 7 control points were used to reconstruct the 3D location of the respective control points through stereo triangulation. The triangulated points were compared with the ground truth as measured from the laser scan (Table 5.1). Table 5.4 shows that RMSE for reconstruction error in all directions decreased as the number of control points increased.

Table 5.4 RMSE for reconstruction in the X, Y and Z directions using pose parameters obtained using 3, 4 and 7 control points.

Reconstruction error direction	Number of control points		
	3	4	7
X (mm)	2.98	2.62	0.75
Y (mm)	0.63	0.46	0.33
Z (mm)	1.56	1.49	0.61

5.6 Repeatability

5.6.1 Repeatability of the slave camera pose

Since the relative pose of the slave camera with respect to the master camera (R_{stereo} and T_{stereo} in Figure 5.6) was already known from chapter 4. There is actually no need

to digitise the seven control points in the view of the slave camera in order to place it in the common local reference frame. Given the relative pose of the master camera with respect to the racket (obtained through digitisation of the seven control points), the pose parameters (R_{stereo} and T_{stereo} in Figure 5.6) can be applied to equation 2.10 (page 25). This places the slave camera in the common local reference frame with the master camera and the racket. The advantage of this method is that digitisation of the seven control points only needs to be performed once in the image plane of the master camera, thus removing the influence of manual digitisation. This method of obtaining the pose of the slave camera will be referred to as the *stereo method*.

The repeatability of obtaining the relative pose of the slave camera in the common local reference frame was investigated using the stereo method and the *2D method*, which involved digitising the seven control points. For the 2D method, the seven control points had already been digitised ten times (with a variance of 0.7 of a pixel, see section 5.5.2) in the image plane of the slave camera. Corresponding image and world coordinates were then used to define the relative pose of the slave camera with respect to the racket.

Table 5.5 presents the repeatability in the relative pose of the slave camera obtained using the two methods. Rotation and translation are represented in Euler angles and millimetres respectively. Table 5.5 demonstrates that the relative pose of the slave camera is more repeatable when obtained using the stereo method compared with the 2D method, by virtue of smaller standard deviation values. The largest difference in repeatability of translation is in the local X-axis where the range for the standard and stereo method is 6.00 and 1.76 mm respectively. This was to be expected as it is the axis heading away from the camera and is the most difficult to obtain in a 3D system (Driscoll 2012). The range values for rotation about the axis of X, Y and Z were under 0.1 and 0.2° for the stereo and standard method respectively.

Table 5.5 Repeatability of the slave camera pose obtained using the stereo and standard method of calibration.

	Standard 2D method	Stereo method
	$\pm s$ [Range]	$\pm s$ [Range]
X (mm)	1.94 [6.00]	0.48 [1.76]
Y (mm)	0.92 [2.72]	0.64 [1.75]
Z (mm)	0.83 [2.47]	0.22 [0.65]
X (degrees)	0.04 [0.13]	0.02 [0.05]
Y (degrees)	0.06 [0.19]	0.03 [0.09]
Z (degrees)	0.02 [0.08]	0.01 [0.05]

5.7 Summary

This chapter presented a method for accurately obtaining the pose of a camera with respect to a racket. This was an important prerequisite for the development of a view fitting method to estimate racket position in 3D. This chapter showed:

- The rig was suitable in order to obtain the relative pose of a camera with respect to a racket. A laser scan of the racket revealed negligible angle offsets of 0.2 and 0.3° at the racket tip in the planes parallel and perpendicular to the racket face plane respectively, with respect to the global Z-axis. Following these measurements the racket remained untouched.
- Digitisation of seven control points on a non-planar calibration object provided the relative pose of the master camera with respect to the racket.
- The standard deviation due to manual digitisation of the control points on the image plane was ± 0.1 pixels. This translated to a relative camera pose error of 0.61 mm for translation and $\pm 0.02^\circ$ for rotation.
- Using seven control points compared to four increased the accuracy of the rotation and translation component of relative camera pose by 54 and 45% respectively. While using seven control points compared to three increased the accuracy of the rotation and translation component of relative camera pose by 76 and 56% respectively.

- Reconstruction of the geometry of the orthogonal 3D calibration object was more accurate when 7 control points were used compared to 4 and 3 of the control points.
- Relative camera pose with respect to the racket was more repeatable when obtained using the stereo method compared with the 2D method. The stereo method is advantageous because it relies on the control points being digitised in the view of the master camera only once. The pose parameters (R_{stereo} and T_{stereo} in Figure 5.6) obtained in chapter 4 are derived from matching checkerboard images containing multiple control points which are automatically detected to a high level of accuracy using the MATLAB calibration toolbox (Bouquet 2010).

5.7.1 Limitations

The method presented in this chapter obtained the relative pose of camera with respect to a racket to an acceptable level of accuracy. However, it could be improved by using a calibration object that occupies more image space. Zhang (1999) and Choppin (2008) recommend making the calibration object as large in the image as possible. This could improve calibration accuracy since control point detection would be made easier.

5.8 Conclusion

By implementing the camera calibration techniques and recommendations outlined in chapter 3 and 4, the method presented in this chapter could be used to capture more views of the racket in order to create a fully calibrated set. Since the intrinsic and extrinsic parameters and a common local reference frame have now been defined, the next step is to apply the epipolar tangency constraint to a fully calibrated set of camera poses and associated racket silhouettes.

6 Measuring the consistency of racket silhouettes

6.1 Introduction

There will always be some error associated with a camera's relative pose; arising from the calibration parameters and manual digitisation of the control points. The literature review identified the epipolar tangency constraint, which can be applied to a pair of fully calibrated cameras and associated silhouettes in order to measure their inconsistency (Wong 2001). In this chapter, synthetic camera poses and silhouettes associated with a 3D racket model were generated. This allowed the epipolar tangency constraint to be applied to a fully calibrated silhouette set associated with the racket model. The influence of discrete pixel rendering and extraction of silhouettes on the epipolar tangency error (ETE) was quantified (Figure 6.1).

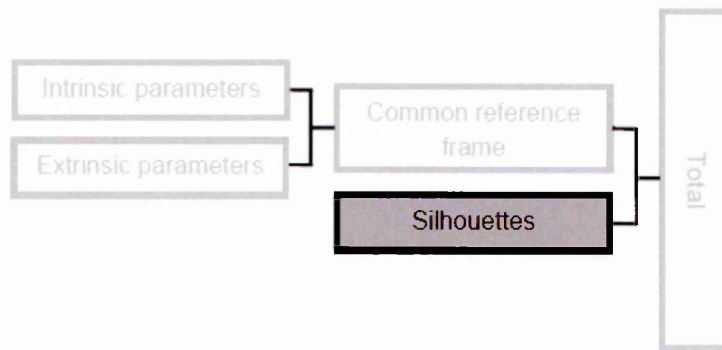


Figure 6.1 Chapter 6 used synthetic data to quantify the influence of discrete pixel rendering and extraction of silhouettes on the epipolar tangency error (ETE).

6.2 Aims and objectives

The aim of this chapter was to apply the epipolar tangency constraint to a synthetic fully calibrated set and quantify the influence of discrete pixel rendering and silhouette extraction on the ETE.

Objectives:

1. To apply the epipolar tangency constraint to a pair of racket silhouettes.
2. To generate synthetic camera poses and silhouettes associated with a 3D racket model.
3. To use the synthetic camera poses to quantify the influence of discrete pixel rendering and silhouette extraction on the ETE.

6.3 The epipolar tangency constraint for a pair of racket silhouettes

Figure 6.2 illustrates the epipolar tangency constraint applied to a pair of racket silhouettes. A full explanation of the epipolar tangency constraint is in section 2.3.6 (page 24-26) of the literature review and in Bradski and Kaehler (2008).

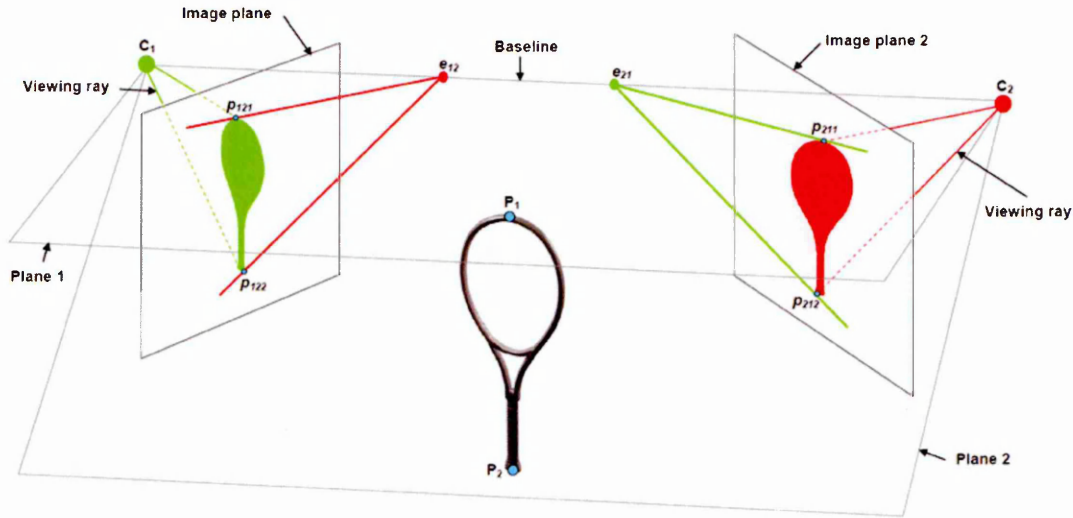


Figure 6.2 The epipolar tangency constraint for a pair of fully stereo calibrated racket silhouettes.

If there is no error in the camera parameters or silhouette extraction then the viewing ray associated with a tangency point must project exactly through the tangency point in the corresponding image (Figure 6.2). In other words, a line between an epipole e_{ijk} and a tangency point p_{ijk} is the same as the line $F_{ji}^* p_{ijk}$ (Figure 6.3), yielding the epipolar constraint in equation 2.14 (see section 2.3.6, page 26).

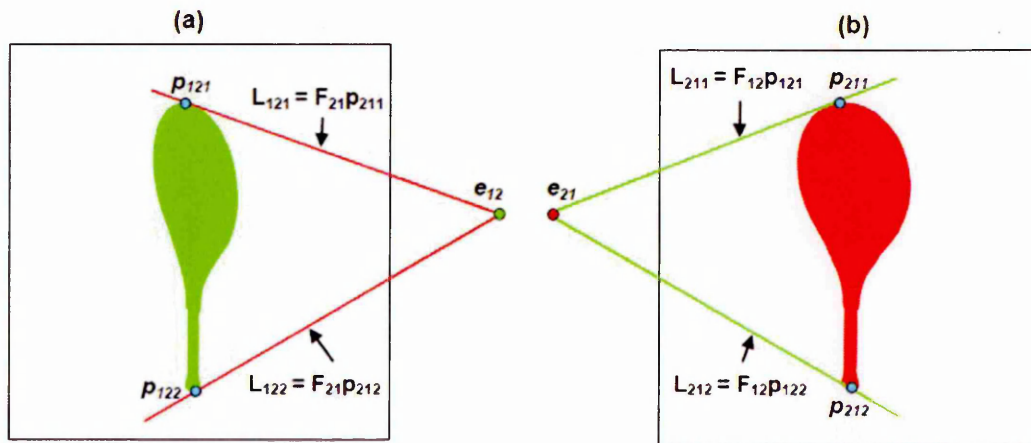


Figure 6.3 A tangency point p_{ijk} is constrained to lie on the line $F_{ji}^* p_{ijk}$ in the image plane of (a) C_1 and (b) C_2 (Figure 6.1).

In reality calibration parameters and extracted silhouette boundaries contain error (Forbes, Voigt and Bodika 2003, Forbes 2007). This causes the epipolar tangent lines not to project exactly through the tangency point in the opposite image (Figure 6.4), resulting in an inconsistent pair of racket silhouettes.

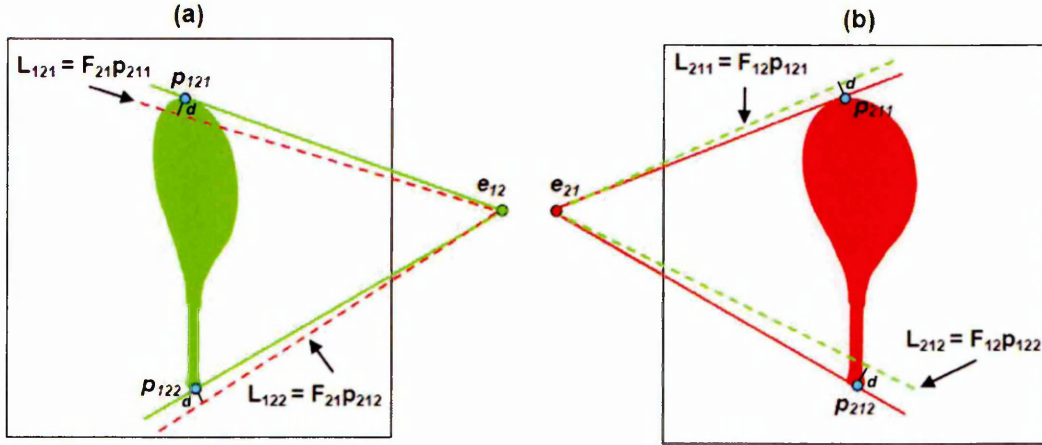


Figure 6.4 Error associated with real camera parameters and silhouette boundaries results in the epipolar tangency lines not projecting exactly through the tangency points in the opposite image for (a) C_1 and (b) C_2 in Figure 6.2.

Reprojection errors can be computed in order to measure the degree of inconsistency between a pair of racket silhouettes. These reprojection errors are the perpendicular distance d between a tangency point and a projected epipolar line (Figure 6.4) and can be computed using equation 2.15 (page 27). For a pair of racket silhouettes there are two tangency points in each image thus four perpendicular reprojection errors. The measure of inconsistency is based on the root mean square (RMS) value of all the reprojection errors and is referred to as the epipolar tangency error (ETE) .

6.4 Epipoles inside silhouettes

When an epipole falls within the boundary of a silhouette, the tangency points are not available and the ETE cannot be computed (Wong 2001, Forbes 2007). In Figure 6.2 this would occur if C_1 and C_2 were moved forward i.e. closer to the racket, so that the baseline passed through the racket. Cases in which the epipole lied inside a silhouette were efficiently identified by placing a rectangle around a racket silhouette boundary. If an epipole fell inside the rectangle it was disregarded when computing the ETE.

6.5 Locating the epipolar tangency points

Epipolar tangency points are computed from silhouette boundaries and are always available except when the baseline passes through the object. The advantage of using the two outer epipolar tangency points is that they are guaranteed to be in correspondence (Wong 2001). For example, using two corresponding points on the inside of the racket frame would mean that from some view orientations the points would become self-occluded by the racket. Using the two outer tangency points avoids self-occlusion and simplifies the matching of corresponding points (Wong 2001).

Lazebnik (2002) and Forbes (2007) advocate using polygonal silhouette boundaries rather than high-order cubic B-spline snakes (Cipolla and Blake 1990, Cipolla and Blake 1992, and Wong and Cipolla 2004), in favour of computational efficiency. Polygonal silhouette boundaries can be computed with sub-pixel accuracy (Forbes 2007) and only needs to be computed once for each silhouette. Whereas tangency points need to be computed repeatedly when the relative pose of a camera is adjusted during an iterative minimisation of the ETE, which will be addressed in chapter 7.

Figure 6.5 shows a simplified polygonal racket silhouette boundary. Epipolar tangency points (shown in blue) exist where tangency lines extending from the epipole touch the silhouette boundary (Lazebnik et al. 2002, Wong and Cipolla 2004).

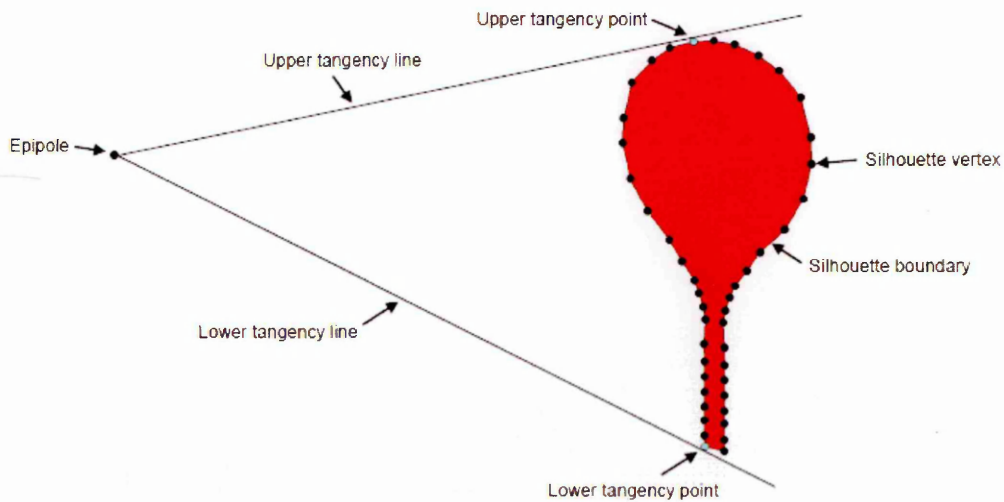


Figure 6.5 Upper and lower tangency point location for a racket silhouette.

In this project a method based on computing the angle of a tangency point relative to an epipole in the image was adopted (Figure 6.6). The angle between the positive x-axis and the tangency line associated with each vertex in the silhouette is calculated

(Figure 6.6). The angles are placed in the correct quadrant using MATLAB's `atan2` function. The y-axis in Figure 6.6 is inverted to coincide with the coordinate axis of an image plane. The two tangency points are defined by the largest and smallest angles, once the angle of all vertices on a silhouette boundary have been calculated.

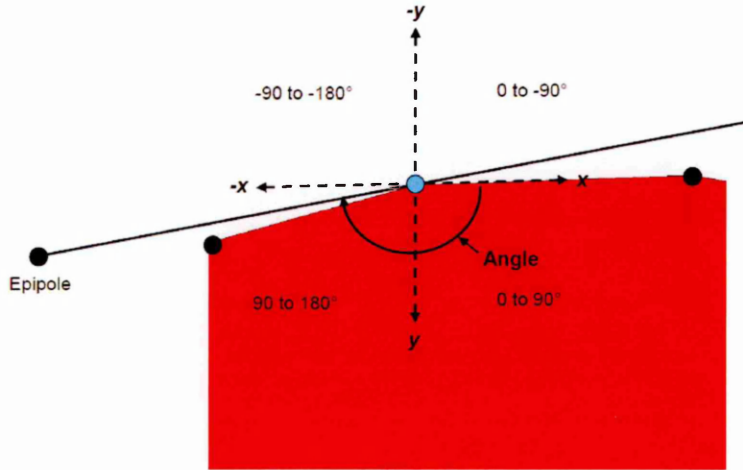


Figure 6.6 Illustration of the method used to locate the two tangencies. The angle on the image between the positive x-axis and the tangency line associated with all vertices in a silhouette boundary is calculated and placed in the respective quadrant. Vertices corresponding to the highest and lowest angles are the tangencies.

The largest and smallest angles do not necessarily correspond to the 'upper' and 'lower' tangency points respectively as seen on the simplified racket silhouette boundary in Figure 6.5. These extreme angles simply indicate the two vertices that are the tangency points. The method used to compute the correct tangency point correspondences across image pairs is explained in the next section.

6.6 Determining racket tangency point correspondences

Correctly determining tangency point matches across racket silhouette pairs was important so that reprojection error distances could be calculated. This project adopted a method used by Forbes (2007) and Wong (2001) that calculated the handedness of the epipole with respect to the tangency point on a racket silhouette boundary. Other methods also exist; Grattarola (1992) used the correspondence which provided the lowest ETE and Hénandez, Schmitt and Cipolla (2007) determined correspondences given the knowledge that camera poses were always orientated upright. However, Grattarola's (1992) method was complex to compute and the camera poses generated in this project were not always guaranteed to be upright.

Figure 6.7 (a) and (b) show a 3D point P captured by camera 1 and 2, the point P is visible to both cameras. In Figure 6.7 (a) the epipole e_{12} is on the opposite side of p_{12}

as e_{21} is of p_{21} , thus the corresponding tangencies have opposite handedness. In Figure 6.7 (b) the epipole e_{12} is on the same side of p_{12} as e_{21} is of p_{21} , thus the corresponding tangencies have the same handedness. A function was written in MATLAB in order to compute the correspondences as follows:

- Compute the handedness given the poses of a camera pair. It did this using the knowledge that if one of the camera's z-coordinate (specified in the opposite camera's reference frame) was below zero, then the cameras had same handedness.
- In Figure 6.7 (a) both cameras are orientated so that they are in front of each other. Therefore, the z-coordinate of both cameras will be above zero when specified in the opposite camera's reference frame, resulting in opposite handedness.
- In Figure 6.7 (b) camera 1 is orientated so that camera 2 is behind it therefore, the z-coordinate of camera 2 will be below zero when specified in the reference frame of camera 1, resulting in the same handedness.
- Correctly paired tangencies were then passed to a function that computes the four reprojection errors using equation 2.15 (page 27).

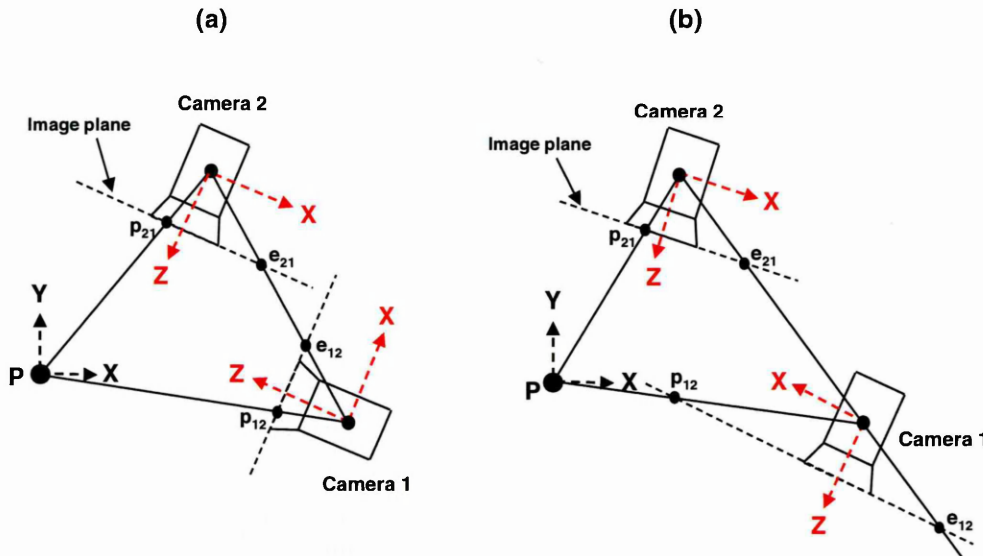


Figure 6.7 Handedness of the epipole with respect to the tangency point is used to decipher whether tangency correspondence is (a) opposite or (b) the same.

6.7 Summary

The main aspects of computing the inconsistency for a pair of fully calibrated racket silhouettes have been identified. Methods for efficiently computing the tangency points and their correspondences for racket silhouettes have been discussed. The following section will use synthetic data to measure the inconsistency of racket silhouettes.

6.8 Synthetic data generation

Camera parameters will always contain some error and are often obtained from complex calibration produces which can be time consuming and inaccurate. Therefore, synthetic camera parameters (internal and external) were used to ensure exact parameter values are considered in order to investigate the influence of discrete pixel rendering and silhouette extraction on the ETE. Previous studies that have adopted silhouette-based pose optimisation methods (Wong 2001, Forbes 2007, Price and Morrison 2007) have all used synthetic data in order to understand different aspects of error and to compare this with real data.

This section will create:

- A 3D racket model in Blender v2.70.
- Synthetic camera views of the model.

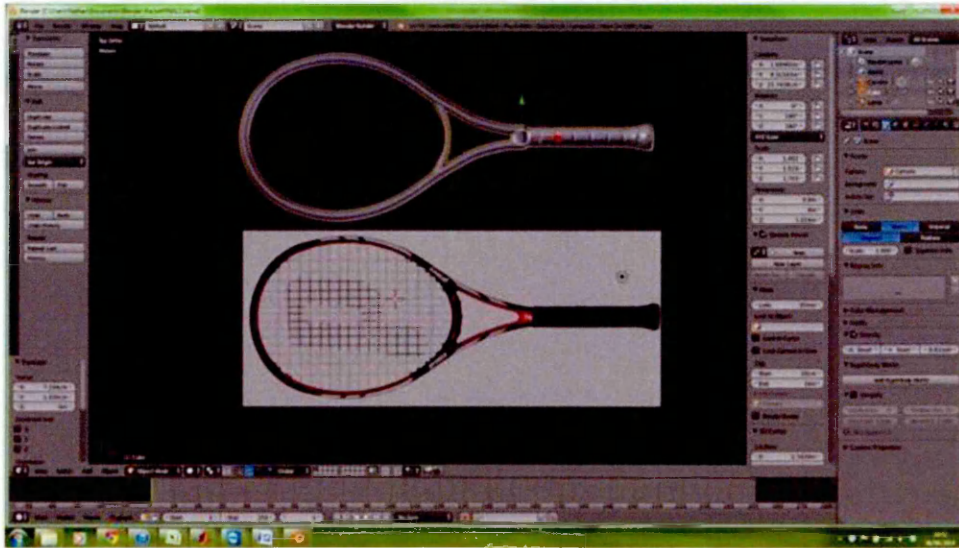
In order to:

- Measure the inconsistency of silhouettes associated with a 3D racket model.
- Measure the effect of discrete pixel rendering and silhouette extraction on the ETE.
- Investigate the effect of image resolution on ETE.

6.8.1 *The 3D racket model*

The 3D racket model was created in an open-source computer graphics software programme called Blender (v2.70). An image of a racket (Prince Warrior 100L ESP) was used as a template to form the shape of the model frame (Figure 6.8 (a)). The frame itself was created using a combination of Blender's inbuilt operations before being scaled to the size of the racket measuring 24 x 262 x 685 mm (depth, width, height). Figure 6.8 (b) is the final rendered racket model.

(a)



(b)



Figure 6.8 The 3D racket model created in Blender (v2.70) showing (a) the template image used to form the shape of the frame and (b) the final rendered model.

6.8.2 Generation of synthetic camera parameters

In Blender, a camera is considered to be an object that provides a means of rendering images of a model. Two cameras with the same relative poses with respect to the racket as the master and slave cameras from chapter 5 (Figure 5.6, page 70) were added to the Blender scene (Figure 6.9), along with the racket model.

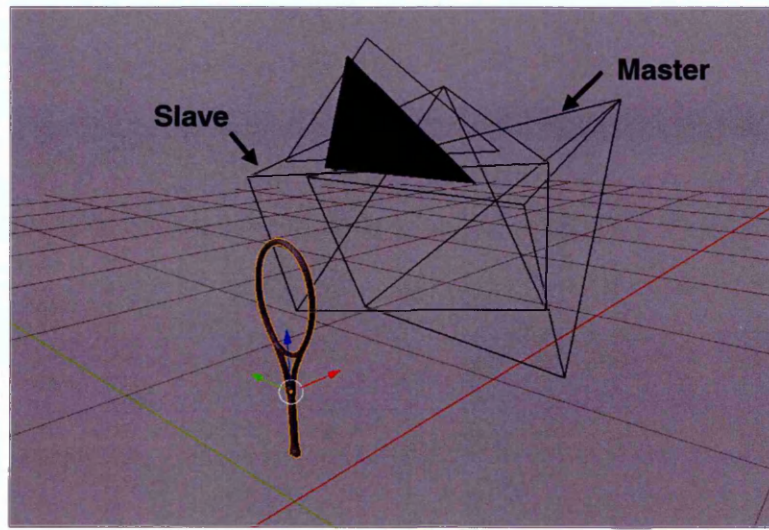


Figure 6.9 Synthetic relative master camera and slave camera poses with respect to the model racket in Blender scene.

Camera poses were constructed using a 3 x 1 translation vector and Euler rotations to comply with Blender's default settings. The synthetic images had a dimension of 1280 x 800 which was the same as the camera images in chapter 5. The intrinsic parameters of the two synthetic cameras closely matched that of the master and slave camera in chapter 5 but had to comply with Blender's settings. The calibration matrix for the master camera was given by

$$\begin{bmatrix} 1802.7 & 0 & 640.5 \\ 0 & 1802.7 & 400.5 \\ 0 & 0 & 1 \end{bmatrix} \quad (6.1)$$

and for the slave camera it was given by

$$\begin{bmatrix} 1778.1 & 0 & 640.5 \\ 0 & 1778.1 & 400.5 \\ 0 & 0 & 1 \end{bmatrix}. \quad (6.2)$$

To comply with Blender's settings an effective focal length was used and the principal point was located at the centre of the image. The effective focal length was based on the mean across f_x and f_y in Table 4.1 (page 58) for the slave and master camera. Lens distortion was not modelled for the synthetic intrinsic parameters. Equations for computing the intrinsic parameters from the data provided by Blender are in the Appendix.

6.9 Experiments

6.9.1 Discrete pixel rendering

This section calculated the four reprojection errors for the synthetic camera poses and silhouettes associated with the 3D racket model generated in the preceding sections of this chapter. In order to quantify the effect of discrete pixel rendering on silhouette consistency, two types of racket silhouettes were generated:

1. Synthetic. This involved extracting the outer 3D coordinates used to build the mesh of the 3D racket model in Blender (130 coordinates in total) and projecting them onto the respective image planes of the master and slave camera. No discrete pixel rendering was used and points were projected using synthetic camera parameters.
2. Rendered. This involved using Blender to render the two views of the model racket silhouettes at a resolution of 1280 x 800 pixels. It required setting foreground and background pixels to white and black respectively and switching off the inbuilt shadow feature. Once the silhouettes of the racket had been rendered, MATLAB's image processing toolbox was used to extract the polygonal silhouette boundaries to sub-pixel accuracy. The polygonal silhouette boundary extraction algorithm is described in section 2.3.3 (page 22-23).

Figure 6.10 shows epipolar lines plotted in the view of the master camera when synthetic parameters and silhouette boundaries were used. For clarity, Figure 6.10 shows 45 of the total 130 projected boundary points. Inset Figure 6.10, shows that the epipolar line projects exactly through the blue tangency point. Rendering the synthetic racket silhouette exactly consistent with that in the opposite slave camera view.

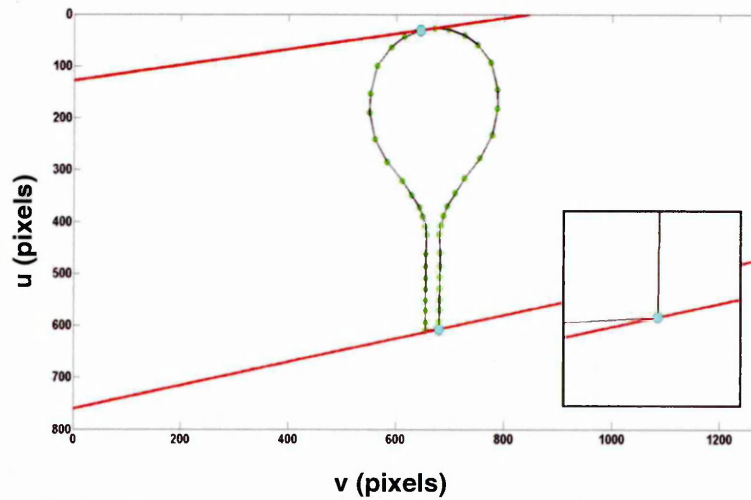


Figure 6.10 Epipolar lines plotted using synthetic camera parameters and synthetic racket silhouette boundary.

The scenario in Figure 6.10 was expected since both synthetic parameters and silhouette boundaries were used. However, when discrete pixel rendering is used and racket silhouettes are extracted using image processing techniques, the epipolar lines did not project exactly through the tangency points. Figure 6.11 shows an example of when this is the case. The quality of the silhouette in Figure 6.11 was quantified by super-imposing the extracted silhouette boundary over the same image of the racket model and visually inspecting how well they match up. The grey threshold was then updated until the best boundary match was observed. The literature review confirmed that subjective assessment of silhouette quality is sufficient when the silhouettes have been obtained in a controlled environment i.e. highly contrasting background and foreground pixels were used to render silhouettes in Blender (Chapter 2, page 24).

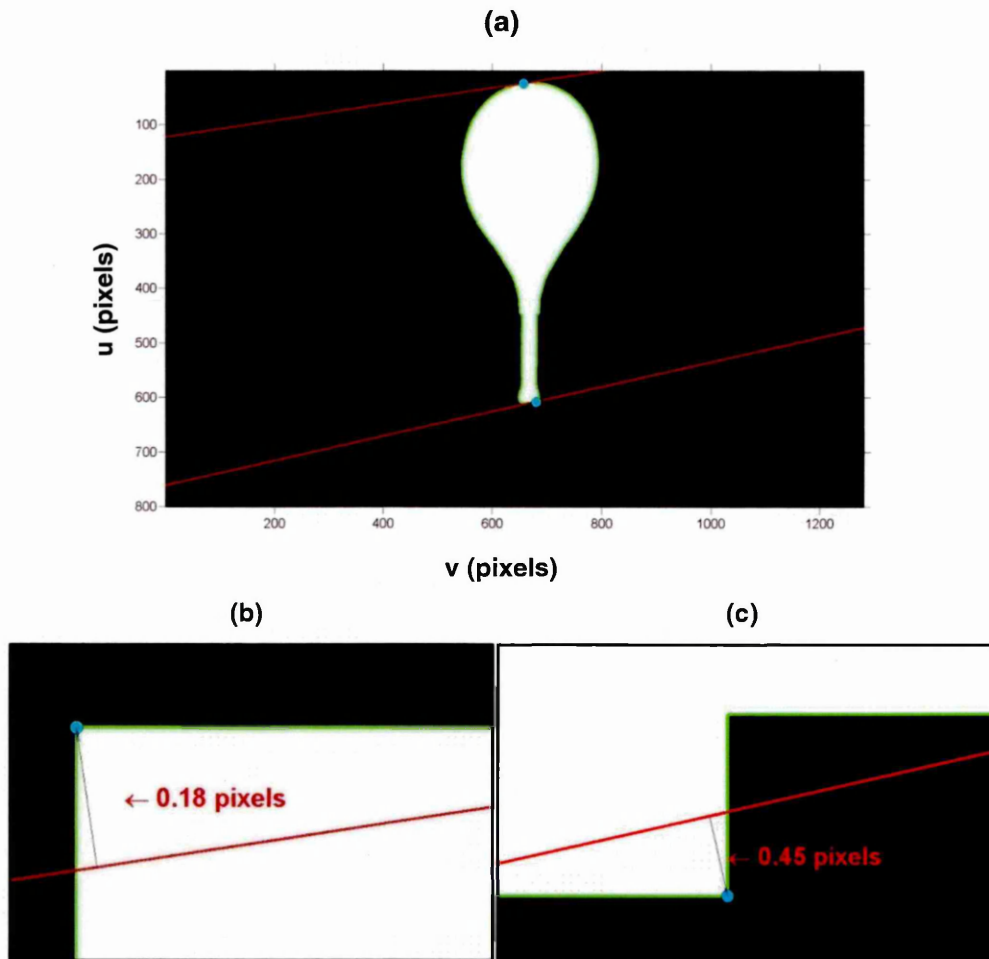


Figure 6.11 Epipolar lines plotted using synthetic camera parameters and rendered silhouette boundaries. The epipolar lines did not project exactly through (a) the upper and (b) the lower tangency points due to the square shape of the pixels.

Figure 6.11 (b) and (c) show the perpendicular reprojection errors between the epipolar tangent lines and the upper and lower tangency points in Figure 6.11 (a) were 0.18 and

0.45 of a pixel respectively. The same perpendicular reprojection error values were measured in the view of the slave camera. Note that perpendicular reprojection error values are not necessarily the same in both views. For the example in Figure 6.11, the RMS ETE was 0.34 pixels due to discrete pixel rendering of the racket silhouettes.

6.9.2 Comparing racket silhouettes with simpler shapes

The previous section showed that the square shape of a pixel caused the racket silhouettes not to be exactly consistent with the 3D racket model. This section compared the consistency of the same racket silhouettes with that of simple shapes i.e. a 2-D plane and a single pixel. Figure 6.12 (a) is a silhouette of a plane made up of a single pixel, it has four available tangency points indicated by the blue dots. Figure 6.12 (b) is a silhouette of a 2D plane with the same height and width as the racket and Figure 6.12 (c) is the silhouette of the racket. Each of the silhouettes in Figure 6.12 are in the view of the master camera, silhouettes in the slave camera view were also generated but are not shown.

Table 6.1 contains the RMS ETE values for the single pixel, the racket size 2D plane and the racket computed using the synthetic master and slave camera poses. It can be seen that the single pixel and racket have largest and smallest ETE values respectively. However, a one-way ANOVA revealed that there was a non-significant difference between the perpendicular reprojection error distances used to calculate the RMS ETE values in Table 6.1, $F(2,9) = 1.37$, $p > .05$. The results confirm that the square shape of a pixel causes silhouettes not to be consistent and that racket shaped silhouettes are no different from simpler shaped objects.

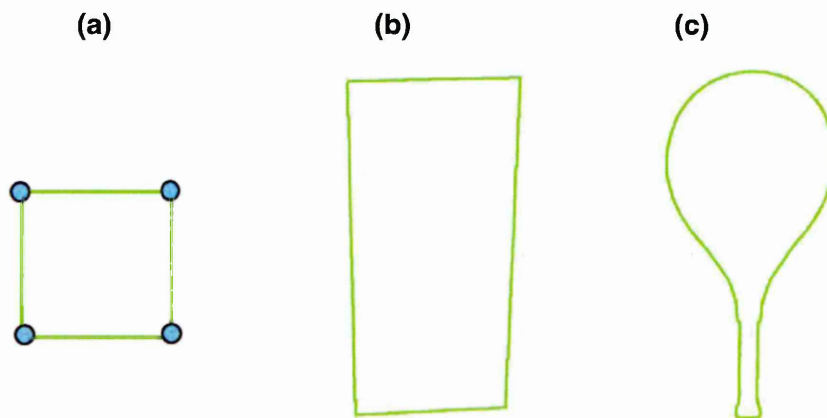


Figure 6.12 Silhouette outlines of (a) a single pixel, (b) a 2D plane with the same height and width as a racket and (c) the racket.

Table 6.1 RMS ETE values for silhouette pairings between a single pixel, a 2D plane with the same height and width as a racket and the racket.

	Single pixel	2D Plane	Racket
RMS ETE (pixels)	0.49	0.41	0.34

6.9.3 Silhouette resolution

In addition to the two synthetic camera poses, four synthetic fully calibrated sets were generated i.e. more synthetic camera poses were added to the Blender scene in Figure 6.9. The four different sets contained 72, 64, 40 and 20 views of the 3D racket model created in Blender. The camera poses in each of the four sets were used to render silhouette images associated with the 3D racket model. Initially, the silhouette images were rendered at a resolution of 2560 x 1600 pixels, and then at successively lower resolutions; 1280 x 800, 640 x 400, 480 x 300 and 320 x 200 pixels.

The RMS ETE in pixels across the four sets for each resolution was calculated and then converted to millimetres using the intrinsic calibration model generated in Blender. A description of how to calculate the number of pixels per millimetre using the Blender calibration model is in the Appendix. Table 6.2 presents the number of pixels per millimetre and millimetres per pixel for the five resolutions tested.

Table 6.2 Number of pixels per millimetre and millimetres per pixel for the five resolutions calculated using the Blender calibration model.

	Resolution				
	2560 x 1600	1280 x 800	640 x 480	480x300	320x200
Millimetres per pixel	0.0125	0.0250	0.0499	0.0665	0.0997
Pixels per millimetre	80	40	20	15	10

Figure 6.13 shows that the RMS ETE (mm) increased with decreasing image resolution. As expected there was no clear relationship between RMS ETE (mm) and the number of synthetic camera poses within a fully calibrated set; more poses simply provides a more reliable RMS ETE. The fully calibrated set with 20 poses displayed a marginally larger RMS ETE (mm) for the middle three resolutions. Figure 6.13 confirms that silhouettes should be rendered at the highest available resolution in order to provide the most accurate fully calibrated set associated with a rackets pose. The maximum resolution of the high-speed Vision Research Phantom Miro M110 camera available for this project is 1280 x 800.

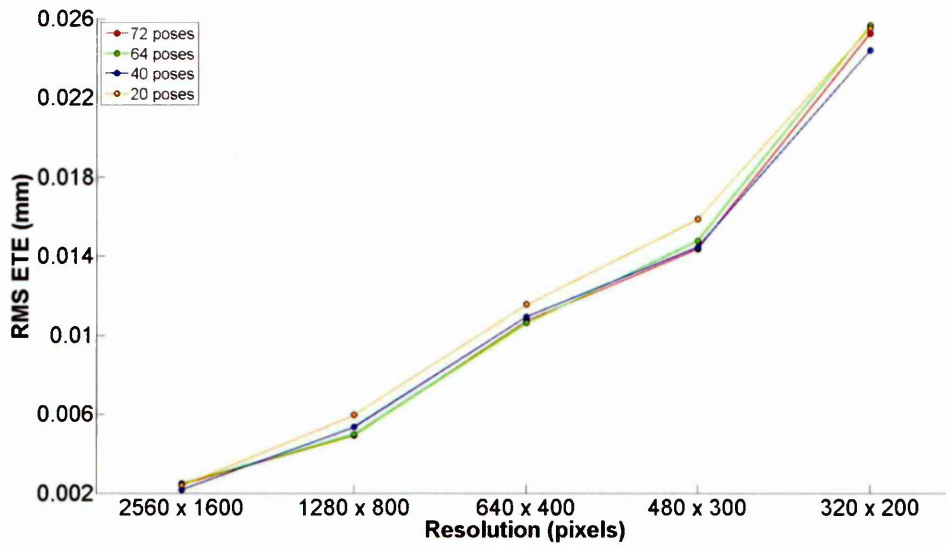


Figure 6.13 The RMS ETE (mm) across synthetic fully calibrated consisting of 72, 64, 40 and 20 camera poses using racket silhouettes rendered at successively lower resolution. The RMS ETE (mm) increases as resolution decreases.

6.10 Discussion

This chapter has applied the epipolar tangency constraint to a pair of racket silhouettes. Synthetic camera poses and rendered silhouettes associated with a 3D racket model were generated, these corresponded with the master and slave camera poses from chapter 5. Allowing the effect of discrete pixel rendering on racket silhouette inconsistency to be quantified. Discrete pixel rendering and extraction of racket silhouettes increased the RMS ETE from zero (when synthetic camera poses and silhouettes were used) to 0.34 pixels. This value would be more reliable if more views of the racket model were used. However, the purpose of the investigation was to simply demonstrate the influence of discrete pixel rendering and racket silhouette extraction on the RMS ETE using these two camera poses.

Figure 6.11 demonstrated that discrete pixel rendering acted to increase the inconsistency of silhouettes. Table 6.1 shows that there was a marginal difference in the RMS ETE between silhouette pairs of a single pixel, a 2D plane with the same height and width as the racket and the racket. This suggests that it is the square shape of pixels that cause the RMS ETE to increase as a result of discrete pixel rendering. Furthermore, the overall shape of a silhouette does not influence the RMS ETE. Use of B-spline snakes (Cipolla and Blake 1990, Cipolla and Blake 1992, and Wong and Cipolla 2004) to extract silhouette boundaries could decrease the RMS ETE but these are more computationally expensive than the polygonal boundaries used in this thesis (Matusik, Buehler, and McMillan 2001, Lazebnik 2002).

Another advantage of using Blender is that synthetic camera poses can be added to a scene at any time. In the last section of this chapter additional synthetic poses were generated to create fully calibrated sets consisting 72, 64, 40 and 20 views. Blender was used to render silhouettes associated with the 3D racket model at successively lower resolutions. Figure 6.13 showed that the RMS ETE (mm) across a fully calibrated set consisting of synthetic camera poses increased as the resolution at which silhouettes associated with the 3D racket model were rendered decreased. This agrees with Forbes (2007) who successively decreased the resolution of polygonal silhouettes associated with stones. Forbes (2007) reported increasing RMS ETE values between 0.1 and 0.4 pixels as resolution was decreased seven successive times from 2592 x 1944. In this study, the largest and smallest RMS ETE values were 0.17 and 0.26 pixels which corresponded to 0.002 and 0.025 mm for resolutions of 2560 x 1600 and 320 x 200 respectively.

It is preferential to use the highest resolution available in order to obtain more consistent fully calibrated silhouette sets. The high-speed Vision Research Phantom Miro M110 cameras available for use in this project capture images at a maximum resolution of 1280 x 800. Figure 6.13 shows that using this resolution produces an RMS ETE comparable to the highest resolution tested of 2560 x 1600 when 72 and 64 camera poses are used for fully calibrated set. However, there is no clear evidence that the effect plateaus at a particular resolution. The findings highlighted that images should be captured using the highest resolution of 1280 x 800 when creating a fully calibrated set using the high-speed Vision Research Phantom Miro M110 cameras. Furthermore, this chapter has shown that obtaining accurate silhouettes is vital for view fitting methods that rely on an accurate fully calibrated set. Previously, it has been demonstrated that discrete pixel rendering and silhouette extraction makes a larger contribution to the RMS ETE than camera parameter error (Forbes 2007).

Finally, Figure 6.13 also indicated that there is no clear relationship between RMS ETE and the number of camera poses in a fully calibrated set. However, the RMS ETE values computed for the two camera poses based on the master and slave camera poses (section 6.9.1) were larger (0.34 pixels) compared to the values presented for sets consisting of more poses in Figure 6.13 (range = 0.17 to 0.26 pixels). This observation is unlikely to be significant since the RMS ETE simply becomes more reliable as the number of camera poses in the fully calibrated set increases. Nevertheless, Price and Morrison (2007) found that the accuracy of view fitting methods increases with the number of poses in a fully calibrated set.

6.11 Conclusion

Discrete pixel rendering contributes to the inconsistency of silhouette sets. However, racket shaped silhouettes exhibit comparable RMS ETE values to simpler shaped silhouettes. Polygonal silhouette quality can be improved by rendering images at the highest resolutions available which in turn decreases the RMS ETE. Therefore, silhouette images should be captured at the highest resolution possible at a cost of increased computational requirements. Furthermore, the number of camera poses in a fully calibrated set has no clear influence on the RMS ETE.

Chapter 7 will determine the most appropriate camera pose configuration for a fully calibrated set associated with a racket. It will then use optimisation methods to minimise the RMS ETE by adjusting camera poses in order to estimate 3D racket position using synthetic data.

7 Optimising the configuration of a fully calibrated set of racket silhouettes using synthetic data

7.1 Introduction

This chapter formulated an objective function in order to minimise the RMS ETE across synthetic fully calibrated sets using iterative optimisation techniques. The literature review identified that the Levenberg-Marquardt optimisation method has previously been used to minimise the ETE (Forbes, Voigt and Bodika 2003, Price and Morrison 2007, Sheets et al. 2011). This optimisation method is preferred because it computes the sum of squared reprojection error distances. A validation scheme whereby each silhouette is removed from the set then its pose estimated using the method was adopted (Price and Morrison 2007). By adjusting parameters of a candidate relative pose (rotation and translation) to minimise the cost, an accurate estimation of the true pose of a camera can be made. Camera pose estimates were used to reconstruct the geometry of the racket face plane. A range of synthetic pose configurations were tested and the results were used to inform the creation of a fully calibrated set using the Vision Research Phantom Miro M110 cameras in the laboratory. Since synthetic camera poses are used, this chapter quantifies the influence of discrete pixel rendering and silhouette extraction on 3D racket position error (Figure 7.1) .

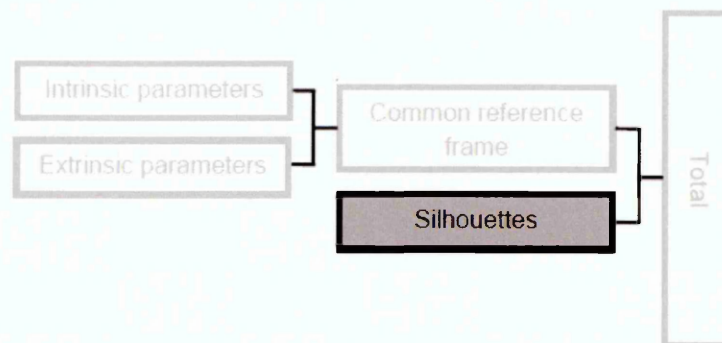


Figure 7.1 Chapter 7 used synthetic camera poses to quantify the influence of discrete pixel rendering and extraction of silhouettes on 3D racket position estimation.

7.2 Aim and objectives

The aim of this chapter was to validate racket pose estimates using synthetic data to determine the most appropriate configuration for a fully calibrated racket silhouette set.

Objectives:

1. To generate candidate relative camera poses for the objective function.
2. To use different fully calibrated set configurations to obtain camera pose estimates using a take-one-out validation method.
3. To reconstruct the racket face plane using the camera pose estimates.

7.3 Formulating the ETE as an objective function

The objective function for optimisation was specified by a concatenated vector of perpendicular reprojection errors (Forbes 2003, Wong 2003, Price and Morrison 2007). Formed in MATLAB by considering all possible silhouette pairings in a synthetic fully calibrated set. This results in $4n$ perpendicular reprojection errors (from $2n$ pairs), where n is the number of silhouettes in the set and each silhouette has two outer tangency points. Therefore, in vector terms the optimisation problem can be stated as

$$\min \|f(x)\|_2^2 = \min(f_1(x)^2 + f_2(x)^2 + \dots + f_n(x)^2) \quad (7.1)$$

where x is the vector of ETE values and $f(x)$ is the cost function that returns a vector of minimised ETE values. Forbes (2007) and Price and Morrison (2007) suggest specifying the nine parameter 3×3 rotational matrix R as a four parameter unit quaternion Q . This provides the optimisation method with fewer parameters to adjust. Therefore, the candidate relative pose can be specified by a vector of seven parameters

$$X = [Q \ t], \quad (7.2)$$

where Q is a 1×4 quaternion and t is a 1×3 translation vector. The Levenberg-Marquardt algorithm was used to adjust the seven pose parameters to minimise the ETE distances. The pose estimation task corresponds to obtaining appropriate values for the pose parameters Q and t .

7.4 Candidate relative poses

A common issue with optimisation frameworks is their sensitivity to the initial parameters; in this case the initial candidate relative camera pose (Wong 2001, Forbes, Voigt and Bodika 2003, Price and Morrison 2007). A good initial candidate relative pose is required i.e. one that is close to the true camera pose. Otherwise the Levenberg-Marquardt algorithm may converge to a local minimum that is far from the global minimum i.e. the true camera pose.

Various methods have been used to initialise candidate relative poses in order to minimise the ETE. Wong and Cipolla (2004) exploited a visual hull of a small clay toy built from a fully calibrated set of coplanar camera poses. Vertices were extracted from the triangular mesh used to build the visual hull and projected onto new silhouette views associated with the clay toy. Wong and Cipolla (2004) found that a good initialisation could be obtained by manually rotating and translating the candidate relative pose until the projection of the visual hull roughly matches the boundary of the new silhouette view. The method developed by Wong and Cipolla (2004) relies on visual judgement.

Forbes, Voigt and Bodika (2003) and Forbes (2007) used merged visual hull models in order to determine initial candidate relative poses. The translational component was initialised using the difference between the centroids and the rotational component was initialised by determining the transformation between the principal axes of two visual hull models. However, the methods used by Wong and Cipolla (2004), Forbes, Voigt and Bodika (2003) and Forbes (2007) relied on a visual hull. The intention of this thesis was not to create a visual hull of a racket since this is computationally expensive and as the literature identified, not necessary for single view fitting methods.

In Price and Morrison (2007), limited information is provided regarding initial parameter selection for ETE minimisation using the Levenberg-Marquardt method. However, Price and Morrison (2007) state that initialisation tasks associated with more regular shaped objects (i.e. symmetrical) is more crucial because local minima are more likely to occur since more than one candidate relative pose could be equally consistent. To overcome this, Price and Morrison (2007) apply multiple optimisations using different initial candidate relative poses. Furthermore, Price and Morrison (2007) identify the rotation component of a candidate relative pose to be the primary cause of the Levenberg-Marquardt optimisation method becoming trapped in a local minima when minimising the ETE. The workaround used by Price and Morrison (2007) was to repeatedly sample the rotational component of candidate relative poses from a normal distribution. The optimisation was then terminated when the RMS ETE was below a threshold of 1.5 pixels.

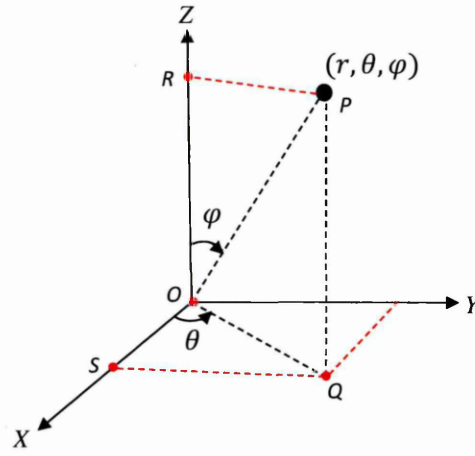


Figure 7.2 A spherical coordinate (adapted from WolframMathWorld 2015).

This project generated multiple initial candidate relative poses using spherical coordinates (Figure 7.2). A point P in a spherical coordinate system is defined in 3-dimensional space by its radial distance (r) from a fixed origin, its azimuthal angle (θ) from the positive x -axis to its orthogonal projection on the x - y plane and the polar angle (φ) measured from a fixed zenith direction. Following the mathematical convention r , θ and φ are referred to as rho, theta and phi respectively.

Spherical coordinates were converted to Cartesian coordinates (x , y , z) using trigonometry. The z -coordinate of point P is equal to

$$z = r \cos \varphi. \quad (7.3)$$

The triangle OSQ (Figure 7.2) also has a hypotenuse of length $r \cos(\varphi)$, the side OS is adjacent the angle θ , therefore

$$x = r \sin \varphi \cos \theta. \quad (7.4)$$

Finally, the side SQ is opposite angle θ hence,

$$y = r \sin \varphi \sin \theta \quad (7.5)$$

Once the spherical coordinates had been converted to X , Y , Z Cartesian coordinates, each was assigned a reference frame. This involved calculating the direction between each spherical coordinate (X, Y, Z) and the centre of the racket model in order to obtain a Z -axis. The butt of the racket model was centred at the global origin of the Blender scene (Figure 7.3). Therefore the centre was defined to be midway up the racket model

i.e. 0, 0 and 342.5 mm in the X, Y and Z directions respectively (Figure 7.3 (a)). An orthogonal X-axis (with respect to the Z-axis) was then calculated followed by the Y-axis which was obtained using the cross product of the Z and X axes (Figure 7.3 (b)). This created an orthogonal reference frame (with respect to the Blender racket model) for each spherical coordinate, forming candidate relative camera poses i.e. each candidate pose had a relative rotation and translation.

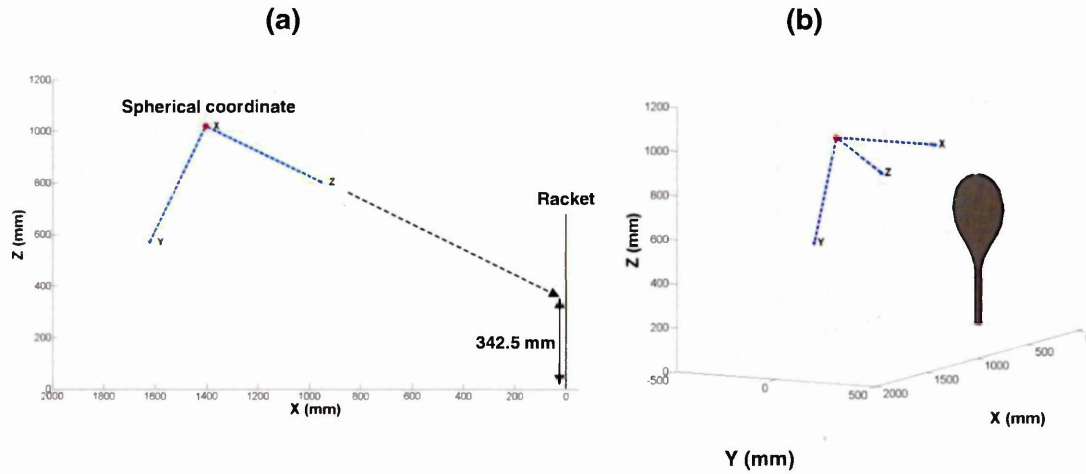


Figure 7.3 Method used to assign each spherical coordinate a reference frame with respect to the racket model. (a) shows a side view and (b) shows a 3D view.

The configuration for the candidate relative poses was then defined. A racket has two perpendicular planes of reflective symmetry, intersecting along its long axis (Srinivasan 2004). Therefore, an entire spherical configuration was not used because local minima are more likely to occur if candidate poses that are far from the solution are used. Instead, angles for theta and phi (Figure 7.2) relative to the racket were manually set so that the configuration of the candidate relative poses surrounded the true pose (Figure 7.4). The blue cameras in Figure 7.4 (a) are the initial relative candidate poses and the red camera is the true pose. Approximate knowledge of the true pose could be obtained by manually measuring the horizontal and vertical distance (X and Z axis in Figure 7.4) of the camera with respect to the racket in order to obtain a resultant distance i.e. a rho value (Figure 7.2). In Figure 7.4, the rho value is 1.7 m. The angles for phi and theta are then set so that the configuration of the relative candidate poses extends 30° either side of the true pose Figure 7.4 (b).

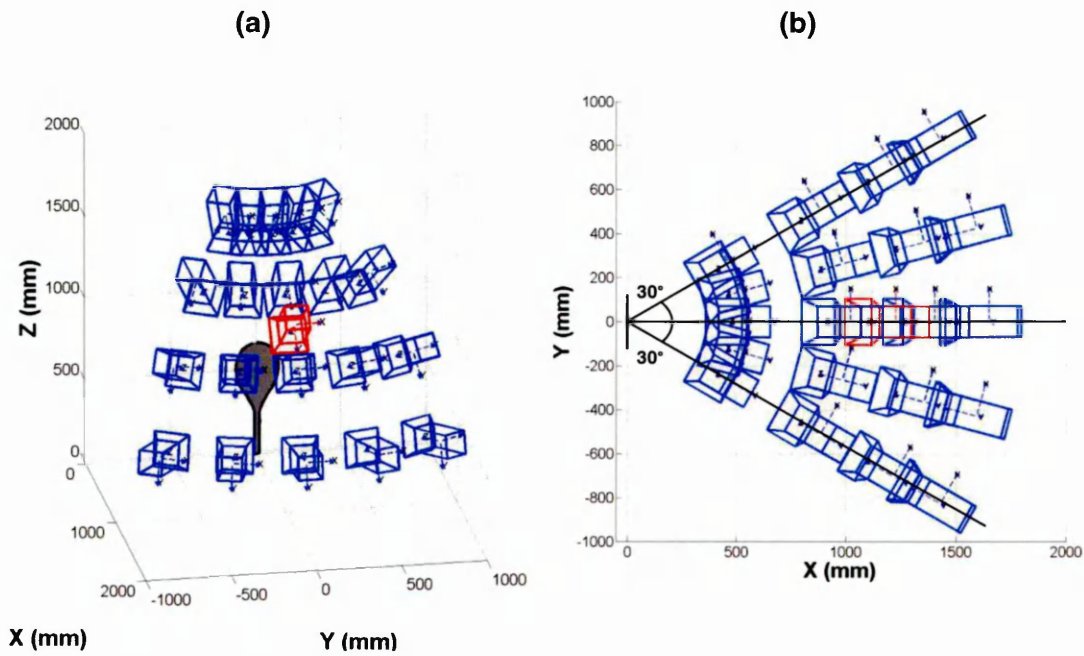


Figure 7.4 Candidate relative poses generated using spherical coordinates showing (a) a 3D view and (b) a top view.

An angle of 30° was chosen because this minimised the chance of the Levenberg-Marquardt converging to a local minima. Angles of up to 90° (either side of the true pose) were tested however, these resulted in the antipodal view being found. Antipodal views occur because random orientations close to the antipodal view appear similar to the actual view and result in a local minimum of the ETE function (Price and Morrison 2007). An error of approximately 180° for camera rotation about the optical axis is characteristic of when the antipodal view has been found (Price and Morrison 2007).

For clarity, Figure 7.4 does not show all the candidate relative poses used. In this project, a maximum of 100 candidate relative poses were used (in the same 60° area), which is the same number of initial conditions as used by Price and Morrison (2007). To obtain sub-pixel accuracy, searches were terminated by setting a threshold of 0.5 pixels on the RMS value of the ETE vector. A threshold of 0.5 produced the best results; a threshold of 1 pixel did not allow the optimisation to fully converge in some instances and a threshold below 0.5 made no difference to the solution. A threshold was required since discrete pixel rendering and silhouette extraction was used therefore, the RMS ETE values across the fully calibrated sets used would not be zero, even if the correct solution was found (see section 6.9.1, page 86-88).

7.5 Method

To assess camera pose estimate accuracy a take-one-out validation method was used in which each silhouette is removed from the set in turn and its pose is estimated using

the method (Price and Morrison 2007). Camera pose estimates were compared with synthetic ground truth camera poses and silhouettes were rendered at a resolution of 1280 x 800 pixels. For each pose estimation, a maximum of 100 candidate relative poses were used in the configuration shown in Figure 7.4. Optimisations were terminated if the RMS ETE went below a threshold of 0.5 pixels since this produced the best results. Other thresholds of 1 pixel and 0.25 of a pixel were also tested. However, the threshold of 1 pixel did not allow the optimisation to fully converge in some instances and the threshold of 0.25 made no difference to the solutions. Price and Morrison (2007) recommend using a threshold that allows for sub-pixel accuracy of the ETE vector i.e. a threshold of 1 pixel or below. A threshold was required because discrete pixel rendering and silhouette extraction was used, which means the RMS ETE value across the fully calibrated sets used will never be zero i.e. there will always be some degree of inconsistency.

Firstly, dome shaped and circular synthetic fully calibrated set configurations containing 80, 72 and 18 poses were tested. Secondly, fully calibrated sets with silhouette views of only one side of the racket face plane were tested, consisting of 36, 27, 18 and 9 views. The area in which camera poses were estimated decreased with the number of racket silhouette views in a synthetic fully calibrated set configuration.

Camera pose estimates obtained using the fully calibrated set configurations were then used to reconstruct 3D coordinates on the racket face plane. In total, 130 coordinates were reconstructed for each camera pose estimate, using a camera-plane model (Dunn et al. 2012). A detailed explanation of computing a camera-plane model can be found in Dunn et al. (2012) and Dunn (2014). Reconstructed coordinates were compared with known 3D coordinates obtained from the racket model mesh generated in Blender (section 6.8, page 83-85).

Finally, the viability of the take-one-out method to validate 3D racket position estimation was investigated. Since the configuration of the fully calibrated set changes by one view for each pose estimate during the take-one-out method. The effect of removing a single silhouette view from a fully calibrated set is quantified. Camera poses associated with ten racket silhouettes that were not originally part of the fully calibrated set were estimated. These camera poses were within the same area as the fully calibrated set configurations. Racket position estimation was carried out using a fully calibrated set consisting of 27 views and different fully calibrated set configurations of 26 views. Difference in racket position estimation error obtained using the sets was calculated.

7.6 Results

Figure 7.5 illustrates a sample optimisation between a pair of cameras using a fully calibrated set consisting of 80 silhouette views associated with the racket model.

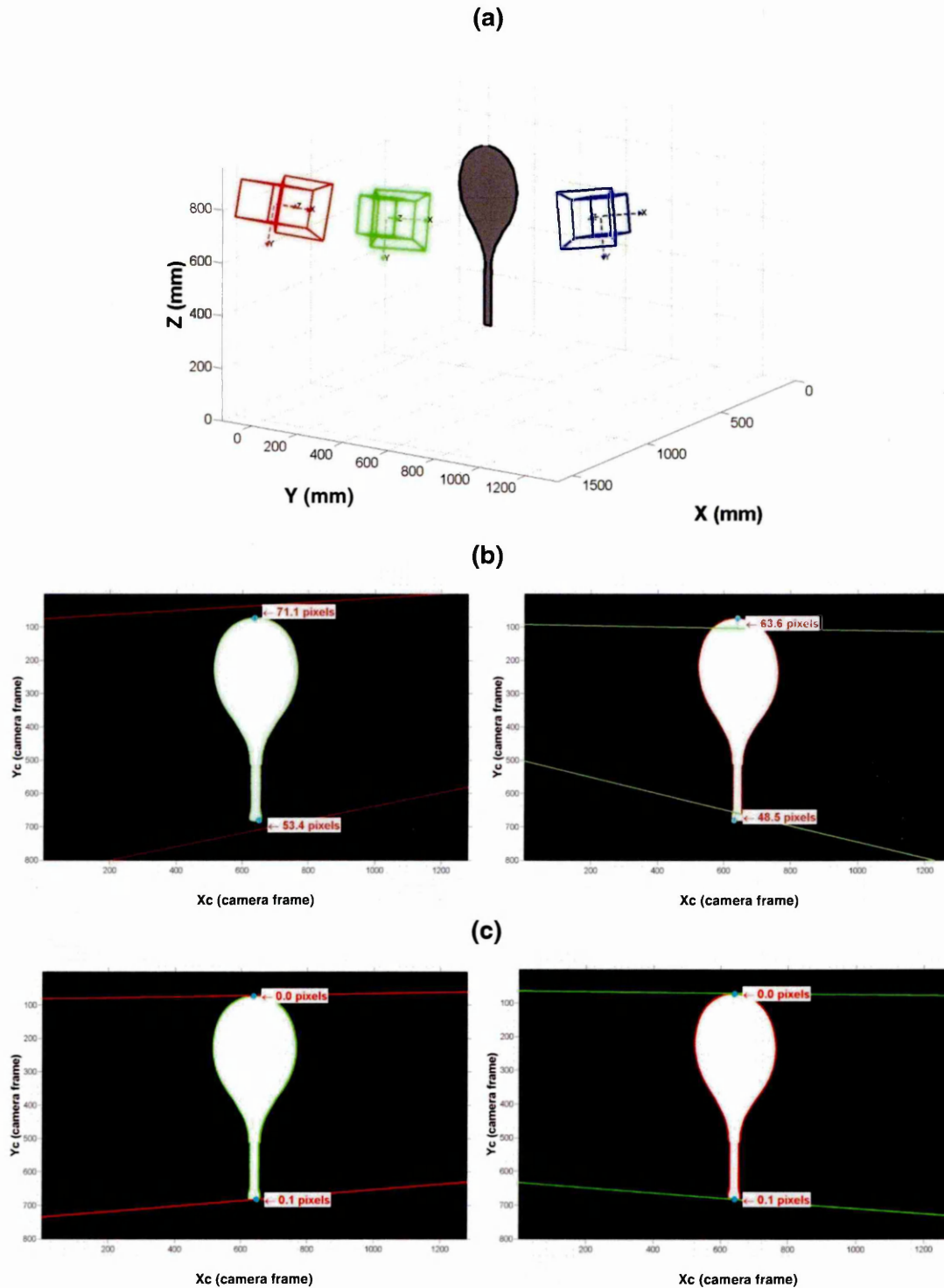


Figure 7.5 Illustration of a sample optimisation using showing (a) positions of the ground truth left camera (red), the initial candidate relative pose (blue) and the camera pose estimate (green), (b) the initial ETes and (c) the optimised ETes.

In Figure 7.5 (a), the synthetic ground truth left camera is shown in red. The blue camera is the initial candidate relative pose. Figure 7.5 (b) shows the epipolar lines plotted using the pose parameters associated with the ground truth left camera (red) and the candidate relative pose (blue). The initial RMS ETE is 59.8 pixels. The Levenberg-Marquardt algorithm is then used to adjust the pose parameters of the candidate relative pose in order to minimise the RMS ETE. The resulting camera pose estimate is shown in green (Figure 7.5 (a)). Figure 7.5 (c) shows the epipolar lines plotted using the ground truth left camera (red) and the camera pose estimate (green). The RMS ETE decreased to 0.07 pixels, requiring 26 iterations of the objective function.

7.6.1 Dome shaped fully calibrated set configurations

Figure 7.6 (a), (b) and (c) show translation error in the X, Y and Z directions respectively for camera pose estimates obtained using a fully calibrated set consisting of 80 synthetic camera poses in a dome shaped configuration.

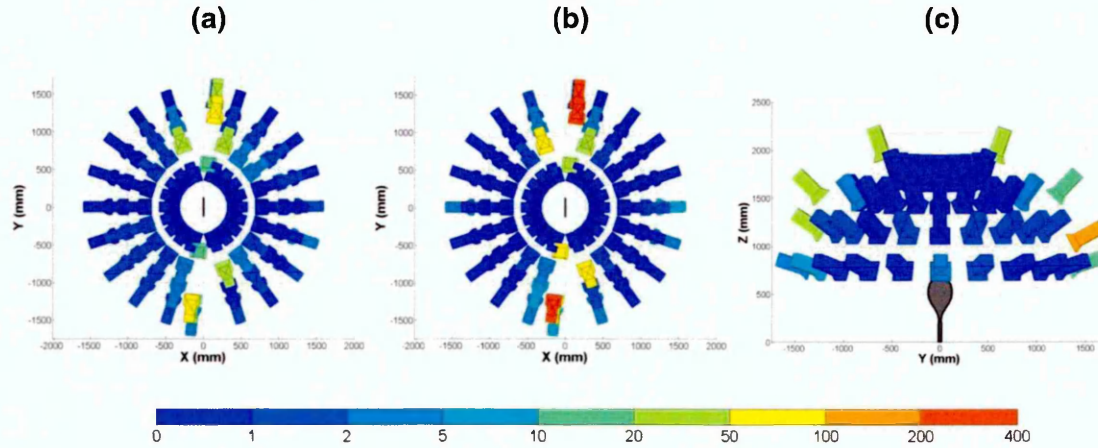


Figure 7.6 Camera pose estimate error for translation in the (a) X, (b) Y and (c) Z directions using an 80 pose synthetic set. The largest errors corresponded to side views of the racket.

Translation error in the X, Y and Z directions was largest for pose estimates associated with side views of the racket. Camera poses associated with side views had an angle larger than 80° with respect to the racket face plane normal. Mean translation errors were 46.7 ± 66.4 , 83.8 ± 77.8 and 25.8 ± 15.9 mm in the X, Y and Z directions respectively. These results occurred because side silhouette views have a markedly different shape compared to silhouette views that include the racket face plane. A silhouette containing a full view of the racket face plane (Figure 7.7 (a)) is made up of 50% more white pixels relative to black pixels than a silhouette containing a side view of the racket (Figure 7.7 (b)). Based on these results, side silhouette views associated with the racket model were removed from future fully calibrated sets.

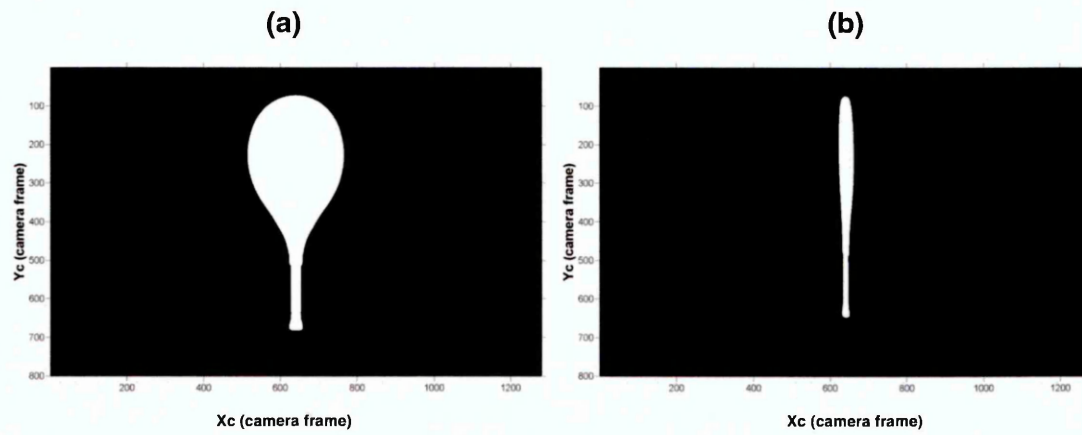


Figure 7.7 Silhouettes associated with (a) a frontal view and (b) a side view of the racket model.

Figure 7.8 (a), (b) and (c) show translation errors in the X, Y and Z directions respectively for camera pose estimates obtained using a dome shaped configuration containing 72 poses with no side silhouette views of the racket model. Removal of the side views from the set increased the accuracy of all the camera pose estimates. In Figure 7.8 all the camera pose estimates had translational errors of less than 15 mm in all directions. Mean translation errors were 0.48 ± 0.45 , 0.50 ± 0.61 and 0.37 ± 0.50 mm in the X, Y and Z directions respectively. Removal of side views greatly decreased mean translation error for all views. This is highlighted by the smaller mean \pm standard deviation values compared to those in Figure 7.6. Table 7.1 details rotation error about the axis' of X, Y and Z. The largest rotation error was 0.03° about the Z axis.

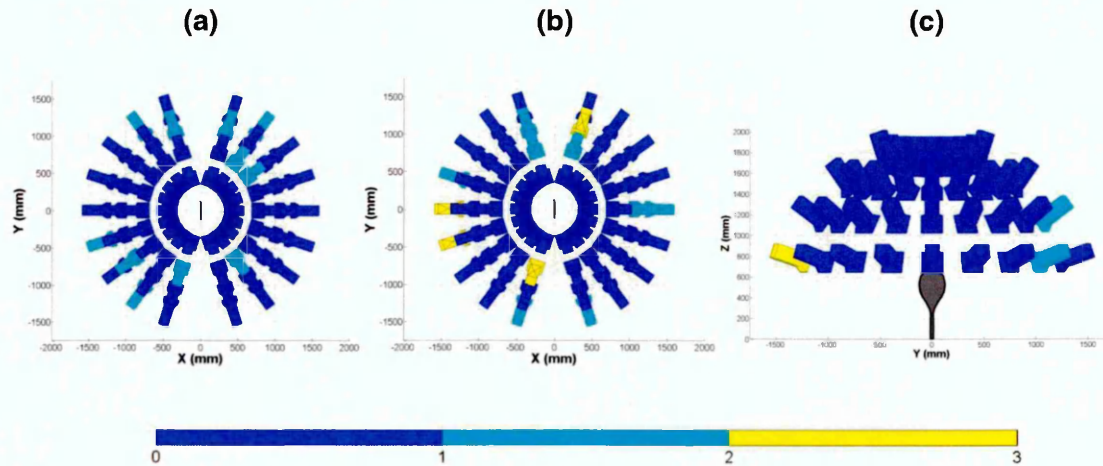


Figure 7.8 Translation errors in the (a) X, (b) Y and (c) Z directions for camera pose estimates made using a 72 pose dome shaped fully calibrated set configuration.

Table 7.1 Rotation error of pose estimates made using a 72 pose dome shaped and 18 pose circular configuration. Computation time for the optimisation to converge to a solution using a 3.4 GHz machine is also shown.

Rotation error direction	72 pose dome set	18 pose set	Difference
X (°)	0.02 ± 0.03	0.29 ± 0.26	0.27
Y (°)	0.01 ± 0.01	0.06 ± 0.05	0.05
Z (°)	0.03 ± 0.03	0.18 ± 0.13	0.15
Computation time (seconds)	2.4	0.9	1.5

Figure 7.9 (a), (b) and (c) show translation errors in the X, Y and Z directions respectively for camera pose estimates obtained using a circular configuration containing 18 poses with no side silhouette views of the racket model. This set was obtained by removing the upper three tiers from the 72 pose set in Figure 7.8. Camera pose estimates were less accurate using this set compared with the 72 pose set (Figure 7.8). Overall, 100, 100 and 83 % of pose estimates made using the 18 pose set had translational errors of less than 15 mm in the X, Y and Z directions respectively. Mean translation errors were 4.02 ± 3.44 , 3.45 ± 3.34 and 7.15 ± 6.42 mm in the X, Y and Z directions respectively. The 72 pose set provides 3 times the number of constraints (views) than the 18 pose set and as a result generates more accurate camera pose estimates. Table 7.1 also details rotation error of the pose estimates for the 18 pose set. The largest rotation error was 0.29° about the X axis.

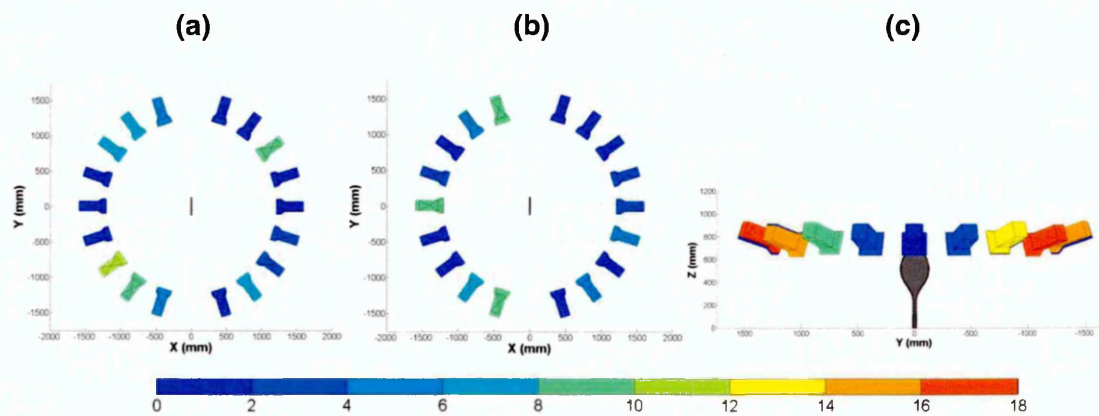


Figure 7.9 Translation errors in the (a) X, (b) Y and (c) Z directions for camera pose estimates made using a circular 18 pose fully calibrated set configuration.

7.6.2 Semi-circular configurations

Since the silhouette of a racket is the same shape when viewed from both sides of the racket face plane. The suitability of fully calibrated set configurations that contain half the number of camera poses compared to those sets in the previous section were

investigated. These sets view only one side of the racket face plane and do not contain side views of the racket therefore are referred to as semi-circular configurations. Four semi-circular configurations were investigated, each with one less tier of poses than the previous i.e. 4,3,2, and 1 tiers.

Figure 7.10 (a), (b) and (c) shows translation error in the X, Y and Z directions respectively for camera pose estimates obtained using a semi-circular configuration consisting of 36 poses, arranged in four tiers. Using this configuration, all of the camera pose estimates had translation errors of less than 15 mm in the X, Y and Z directions. On average, camera pose estimate translation error was only 0.5 mm less accurate in all directions, compared to when the dome shaped 72 pose set was used (Figure 7.8). Mean convergence time of the optimisation decreased from 2.4 (Table 7.1) to 1.9 (Table 7.2) seconds when this set was used rather than the 72 pose set on a 3.4 GHz Intel Core i7 machine.

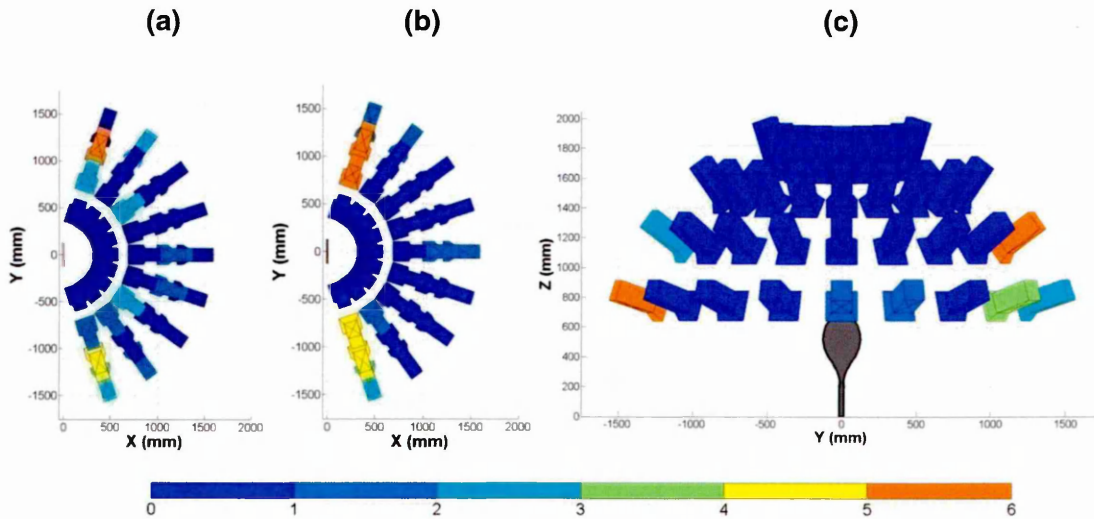


Figure 7.10 Translation errors in the (a) X, (b) Y and (c) Z directions for camera pose estimates made using a semi-circular fully calibrated set configuration containing 36 poses.

Figure 7.11 (a), (b) and (c) shows translation error in the X, Y and Z directions respectively for camera pose estimates obtained using a semi-circular configuration consisting of 27 poses, arranged in three tiers. Using this configuration, 96 % of pose estimates had translational errors of less than 15 mm in all directions. Of note is the comparatively higher translation error of the pose estimate second left on the lower tier. An acceptable solution for this view was not found after 100 candidate relative pose trials. This was reflected in the value of its optimised RMS ETE value of 1.4 pixels, which is around 7 times larger than the RMS ETE of 0.19 pixels across the set once

this view was removed. Only estimates with an RMS ETE less than 0.5 were accepted as solutions. Mean convergence time was 1.5 seconds using this set (Table 7.2).

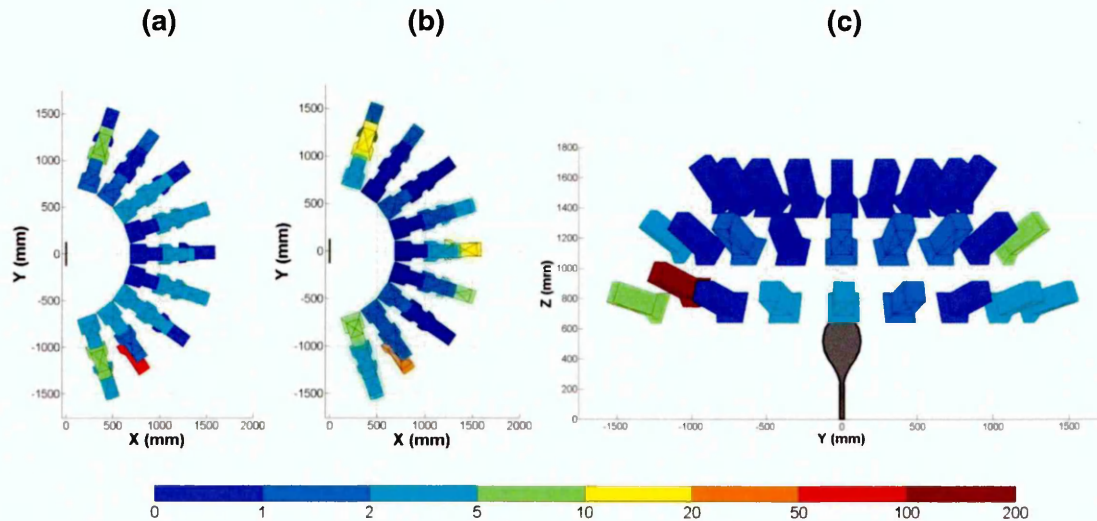


Figure 7.11 Translation errors in the (a) X, (b) Y and (c) Z directions for camera pose estimates made using a semi-circular fully calibrated set configuration containing 27 poses.

Figure 7.12 (a), (b) and (c) shows translation error in the X, Y and Z directions respectively for camera pose estimates obtained using a semi-circular configuration consisting of 18 poses, arranged in two tiers. The accuracy of the camera pose estimates clearly degrades with decreasing number of poses in a semi-circular fully calibrated set. With this set, 67 % of pose estimates displayed translation errors less than 15 mm in all directions. The largest camera pose translation errors were associated with the outer views on the top tier. Using this set mean convergence time was 0.9 seconds (Table 7.2).

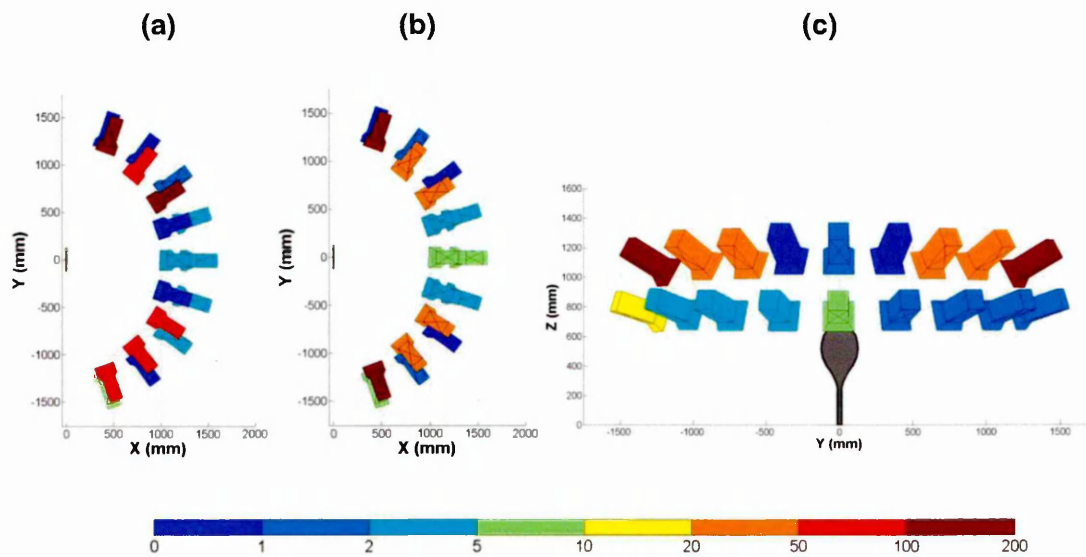


Figure 7.12 Translation errors in the (a) X, (b) Y and (c) Z directions for camera pose estimates made using a semi-circular fully calibrated set configuration containing 18 poses.

Figure 7.13 (a), (b) and (c) shows translation error in the X, Y and Z directions respectively for camera pose estimates obtained using a semi-circular configuration consisting of 9 poses, arranged in a single tier. Using this set, 67, 22 and 11 % of pose estimates displayed translation errors less than 15 mm in the X, Y and Z directions respectively.

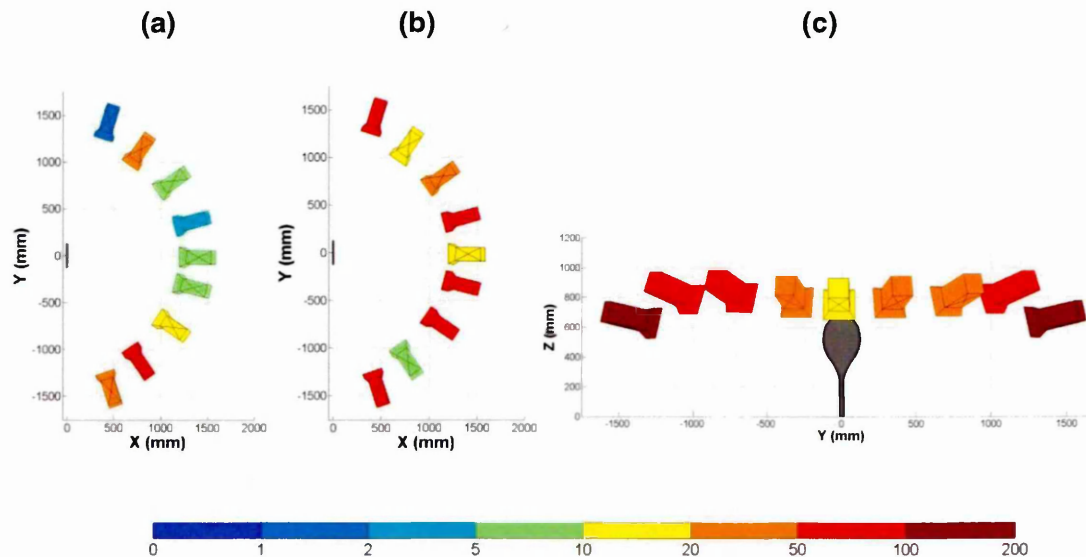


Figure 7.13 Translation errors in the (a) X, (b) Y and (c) Z directions for camera pose estimates made using a semi-circular fully calibrated set configuration containing 9 poses.

Table 7.2 presents mean values for rotation error about the axis' of X, Y and Z for all the semi-circular configurations tested. Camera pose estimate rotation accuracy decreased with the number of poses in the configuration. The semi-circular configurations consisting of 36 and 27 views estimated poses to within less than half a degree of accuracy. Furthermore, once the camera pose estimate whose optimised RMS ETE did not fall below the sub-pixel threshold using the semi-circular configuration with 27 views had been removed, mean rotation error values reduced by more than 50% about all axes.

Table 7.2 Rotation error of pose estimates made using semi-circular configurations consisting of 36, 27, 18 and 9 poses. Computation time for the optimisation to converge to a solution using a 3.4 GHz machine is also shown.

Rotation error (°)	Semi-circular configuration			
	36 pose set	27 pose set	18 pose set	9 pose set
X	0.05 ± 0.08	0.29 ± 0.97	1.40 ± 2.14	2.47 ± 1.70
Y	0.02 ± 0.02	0.05 ± 0.18	0.23 ± 0.29	0.59 ± 0.33
Z	0.04 ± 0.04	0.22 ± 0.61	1.22 ± 1.60	1.72 ± 1.29
Computation time (seconds)	1.9	1.5	0.9	0.5

7.6.3 Reconstructing the racket face plane

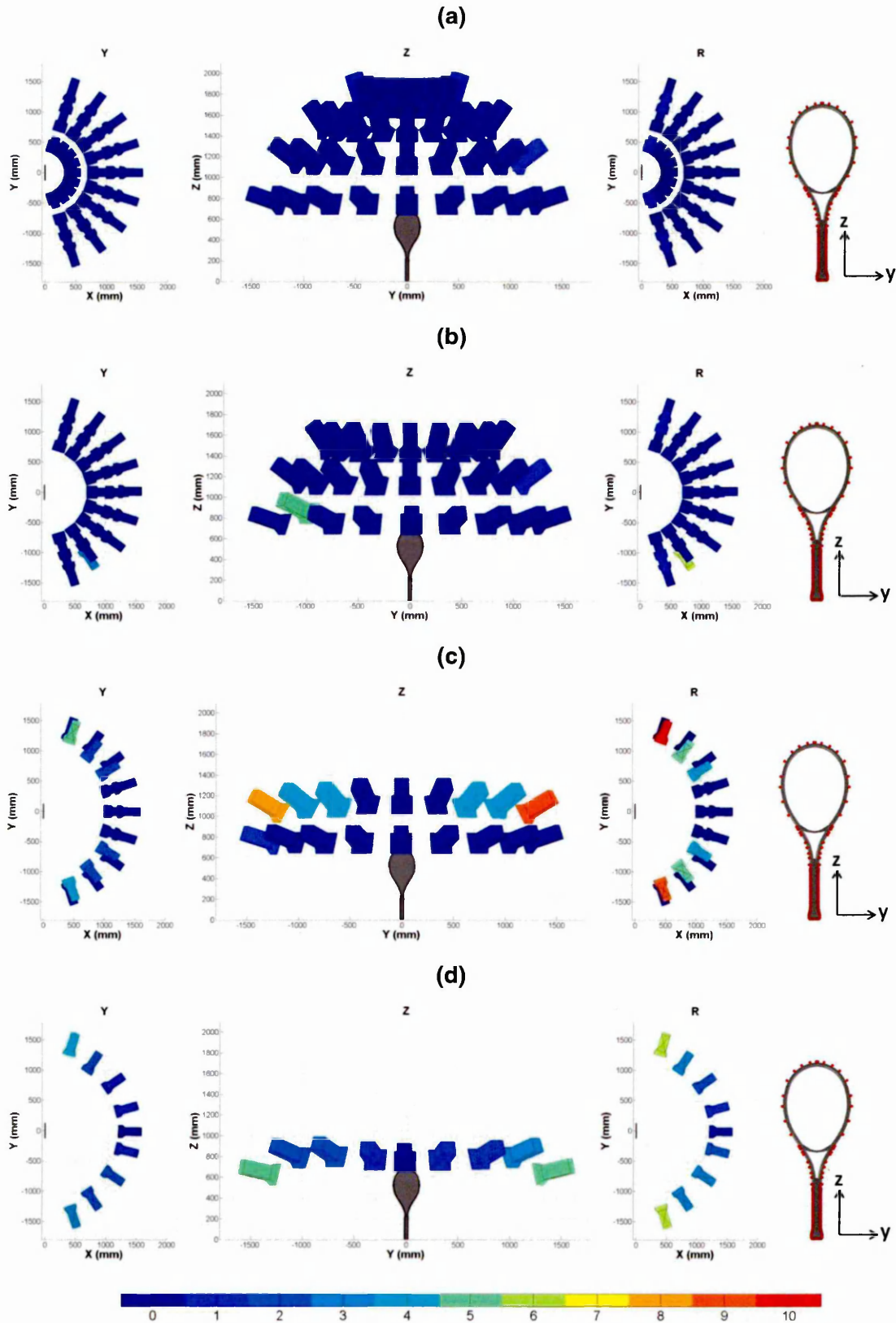


Figure 7.14 RMSE for reconstructed coordinates on the racket face plane in the X, Y and R directions using semi-circular fully calibrated sets with (a) 36, (b) 27, (c) 18 and (d) 9 poses.

Figure 7.14 (a), (b), (c) and (d) show root mean square reconstruction error (RMSE) for

coordinates on the racket face plane in the Y, Z and resultant (R) directions, using semi-circular fully calibrated sets consisting of 36, 27, 18 and 9 poses respectively.

Reconstruction error increased as the number of poses in the fully calibrated set decreased. The largest errors occurred for the outer views when using the sets consisting of 18 and 9 poses. However, all reconstruction errors were below the accuracy criteria of 15 mm outlined in the literature review (Chapter 2, page 40). There is a marked decrease in the magnitude of camera pose estimate error (Figure 7.10 to Figure 7.13) and reconstruction error (Figure 7.14). This is because camera pose estimates were located approximately 1.7 m from the local origin whereas the reconstructed points on the racket face plane were closer to the local origin (668 mm at the furthest point). Figure 7.15 (a), (b) and (c) show that resultant reconstruction error increases with camera pose estimate translation error in the X, Y and Z directions respectively.

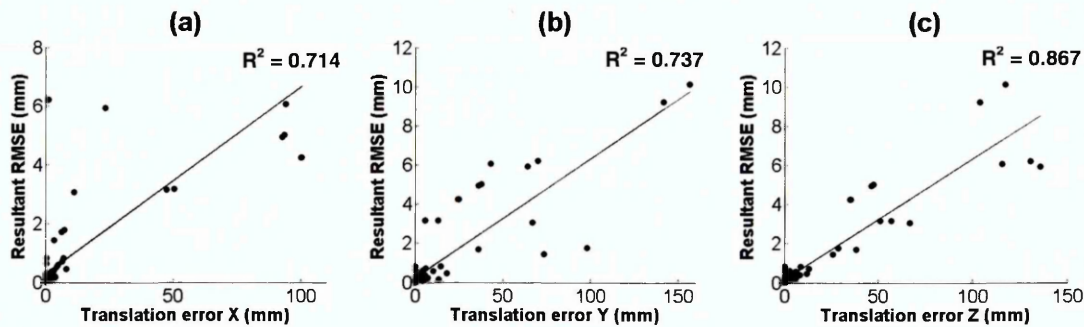


Figure 7.15 Resultant RMSE for coordinates on the racket face plane increases with camera pose estimate translation error in the (a) X, (b) Y and (c) Z direction.

Figure 7.16 (a), (b) and (c) show that resultant reconstruction error increases with camera pose estimate rotation error in the X, Y and Z directions respectively. Therefore, camera pose estimate accuracy directly influences reconstruction accuracy of coordinates on the racket face plane and the accuracy of 3D racket pose estimation is likely to decrease with distance from the local origin.

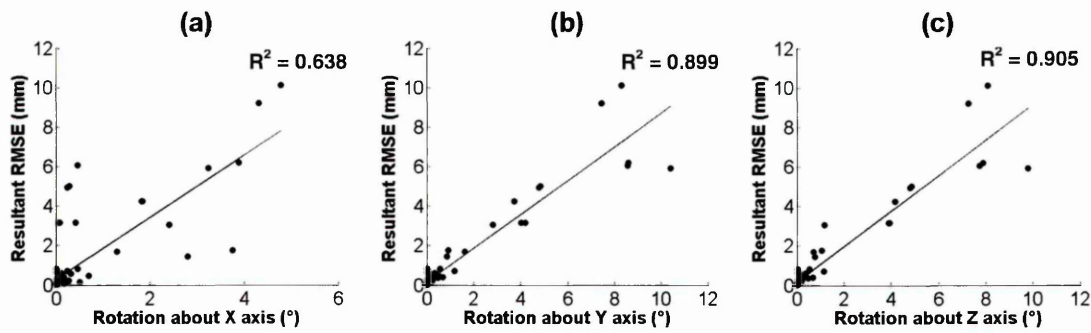


Figure 7.16 Resultant RMSE for coordinates on the racket face plane increases with camera pose estimate rotation error in the (a) X, (b) Y and (c) Z direction.

7.6.4 Quantification of removing a pose from a fully calibrated set

The semi-circular fully calibrated set with 27 views was used for this study. In practice, this set would be easier to replicate in the laboratory, which is an important consideration for the work in chapter 8. This is because the top tier of the set with 36 poses would be difficult to replicate; placing cameras this high would not allow for adequate checkerboard coverage thus accurate calibration parameters. A workaround could be to tilt the racket forwards however, moving the racket and camera may introduce more error.

Firstly, this study involved reconstructing coordinates on the racket face plane for each of the ten views (which were not part of the set) using the fully calibrated set consisting of 27 poses. Secondly, reconstruction was then performed using 27 different configurations with 26 racket silhouette views i.e. a different view was removed from the fully calibrated set for each reconstruction.

Reconstruction of the racket face plane differed by 0.03, 0.01 and 0.01 % in the Y, Z and resultant (R) directions respectively, using camera pose estimates obtained from fully calibrated sets consisting of 27 and 26 views. Therefore, it is clear that removal of one view from a fully calibrated set (required for the take-one-out validation method) does not affect the accuracy of reconstructed coordinates on the racket face plane. This confirms that the results reported in this chapter are valid for the synthetic pose configurations tested and the take-one-out method can be used for validation in the laboratory.

7.7 Discussion

In this chapter a take-one-out validation method was used. It involved removing each silhouette from the set in turn and estimating its relative pose using the method. The method works by adjusting candidate relative poses in order to minimise the ETE

distance (Figure 7.5), using a fully calibrated set. The fully calibrated set configurations consisted of tiers of coplanar camera poses. Mündermann et al. (2005) found that coplanar configurations produced accurate visual hulls of human participants who were not arbitrarily orientated with respect to the cameras. The racket model was strategically placed (upright on its butt) in the Blender scene to allow for a wide range of views to be generated. If the racket had been laid on its side, a range of views that covered the entire shape of the racket would have been more difficult to obtain.

Testing of a dome shaped fully calibrated set configuration revealed that the method was unable to estimate camera poses associated with side silhouette views of the racket to the same degree of accuracy as silhouette views that included the majority of the racket face plane. These results are specific to a racket and occurred because side views have a notably different shape compared to views that include the racket face plane. Wong (2001) noted that pose estimate accuracy decreases with the size of the region occupied by a silhouette in the image. At a resolution of 1280 x 800, a silhouette view of the entire racket face plane was comprised of 50% more white pixels compared to a side view of a racket (Figure 7.7). Therefore, side views were removed from fully calibrated sets in order to improve camera pose estimate accuracy, this was demonstrated using a 72 pose set in Figure 7.8. Future developments of the method could calculate the percentage of white pixels relative to black pixels in a silhouette before deciding to run an optimisation. Silhouettes with $\geq 50\%$ less white pixels than views that contain the entire racket face plane could be discarded.

Since a visual hull of a racket is not required for the single view fitting method developed in this project, it was viable to use semi-circular fully calibrated sets which viewed one side of the racket face plane. For an accurate racket visual hull, camera poses would be required at both sides of the racket face plane to provide enough information about the volume of space that the racket does not occupy. Wong (2001) used 11 silhouette images of a small clay sculpture captured in an incomplete circular coplanar configuration to estimate the pose of additional silhouette views.

Four and three tiered semi-circular fully calibrated set configurations consisting of 36 and 27 racket views respectively produced the most accurate pose estimates. On average, only 5 and 14% of poses were not estimated to within 10 mm of the ground truth using the 36 and 27 pose configurations respectively. This was a good result considering the time for the Levenberg-Marquardt method to converge decreased from 2.7 seconds using the 72 pose set to 1.9 and 1.5 seconds using the 36 and 27 pose configurations respectively using a 3.4 GHz Intel Core i7 machine. As expected, pose

estimate accuracy decreased when 18 and 9 racket silhouette views were used. Price and Morrison (2007) also observed a decrement in pose estimate accuracy from a configuration consisting of 60 views to ones consisting of 24 and 6 views of a rigid particle.

In order to compare the method's ability to estimate racket position with the marker-based methods (Elliott, Marsh and Blanksby 1986, Elliott, Marsh and Overheu 1989a and Elliott, Marsh and Overheu 1989b, Choppin et al. 2007a), pose estimates were used to reconstruct 3D points on the racket face plane. Choppin, Goodwill and Haake (2011) reported that, on average, racket marker position could be reconstructed to within ± 2.5 mm. This included human error from manual marker selection on the racket frame and systematic error from the calibration. In this chapter coordinates on the racket face plane were reconstructed to within ± 0.15 , ± 0.33 , ± 1.18 and ± 2.27 mm using semi-circular fully calibrated set configurations consisting of 36, 27, 18 and 9 racket silhouette views respectively. However, the scenario tested in this chapter was more constrained; cameras were approximately 1.7 m from the racket which was located at the local origin. Choppin, Goodwill and Haake (2011) reconstructed 3D coordinates on the racket frame during baseline rallies using cameras placed at the net i.e. approximately 14 m away from the racket.

7.7.1 Conclusion

This chapter has presented the initial step in the validation of a method for markerless estimation of tennis racket position. The approach adopted a take-one-out method which involved removing each silhouette from the a fully calibrated set in turn and estimating its pose. It was found that:

- Semi-circular fully calibrated configurations (which viewed one side of the racket plane) were capable of estimating camera poses to a comparable degree of accuracy to that of a full dome shaped configuration, with a shorter convergence time.
- Racket position estimation accuracy decreased with camera pose estimate accuracy.
- A semi-circular fully calibrated set containing 27 racket silhouette views arranged in three tiers positioned 1.7 meters from the racket is most appropriate to be replicated in the laboratory.

In chapter 8, a laboratory-based fully calibrated set is created using the Vision Research Phantom Miro M110 cameras.

8 Validation of racket pose estimates using a fully calibrated set created in the laboratory

8.1 Introduction

This chapter adapted the calibration methods outlined in chapters 3, 4 and 5 to create a laboratory-based fully calibrated set (LFCS) i.e. real calibration parameters and silhouettes. The configuration of the LFCS was based on findings from chapter 7. Camera poses estimated using the LFCS were used to reconstruct 3D racket position. Therefore, this chapter assesses the total error from the calibration parameters and silhouette extraction on 3D racket position estimation (Figure 8.1).

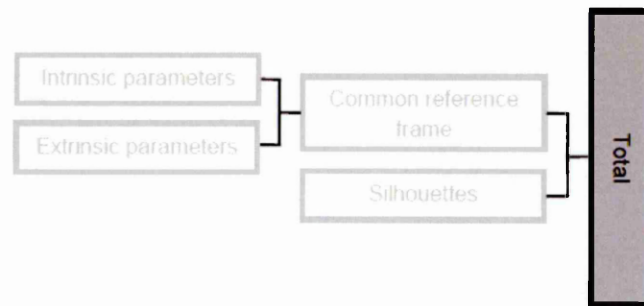


Figure 8.1 Chapter 8 calculated the total error resulting from calibration parameters and silhouette extraction on the ability of a view fitting method to estimate 3D racket position.

8.2 Aim and objectives

The aim of this chapter was to validate 3D racket position obtained using camera pose estimates obtained from a LFCS.

Objectives:

1. To create a LFCS using calibration methods outlined in chapters 3, 4 and 5.
2. To assess the accuracy of the LFCS by measuring its consistency.
3. To use the camera pose estimates obtained with the LFCS to reconstruct 3D racket position.

8.3 Creating the LFCS

8.3.1 Method

Corresponding checkerboard images were captured using two black and white high-speed Vision Research Phantom Miro M110 cameras. The images were passed to Bouguet's (2010) camera calibration toolbox for MATLAB in order to compute the intrinsic and extrinsic parameters (see chapter 4). Twenty one stereo calibrations were

performed and a common local reference frame was defined using the methods presented in chapter 5 (Figure 8.2). The master camera remained untouched and was positioned at approximately -10° relative to the racket stringbed normal so that it did not obstruct any of the slave camera positions. The slave camera was positioned at 7 different angles with respect to the racket stringbed normal i.e. 0° in Figure 8.2. The angles were measured approximately using a protractor and a piece of string that was 1.7 m in length. The slave camera angles ranged from -60° to 60° in 20° increments.

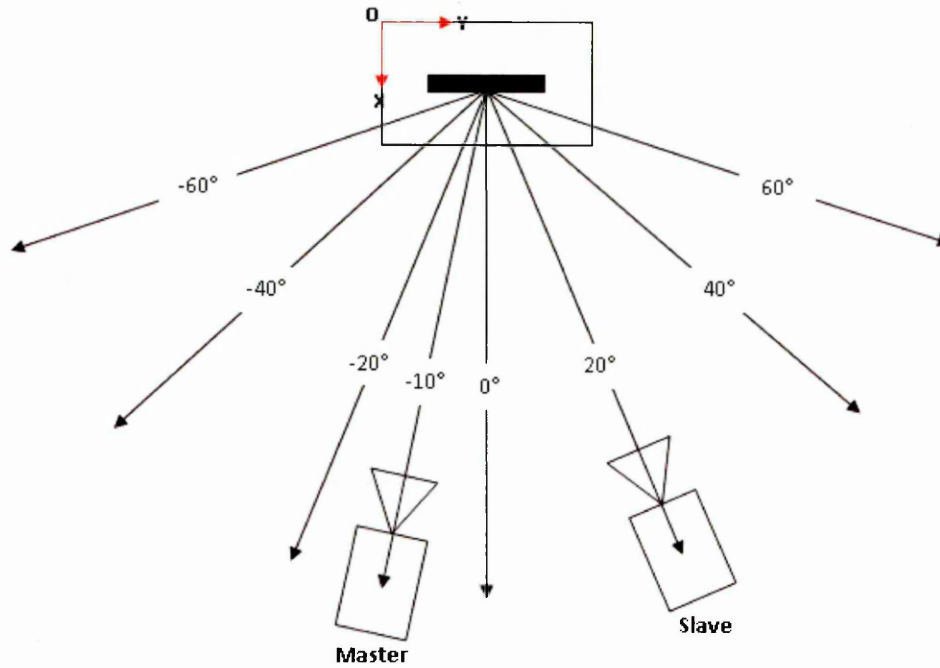


Figure 8.2 Approximate angles relative to the racket stringbed normal (0°) for the 21 slave camera poses.

The angle of the synthetic poses ranged from -80° to 80° in 20° increments with respect to the racket model face plane normal for 27 pose set (Figure 7.11, page 105). However, these extreme angles (-80° and 80°) were not used for the LFCS for two reasons:

1. Adequate checkerboard coverage in both views was not possible at these angles therefore the calibration would not be accurate.
2. There was not enough space in the laboratory to position the slave cameras at -80° and 80° with respect to the racket model face plane normal.

The influence of removing these views was checked using the synthetic data and the differences in 3D racket position estimation was negligible. At each angle, the slave camera was positioned at 3 different heights; 1.15, 1.55 and 1.85 m from the floor,

forming the 21 different positions. The camera heights were the same as those used for the synthetic three tiered semi-circular configuration with 27 poses in chapter 7.

At each of the 21 slave camera positions, images of the racket were captured so that silhouettes could be extracted. Image thresholding was used to extract the racket silhouettes and polygonal silhouette boundaries were segmented (see section 2.3.3, page 20-23). Silhouette extraction quality was assessed visually by plotting the extracted boundary pixels onto the original greyscale image. The threshold value was adjusted if the silhouette boundary did not match the racket shape in the original image.

8.3.2 Results

Table 8.1 shows the mean \pm standard deviation for the intrinsic parameters (f_x , f_y , c_x , c_y) for the master and slave camera over the 21 stereo calibrations. For both cameras, focal length was less repeatable than principal point. However, there was no significant difference ($p > .05$) between the mean intrinsic parameter values in Table 8.1 and those in Table 4.1, given that the same camera and lens settings were used.

Table 8.1 Mean \pm standard deviation for the intrinsic parameters (f_x , f_y , c_x , c_y) for the master and slave cameras over the 21 stereo calibrations.

Intrinsics (pixels)	Master camera	Slave camera
f_x	1804.30 \pm 1.72	1781.60 \pm 1.30
f_y	1802.20 \pm 2.31	1779.50 \pm 1.26
c_x	642.64 \pm 0.75	651.53 \pm 0.72
c_y	440.76 \pm 0.84	399.30 \pm 0.93

Figure 8.3 is a schematic showing the LFCS. The relative location of the master camera (red), 21 slaves (blue) and the racket in the common local reference frame is shown. Mean slave camera distance from the racket was 1.5 \pm 0.76 m.

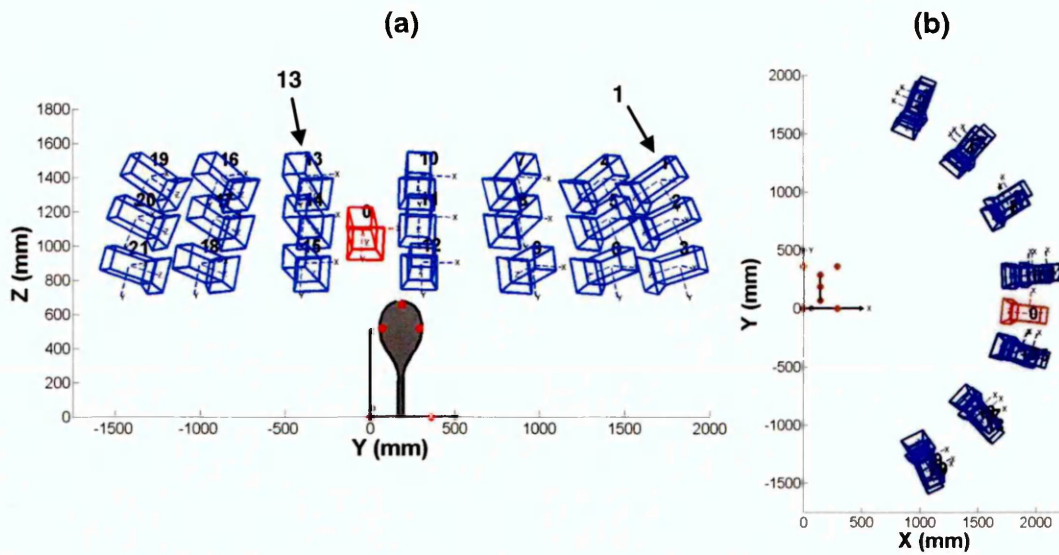


Figure 8.3 Schematic showing (a) a front view and (b) a top view of the master camera pose (red), the fully calibrated set of 21 slave camera poses (blue) and the racket in a common reference frame.

8.4 Assessing the accuracy of the LFCS

This section assessed the accuracy of the LFSC created in the previous section by using the computed parameters to calculate the RMS ETE across the set. The accuracy of the LFSC affects how close the projected tangent lines are to being tangential to the racket's silhouette. With exact parameters and silhouettes, the epipolar tangency lines will be exactly tangent to the silhouettes (see section 6.9.1, page 86-88). The RMS ETE for the LFCS was compared with a synthetic replica.

8.4.1 Synthetic data generation

The synthetic set was exactly the same as the LFCS in Figure 8.3. Twenty one synthetic camera poses were added to the Blender scene. The intrinsic parameters were based on those for the slave camera in Table 8.1. For each pose, an effective focal distance was used and the principal point was located at the centre of each silhouette image. The standard deviation in focal distance across the 21 poses was ± 1.28 pixels. This corresponded to the mean standard deviation across f_y and f_x in Table 8.1. Silhouettes associated with the racket model (created in Blender Figure 6.8) were rendered at a resolution of 1280 x 800 pixels.

8.4.2 Method

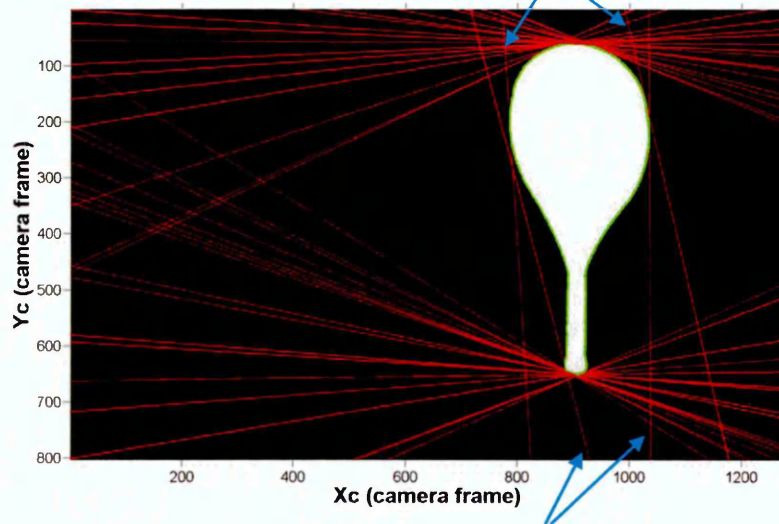
- The accuracy of the LFCS was quantified by using the computed parameters to calculate the RMS ETE. The LFCS's accuracy was assessed using the 7 control points to define the local common reference frame (P_1 to P_7 in Figure 5.6).

- The RMS ETE values for the LFCS were compared with the synthetic data created in section 8.4.1. This was measured using three levels of syntheticity:
 1. Fully synthetic (see section 6.9.1, page 86).
 2. Synthetic parameters and rendered silhouettes. This involved using Blender to render the 21 silhouette views associated with the model at a resolution of 1280x800 pixels.
 3. LFCS parameters and synthetic silhouettes. This involved using the calibration parameters associated with the LFCS to project the 130 outer 3D points of the racket model, forming exact projected silhouette boundaries. The master camera was not included in the LFCS because the configuration was based on the synthetic set containing 27 evenly spread views (Figure 7.11, page 105).

8.4.3 Results

Figure 8.4 shows the projected epipolar tangents derived using the LFCS calibration parameters and silhouettes. The silhouette in Figure 8.4 corresponds to camera pose 8 in Figure 8.3 (a). The majority of tangents touch the tip and butt of the racket. However, four of the tangents touch the side of the racket. These four tangents are a result of when camera pose 8 is paired with poses 7 and 9 in Figure 8.3 (a). These two poses are located above and below (Z axis direction) pose 8 respectively. All the other poses in the LFCS are either side (Y axis direction) of pose 8.

Epipolar lines converging above the racket (pose 8 paired with pose 7 Figure 8.3 (a)).



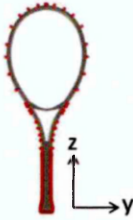
Epipolar lines converging below the racket (pose 8 paired with pose 9 Figure 8.3 (a)).

Figure 8.4 Projected epipolar tangent lines on the image plane of camera pose 8 in Figure 8.3 (a) derived from the all camera poses. The data has some error therefore the lines are not exactly tangential.

Table 8.2 presents the RMS ETE for the synthetic and real data. The racket in Table 8.2 shows the 130 outer 3D points of the racket model used to obtain exact projected silhouette boundaries. As expected the RMS ETE is zero when synthetic parameters and synthetic silhouette boundaries were used. Using synthetic parameters and rendered silhouettes at a resolution of 1280 x 800 pixels, increased the RMS ETE to 0.22 pixels. Exact polygonal silhouette projections generated using LFCS calibration parameters provided an RMS ETE of 0.19. The RMS ETE across LFCS (real parameters and silhouettes) was 0.41 pixels.

Synthetic parameters and rendered silhouettes were more inconsistent (RMS ETE = 0.22) than the real parameters and the synthetic silhouettes (RMS ETE = 0.19), demonstrating the accuracy of the calibration method. Discrete pixel rendering and silhouette extraction is a marginally greater contributor to the inconsistency of the LFCS than the calibration parameters. Higher resolution silhouettes may improve the consistency of the LFCS (see section 6.9.3, page 89).

Table 8.2 RMS ETE (pixels) values for the 21 pose synthetic set using synthetic parameters with synthetic silhouettes, synthetic parameters with rendered silhouettes and real parameters with synthetic silhouettes.

	Synthetic parameters synthetic silhouettes	Real parameters synthetic silhouettes	Synthetic parameters rendered silhouettes	Real parameters rendered silhouettes	
RMS ETE	0.00	0.19	0.22	0.41	

8.5 Racket position validation

In this section camera pose estimates were obtained using the take-one-out method and then used to reconstruct 3D coordinates on the racket frame surface. Chapter 7 showed that the take-one-out method is suitable to be used for the validation of camera pose estimates; changing the configuration of a set by removal of a single pose did not affect pose estimate accuracy (see section 7.6.4, page 110).

8.5.1 Method

Camera pose estimates were obtained using:

- LFCS parameters and synthetic silhouettes.
- Synthetic parameters and rendered silhouettes.
- The LFCS.

A maximum of 100 relative candidate poses were adjusted for each camera pose estimation and an ETE threshold of 0.5 pixels was used (see section 7.4, page 94-98). Camera pose estimates were compared with the corresponding poses in the LFCS, which were deemed as ground truth. For the fully calibrated set that used synthetic parameters and rendered silhouettes, pose estimates were compared with the corresponding synthetic poses.

The camera pose estimates were then used to reconstruct 106 3D coordinates on the racket frame surface. Reconstructed 3D coordinates were compared with coordinates extracted from the racket laser scan, which were deemed as gold standard. An account of how the scan was performed is detailed in section 5.3 (page 65-69). This section of the study was carried out for two reasons:

1. To compare data with previous methods identified in the literature review that estimated 3D racket position.
2. To inform potential users (players, coaches and manufacturers) of the accuracy and potential of the method.

As the 3D coordinates extracted from the laser scan were not coplanar i.e. not on the racket face plane, stereo triangulation was used for the reconstruction. Exact pixel projections of the 3D coordinates were obtained using the calibration parameters. This allowed for triangulation using the master camera (Figure 8.3), which was deemed as ground truth and each camera pose estimate. The master camera always had the same pose therefore reconstruction accuracy using the camera pose estimates was quantified.

8.5.2 Results

Figure 8.5 illustrates a sample reconstruction using the LFCS. Figure 8.5 (a) and (b) show views of the racket as seen by camera 1 and camera 13 in Figure 8.3 (a) respectively. Pixel projections of the 3D coordinates have been plotted on the images. The green dots are ground truth while the red dots are projections of the reconstructed coordinates. The visible red dots in Figure 8.5 (a) indicate lower reconstruction accuracy compared with Figure 8.5 (b). In Figure 8.5 (a), RMSE for reconstruction in the X, Y and Z directions was 2.96, 1.28 and 4.05 mm respectively. In Figure 8.5 (b), RMSE for reconstruction in the X, Y and Z directions was 1.13, 0.20 and 0.46 mm respectively. Figure 8.5 (a) and (b) are extreme examples; camera pose estimates corresponding to views 1 and 13 had resultant translation errors of 214.28 and 9.87 mm respectively.

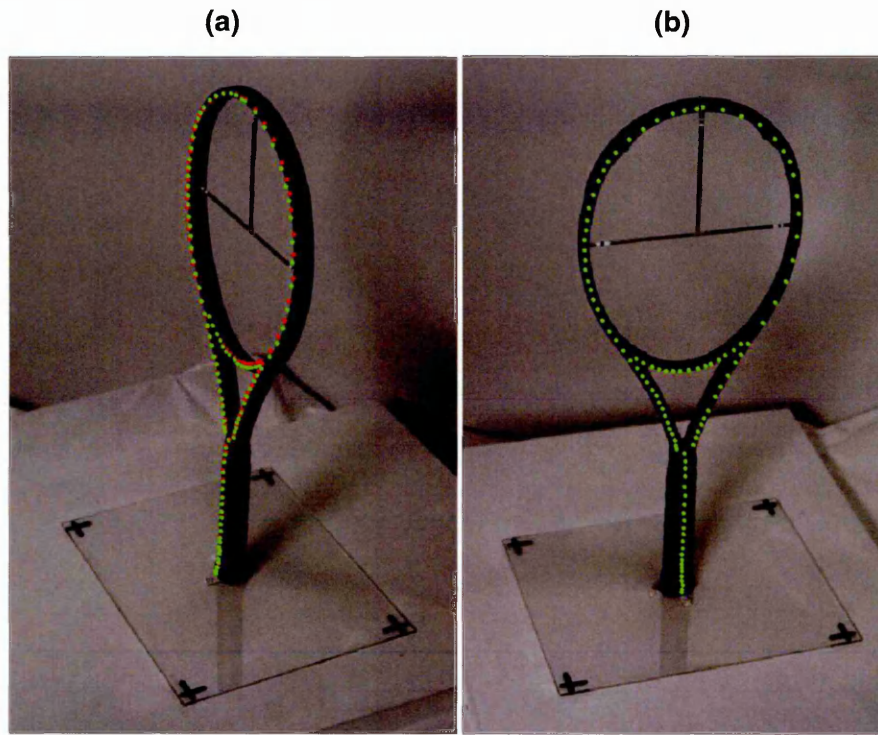


Figure 8.5 Racket views corresponding to (a) camera pose 1 and (b) camera pose 13 from Figure 8.3. The green dots are exact projections of 3D points obtained from the racket scan and the red dots are projections of 3D coordinates reconstructed using pose estimates.

Figure 8.6 shows RMSE (mm) for reconstruction in the X, Y and Z directions of 3D coordinates on the racket frame surface. Reconstruction error in all directions was lowest when the LFCS parameters and synthetic silhouettes were used (Figure 8.6 (a)). In Figure 8.6 (a), mean reconstruction error was 0.42, 0.06 and 0.21 mm in the X, Y and Z directions respectively. Reconstruction error increased when synthetic poses and rendered silhouettes were used (Figure 8.6 (b)). In Figure 8.6 (b), mean reconstruction error was 0.76, 0.15 and 0.48 mm in the X, Y and Z directions respectively.

In Figure 8.6 (c), 3D racket position was estimated using the LFCS. Camera pose estimate 1 produced the largest reconstruction errors of 2.9, 1.3 and 4.1 mm in the X, Y and Z directions respectively. The camera pose associated with this view was located at approximately 60° with respect to the racket face plane normal. In general, racket position accuracy decreases as the angle of the camera pose estimate with respect to the racket face plane normal increases. Camera pose estimate 5 produced the lowest reconstruction errors of 0.1, 0.1 and 0.3 mm in the X, Y and Z directions respectively. Mean reconstruction error was 1.46, 0.29 and 1.03 mm in the X, Y and Z directions respectively, across all the camera pose estimates.

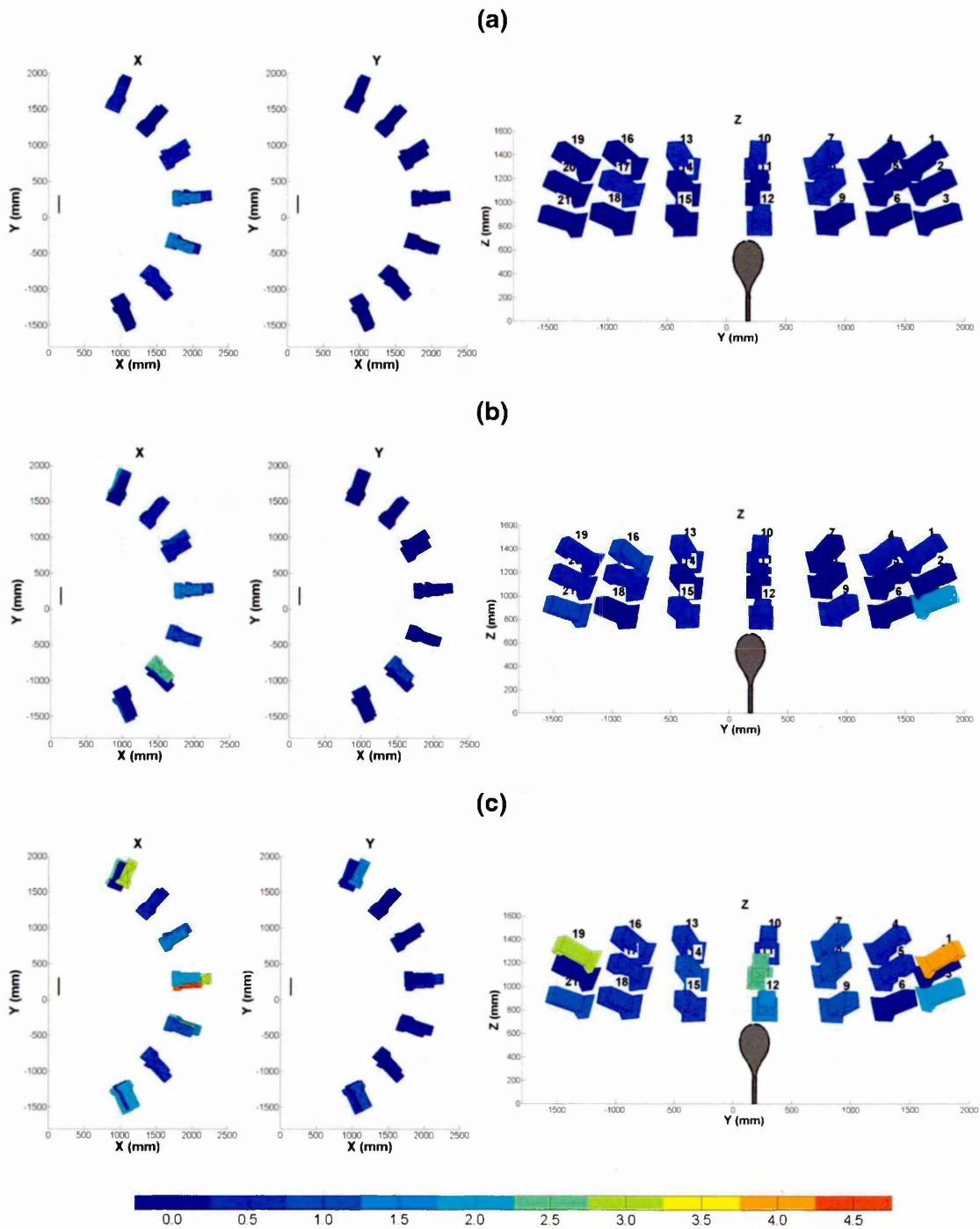


Figure 8.6 RMSE (mm) for reconstruction in the X, Y and Z directions of 3D points on the racket frame surface using camera pose estimates obtained from (a) LFCS parameters and synthetic silhouettes, (b) synthetic poses and rendered silhouettes and (c) the LFCS.

8.6 Discussion

In this chapter, a LFCS based on pose configuration findings from chapter 7 was created. The LFCS used the calibration methods developed in chapters 3, 4 and 5 to obtain 21 camera poses and a common local reference frame including a racket. It contained three tiers of seven camera poses in a semi-circular configuration (Figure

8.3). The accuracy of the LFCS was quantified by using the computed parameters to calculate the RMS ETE across the set. An RMS ETE value of 0.41 pixels was measured for the LFCS. Forbes (2007) reported an RMS ETE value of 0.4 pixels using silhouettes of stones and parameters obtained using a self-calibration technique. Forbes (2007) obtained this RMS ETE value for a fully calibrated set consisting of 90 silhouette views associated with different stones. Wong (2001) calculated an RMS ETE of 0.39 pixels across 192 control points on a calibration grid used to calculate parameters. Therefore, the accuracy of the LFCS created in this chapter was similar to those used in previous successful silhouette-based pose optimisation studies, despite that these comparisons are for different scenarios.

Use of synthetic data allowed the influence of silhouette segmentation and camera calibration on the RMS ETE to be quantified in isolation. It was found that silhouette segmentation was marginally, a larger contributor to the inconsistency of the LFCS. This agrees with Forbes (2007) who measured the RMS ETE across fully calibrated silhouette sets associated with stones by employing the same levels of syntheticity used in this chapter. Highlighting that accurate silhouette extraction is important for the accuracy of a view fitting method. Chapter 6 (section 6.9.3, page 89) showed that when converted to millimetres, the RMS ETE across a fully calibrated set increased as silhouette resolution decreased. Therefore, silhouettes should be obtained at the highest possible resolution in order to produce accurate fully calibrated sets. At these high resolutions, quality polygonal silhouette boundaries can be obtained using the efficient segmentation method outlined by Forbes (2007) (section 2.3.3, page 20-23). Which relies on low level image processing functions such as morphological opening and grey thresholding to obtain the binary silhouette image. The use of B-spline snakes (Wong 2001) to segment the boundaries is more complex and less efficient. Since manual interaction with a user-supplied input function is required to identify an initial contour resulting in slow convergence speeds because of large numbers of sample points (Brigger, Hoeg and Unser 2000). Furthermore, B-spline snakes are required to identify boundaries in noisy images. The racket silhouette images in this chapter were obtained under controlled conditions using a backlight to ensure background pixels were substantially brighter than foreground pixels, simplifying boundary extraction.

Camera pose estimates obtained using the LFCS and the two synthetic fully calibrated sets were used to estimate the position of the racket in 3D. This involved reconstructing a number of 3D coordinates on the surface of the racket frame which were validated against coordinates obtained from the racket scan. Reconstruction error of coordinates on the racket frame surface increased with camera pose estimate error (Figure 8.5),

this was also demonstrated in chapter 7 (section 7.6.3, page 108-110). This was because camera poses are further from the local common origin than the racket. Racket position estimation uncertainty is therefore likely to increase with the distance of the racket from the local origin. In real-play conditions, the local origin could be placed at the centre mark of the baseline so that it is close to the racket during serves and groundstrokes.

Figure 8.6 showed that the fully calibrated set consisting of synthetic parameters and rendered silhouettes estimated racket position less accurately compared to the set consisting of LFCS parameters and synthetic silhouettes. Therefore, discrete pixel rendering and silhouette extraction is a larger contributor to 3D racket position error than the calibration parameters, using this configuration. This finding highlights the accuracy of the calibration method used to obtain the LFCS and the need to extract accurate silhouettes.

Choppin, Goodwill and Haake (2011) assessed the accuracy of a marker-based racket tracking method by comparing reconstructed distances between markers to those measured directly on the frame. On average, Choppin, Goodwill and Haake (2011) reconstructed individual marker position on the racket frame to within ± 2.5 mm. Choppin et al. (2011) used reconstructed coordinates from high-speed video sequences to calculate racket velocity and angle. The reconstruction errors corresponded to an average angular error of $\pm 1^\circ$ and velocity error of $\pm 0.5 \text{ m s}^{-1}$. Using the LFCS, mean reconstruction error for coordinates on the racket frame surface was 1.46, 0.29 and 1.03 mm in the X, Y and Z directions respectively. This rounds to a resultant reconstruction error of ± 2 mm across all 21 camera pose estimates in Figure 8.6 (c). This is 13 mm less than the accuracy criteria of 15 ± 10 mm achieved by the markerless method developed by Corazza et al. (2010), which was used by Sheets et al. (2011) to measure tennis serve kinematics in practice conditions (Chapter 2, page 40). However, it is not possible to compare directly compare the results of this study with the values reported by Choppin, Goodwill and Haake (2011). Since, Choppin, Goodwill and Haake (2011) reconstructed markers on the racket in a much less constrained scenario i.e. from camera distances up to 14 m between the net and the baseline. In this study reconstruction was performed from approximately 1.7 m from the racket.

Of note is the relatively large reconstruction errors in the X and Z directions for camera poses associated with frontal racket silhouette views (Figure 8.6 (c)). One reason could be that the convergence angle between the master camera and these frontal slave

camera positions was low, making the calibration parameters less accurate. Low convergence angles have been linked with inaccurate estimations of depth in stereo calibrations (Choppin 2008, Kelley 2014).

Another reason could be that these frontal racket silhouette views have reflective symmetry, with a line of symmetry running along the long axis of the racket. Reflective symmetry is the simplest type symmetry and is apparent when one half of an object is the reflection of the other half. Rotational symmetry also exists, which is apparent if there is a centre point around which the object is rotated and looks the same. The number of positions in which the object looks exactly the same is referred to as the order of symmetry. Price and Morrison (2007) found that single view poses from irregular shaped (non-symmetrical) particles were more consistently estimated than that of regular shaped particles. The epipolar constraint can be considered as an envelope of intersecting planes that encloses an object's convex hull (Sinha et al. 2004). Therefore, an object with a symmetrical convex hull will be more prone to suboptimum alignment because there will be multiple candidate poses that are equally consistent but do not necessarily correspond to the correct solution. However, this would be more problematic for silhouettes associated with an object that has rotational symmetry e.g. a cone or sphere. This is because there would be multiple poses associated with the object that resemble the same convex hull which can effectively become "jammed" in the epipolar tangency envelope. Thus, causing the optimisation to become trapped in a local minimum, resulting in the camera pose parameters never being able to move toward the correct solution. For a racket, the two frontal views have reflective symmetry (side racket views also have reflective symmetry but are not included in this project (see chapter 7, page 101)) and the same convex hull therefore, the chance of the optimisation converging to a local minimum is much less. Price and Morrison (2007) contend the problem can be alleviated if lots of constraints are used i.e. using a fully calibrated set containing a large range of views of an object. Chapter 7 demonstrated that pose estimate accuracy decreased with the number of views in a set. Therefore, the fact a racket has reflective symmetry (when viewed from the front and back) is unlikely to influence the accuracy of the method in estimating 3D racket position. Furthermore, reflective symmetry can easily be accounted for by using a fully calibrated set that includes a range of views of the racket.

Figure 8.7 shows that the method works well for a non-symmetrical object, agreeing with Price and Morrison (2007). Using the same synthetic pose configuration, rendered silhouette views of an irregular plane (with the same height and width as the racket) were used to estimate its relative pose. Overall, mean reconstruction error for the

irregular plane was low (Figure 8.7). However, direct comparison between results in Figure 8.7 and those for the racket in Figure 8.6 (b) is not possible. A frontal silhouette view of the racket face plane is made up of 50% less white pixels with respect to black pixels compared with the same view of the irregular plane. Therefore, the lower error for the irregular plane is more likely due to the fact that more pixels are used.

Nevertheless, future research could focus on making the racket silhouette appear more irregular in the silhouette image in order to improve results using the method. An approach could be to paint sections of the frame with a particular contrasting colour or pattern. The main challenge surrounding this would be to not impede racket silhouette extraction. Another way could be to use grommets (holes through which the string fits) that protrude from the side of the racket, effectively changing the shape and irregularity of the rackets silhouette. However, this could add mass to the racket and increase the moment of inertia (MOI).

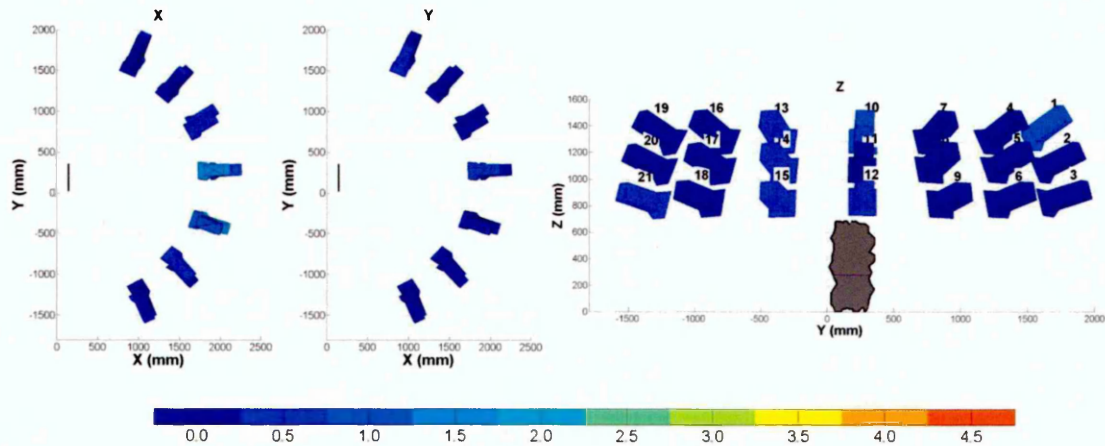


Figure 8.7 Camera pose estimate translation error in the X, Y and Z directions using rendered silhouettes of an irregular shaped plane with the same height and width as a standard size racket.

8.7 Conclusion

This chapter used a fully calibrated silhouette set of real data to estimate 3D racket position and found:

- A method to create a fully calibrated silhouette set consisting of 21 different camera poses in a semi-circular configuration was demonstrated.
- The consistency of the set was comparable with previous silhouette-based pose optimisation studies albeit for different scenarios. This highlights the accuracy of

the calibration method developed in this project and the importance of accurate silhouette extraction.

- Camera pose estimates were used to estimate 3D racket position using the laboratory-based fully calibrated set. This involved reconstructing 3D coordinates on the racket frame surface. Synthetic replication of the LFCS showed that discrete pixel rendering and silhouette extraction was a larger contributor to racket position error than the calibration parameters.
- Coordinates on the racket frame surface were reconstructed to within ± 2 mm using the LFCS (real parameters and silhouettes), which is less than the accuracy criteria of 15 mm. Therefore, under controlled conditions, the method estimates 3D racket position more accurately than the markerless method used by Sheets et al. (2011) to measure serve kinematics during practice play.
- Silhouettes associated with non-symmetrical objects provide marginally more accurate results using a view fitting method. For a non-symmetrical plane with the same height and width as the racket, this is likely due to there being more visible pixels in the silhouettes.

9 Estimating 3D racket position from different camera positions

9.1 Introduction

This chapter demonstrates the potential of the method to be applied in real-play conditions. Real-play conditions occur when players compete over a number of points in order to win sets and any coaching during this time is strictly prohibited. Racket position from a range of camera orientations between 0 and 90° and distances up to 14 m is estimated. Camera orientation is the angle of rotation about the optical axis in the plane aligned with the racket face plane and camera distance is measured with respect to the racket. Finally, racket motion expected to occur during a serve was simulated by rotating the racket about an axis 10.16 cm (4 inches) from the butt, in the plane perpendicular to the racket face plane. The method was then used to estimate 3D racket position across the series of silhouette images.

9.2 Aim and objectives

The aim of this chapter was to simulate real play conditions in Blender and apply the method to estimate 3D racket position.

Objectives:

1. To estimate 3D racket position over a range of camera orientations.
2. To investigate whether the orientation of camera poses within a fully calibrated set influences 3D racket position accuracy.
3. To estimate 3D racket position from realistic camera distances that are appropriate to be used during real play conditions.
4. To simulate racket motion expected to occur during a serve and apply the method to estimate racket position over a series of silhouette frames.
5. To make recommendations for the method to be applied in real play conditions.

9.3 The fully calibrated set

This chapter used synthetic camera parameters and rendered silhouettes. The fully calibrated set camera pose configuration was the same as the lab-based fully calibrated set (LFCS) in chapter 8 (Figure 8.3). For each camera pose, an effective focal distance was used and the principal point was located at the centre of each silhouette image. Variance in focal distance across the 21 camera poses was ± 1.2 pixels, corresponding to the mean standard deviation across f_y and f_x in Table 8.1. Lens distortion was not simulated for this intrinsic model.

9.4 Method

9.4.1 Camera orientation

Sixteen synthetic racket views were generated (Figure 9.1). These sixteen racket views were not part of the fully calibrated set described in section 9.3 and therefore will be referred to as unknown views. Figure 9.1 (a) and (b) show a front and top views of the unknown view configuration respectively. Mean camera distance from the racket was 1.55 ± 0.4 m. The configuration of the unknown views covered a similar area to that of the fully calibrated set in Figure 8.3.

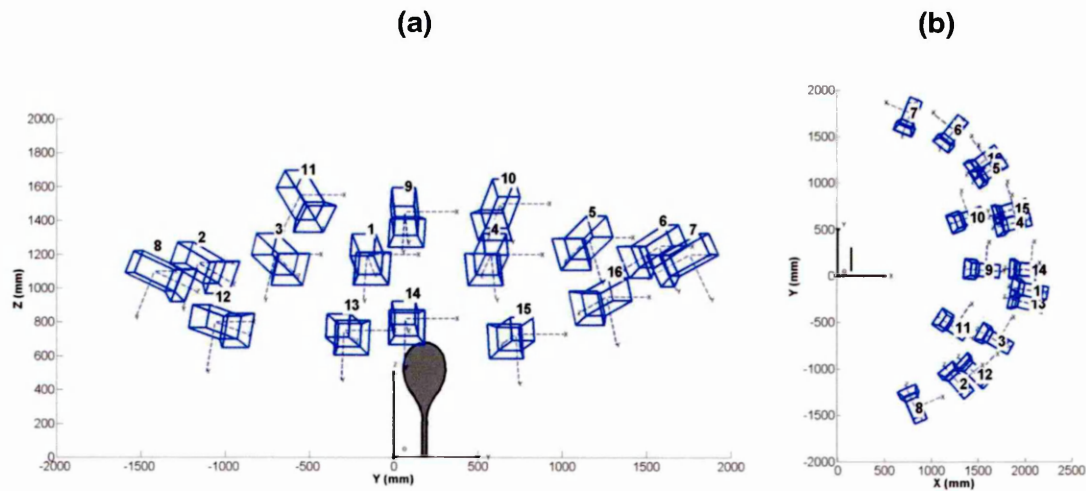


Figure 9.1 The sixteen synthetic unknown views. (a) shows a front view and (b) shows a top view.

The camera poses associated with the unknown views were assigned an effective focal length of 1780.6 pixels. This corresponded to the mean focal length across f_x and f_y for the slave camera in Table 8.1. Mean racket silhouette height (number of white pixels in the u image axis) was 72 and 71 % of the image for the unknown views and views in the fully calibrated set respectively.

Ten sets of the sixteen unknown views were tested; Figure 9.1 is an example of one of the sets. The ten sets differed by their rotation angle about the optical axes in the plane aligned with the racket face plane i.e. orientation about the Z-axis. Camera pose orientation was increased from 0 to 90° in 10° increments (Figure 9.2). In Figure 9.2 the rectangles represent the cameras image plane, which rotates while the racket model remains stationary. The blue, red and green rectangles show the image plane when it is rotated to orientations of 0, 50 and 90° respectively.

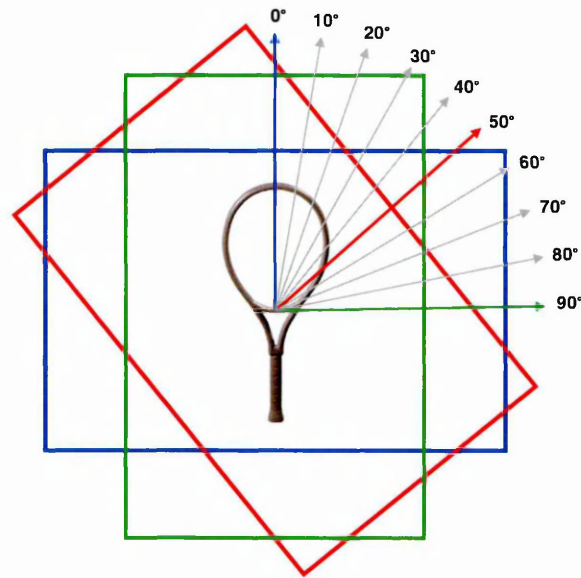


Figure 9.2 For each of the 16 unknown views in Figure 9.1 camera pose orientation was increased from 0 to 90° in 10 increments, forming ten sets of unknown views. The blue, red and green rectangles show the image plane at orientations of 0, 50 and 90° respectively.

Figure 9.3 (a), (b) and (c) show example racket silhouette images captured from camera pose 5 in Figure 9.1 using camera pose orientations of 0, 50 and 90° (Figure 9.2). The silhouette orientations were intended to cover a range of racket orientations expected to occur in tennis strokes.

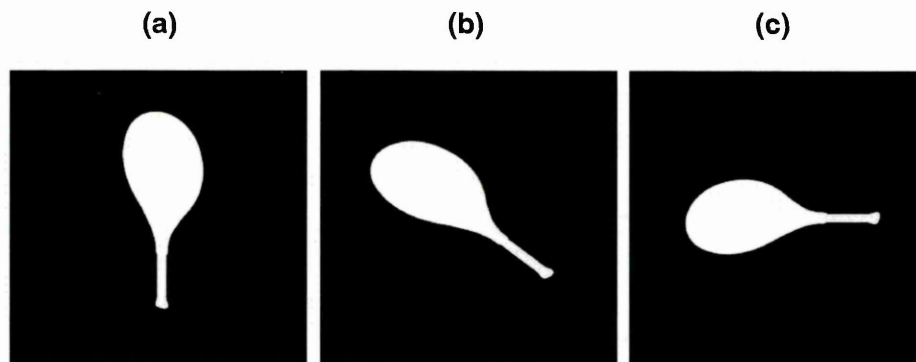


Figure 9.3 Racket silhouette images captured from view point 5 in Figure 9.1 at camera pose orientations of (a) 0, (b) 50 and (c) 90°.

The 3D racket position associated with the ten sets of unknown views was estimated using the synthetic fully calibrated set outlined in section 9.3. The fully calibrated set was then modified by randomly changing the orientation of the camera poses. Mean orientation of this fully calibrated set was $61.19 \pm 23.94^\circ$. The purpose was to investigate if 3D racket position accuracy could be influenced by using fully calibrated sets consisting of silhouettes that were orientated differently compared to the original

set. The 3D racket position associated with the ten sets of unknown views was then estimated using the randomly orientated fully calibrated set.

The same method for generating candidate relative poses as outlined in section 7.4 (page 94-98) was used. Racket position was estimated by reconstructing 130 coordinates on the racket face plane as per chapter 7. Reconstructed coordinates were validated against ground truth coordinates obtained from the 3D racket model mesh created in Blender (section 6.8, page 83).

9.4.2 Camera distance

The camera distances tested, were based on tennis court dimensions (Figure 9.4). All the camera poses were trained on the centre mark at the baseline since this is where serves and the majority of groundstroke take place during a tennis match (Choppin, Goodwill and Haake 2011).

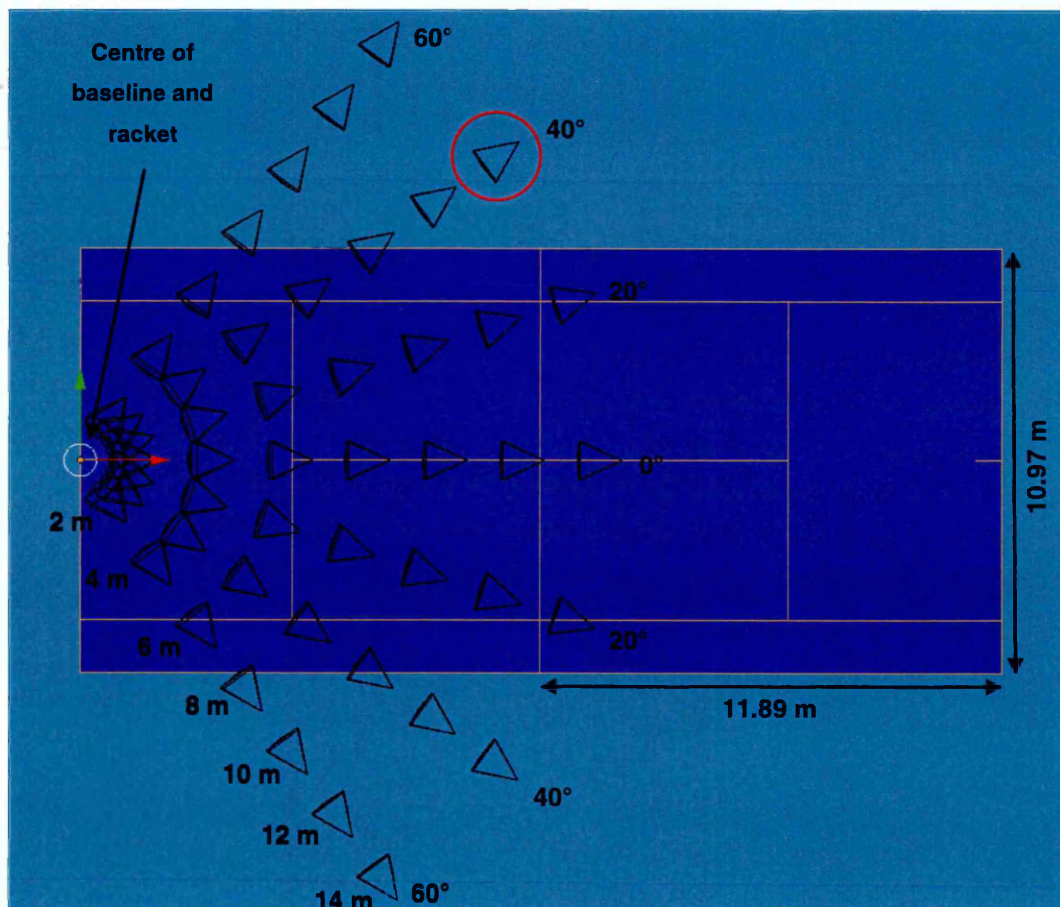


Figure 9.4 Camera distances tested and their relative positions on the court.

The camera distances were measured relative to the racket which was located at the centre of the baseline in Figure 9.4. The camera distances tested were 2, 4, 6, 8, 10, 12, and 14 m. At each distance there were seven camera poses. Across each set of

seven poses, camera angle with respect to the local X-axis (indicated by the red arrow in Figure 9.4) was increased from 0 to 60° in 20° increments.

Each camera distance was allocated a unique focal length. Focal length was determined by zooming-in until the racket appeared to be a similar size as the silhouettes in the fully calibrated set. Racket silhouette size was judged visually because is the most efficient method when filming in real play conditions rather than calculating the percentage of white pixels in each image. The fully calibrated set used in this section is outlined in section 9.3 i.e. all the poses were orientated upright, and candidate relative poses were determined using the method in section 7.4 (page 94).

Finally, a flat serve was simulated. The simulation was performed using the camera pose circled red in Figure 9.4. This pose was used because it lies outside the court and therefore would not be intrusive during real play conditions. Figure 9.5 shows the racket model relative to the centre mark of the baseline (local origin) in the view of the camera pose circled red in Figure 9.4. The butt and tip of the racket model were located at 0 and 0.7 m in the local Z direction respectively. The camera pose in Figure 9.4 was positioned at a height of 1.2 m which made it 0.5 m higher than the tip of the racket and at 40° with respect to the racket stringbed normal. It was orientated so that its optical axis intersected the centre of the image (red dot in Figure 9.5) i.e. it was looking head on to the racket. In real play conditions, the camera could be positioned so that it is slightly higher (0.5 m) than the racket. This is a simulation of how the method could be applied in real play conditions. However, other camera positions would need to be simulated to ascertain whether camera height i.e. if the camera looks up or down with respect to the racket, influences the accuracy of the method.

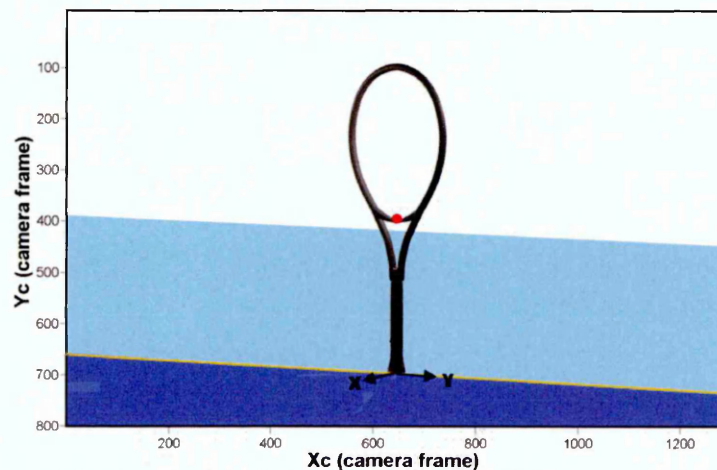


Figure 9.5 View from of the camera pose circled red in Figure 9.4. The red dot indicates where the camera view was centred i.e. midway up the racket.

To simulate racket motion expected to occur during a serve, the racket model was rotated about an axis 10.16 cm (4 inches) from the butt in the plane perpendicular to the racket face plane. This is the location of the axis of rotation used to define the swing weight of a tennis racket in laboratory conditions (Cross and Bower 2006). The racket was rotated about this axis between -40° to 30° in 2° increments in order to cover a range of racket positions important for racket-ball impact during a serve (Figure 9.6 (a)). For a serve, impact occurs above the head when the racket is orientated upright (Elliott, Marsh and Blanksby 1986, Kotze, Mitchell and Rothberg 2000, Abrams et al. 2011). In Figure 9.6 (a) this would be when the racket is orientated upright at 0° .

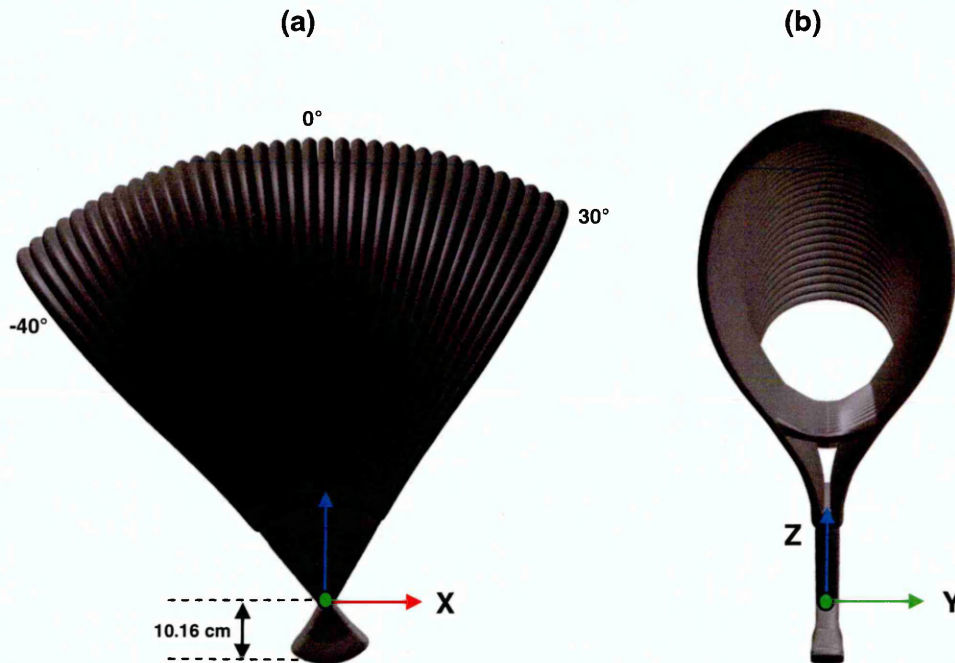


Figure 9.6 Simulated racket position during a serve around ball impact. The racket was rotated about an axis 762 mm from the butt in 2° increments. (a) shows a side view and (b) a front view of the simulated racket positions in Blender.

Silhouette images were rendered at each of the 36 racket angles in Figure 9.6 (a) and 3D racket position associated with each silhouette was estimated using fully calibrated sets with the camera poses orientated upright and randomly. The first analysis frame was when the racket was orientated at 0° because this silhouette provides the best view of the racket face plane. For both fully calibrated sets, the algorithm was instructed to perform two optimisations; the first worked backwards from when the racket was oriented at 0 to -40° , the second worked forwards from 0 to 30° . For the first optimisations i.e. when the racket was orientated at 0° , the candidate relative pose was obtained using the method described in section 7.4 (page 94-98). However, the following optimisations i.e. when the racket was not orientated at 0° , were initialised

using the solution of the previous. The 3D racket position was obtained by using the camera pose estimates to reconstruct the 130 coordinates on the racket face plane for each racket angle (Dunn et al. 2014). Reconstruction results were validated against known 3D coordinates obtained from the racket model mesh (section 6.8.1, page 83-84).

9.5 Results

9.5.1 Orientation

Figure 9.7 shows resultant reconstruction error on the racket face plane for unknown views orientated between 0 and 90°. Two fully calibrated sets were used to obtain these results; a set with camera poses orientated upright (at 0°) and a set with camera poses orientated randomly. Each bar in Figure 9.6 represents the RMSE for 2080 reconstructed coordinates. These were obtained using each of the sixteen camera pose estimates to reconstruct 130 coordinates on the racket face plane for each orientation i.e. 160 optimisations were performed using each fully calibrated set. The exact location of the 130 reconstructed coordinates are shown on the racket in Figure 9.7.

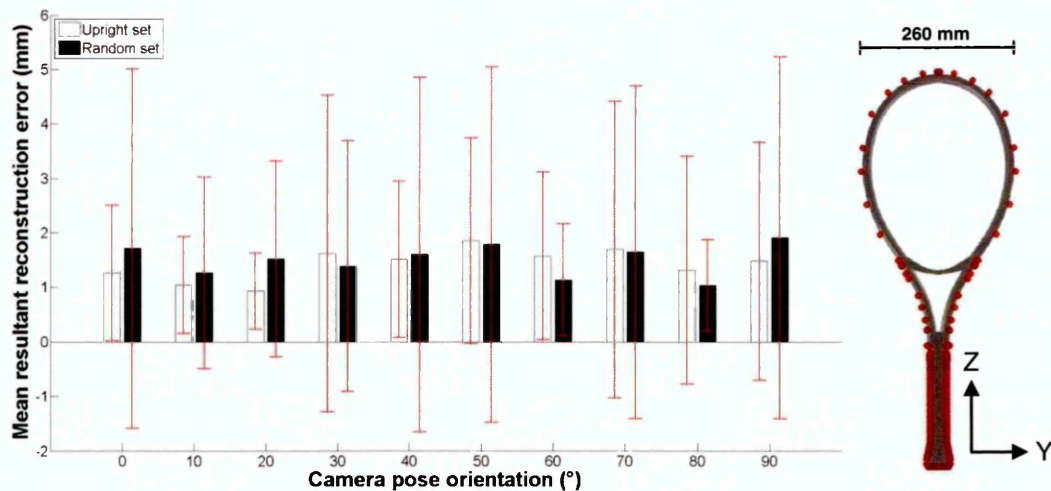


Figure 9.7 Resultant reconstruction error on the racket face plane for unknown view orientations between 0 and 90° obtained using fully calibrated sets consisting of camera poses orientated upright and randomly.

The largest mean resultant reconstruction error was 1.91 ± 3.32 mm using the randomly orientated set when the unknown views were orientated at 90°. The lowest mean resultant reconstruction error was 0.93 ± 0.69 mm using the fully calibrated set orientated upright when the unknown views were orientated at 20°. However, the randomly oriented fully calibrated set provided the lowest mean resultant reconstruction

error for six of the ten unknown view orientations tested. These orientations were 30, 40, 50, 60, 70 and 80°. Despite this there was not a significant difference ($p > .05$) between using a fully calibrated set with camera poses orientated upright or randomly, implying the method is robust. However, Figure 9.7 indicates it is beneficial to use a randomly orientated fully calibrated set for unknown views whose orientation exceeds 20° but not at 90°.

The following colour maps demonstrate the best results for each of the fully calibrated sets in Figure 9.7. Figure 9.8 (a) and (b) show resultant RMSE for unknown views orientated at 20° obtained using the fully calibrated sets orientated upright and randomly respectively. The fully calibrated set orientated upright produced more accurate results for unknown views at this orientation. Mean resultant reconstruction error was 0.93 ± 0.69 mm and 1.52 ± 1.80 mm using the fully calibrated orientated upright and randomly respectively. Furthermore, reconstruction for view 7 was 84% more accurate when using the fully calibrated set orientated upright.

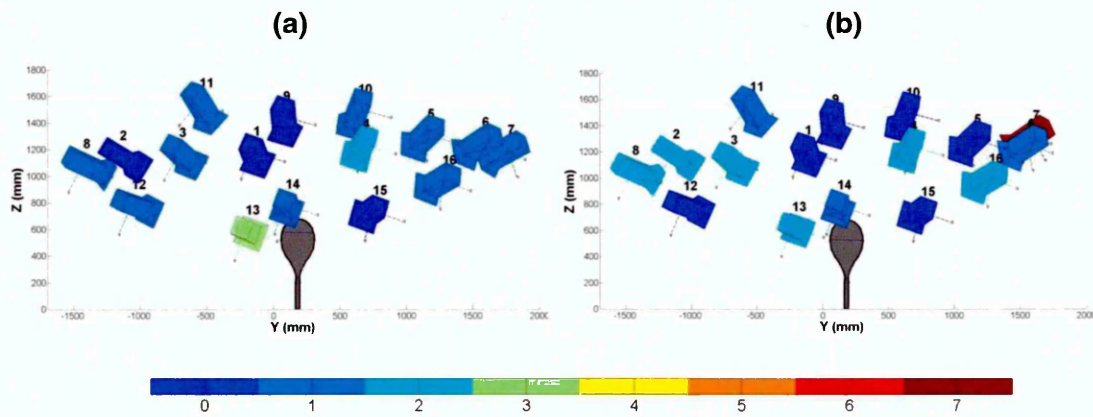


Figure 9.8 Resultant RMSE for unknown views orientated at 20° obtained using fully calibrated sets containing camera poses orientated (a) upright and (b) randomly.

Figure 9.9 (a) and (b) show resultant RMSE for unknown views orientated at 60° obtained using the fully calibrated sets orientated upright and randomly respectively. The fully calibrated set containing camera poses of random orientations provided more accurate results. Mean resultant reconstruction error was 1.57 ± 1.54 mm and 1.13 ± 1.03 mm using the fully calibrated orientated upright and randomly respectively.

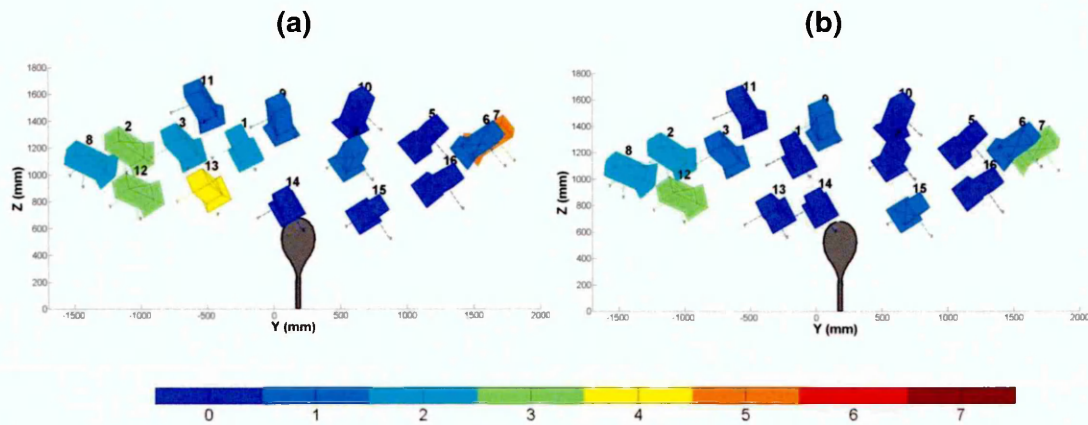


Figure 9.9 Resultant RMSE for unknown views orientated at 60° obtained using fully calibrated sets containing camera poses orientated (a) upright and (b) randomly.

9.5.2 Distance

Figure 9.10 shows RMSE for reconstruction of coordinates in the Y, Z and resultant (R) directions on the racket face plane. Over all camera distances, mean RMSE in the R direction varied by 0.52 mm, demonstrating the robustness of the method. The RMSE in the Z direction represented the largest component of error.

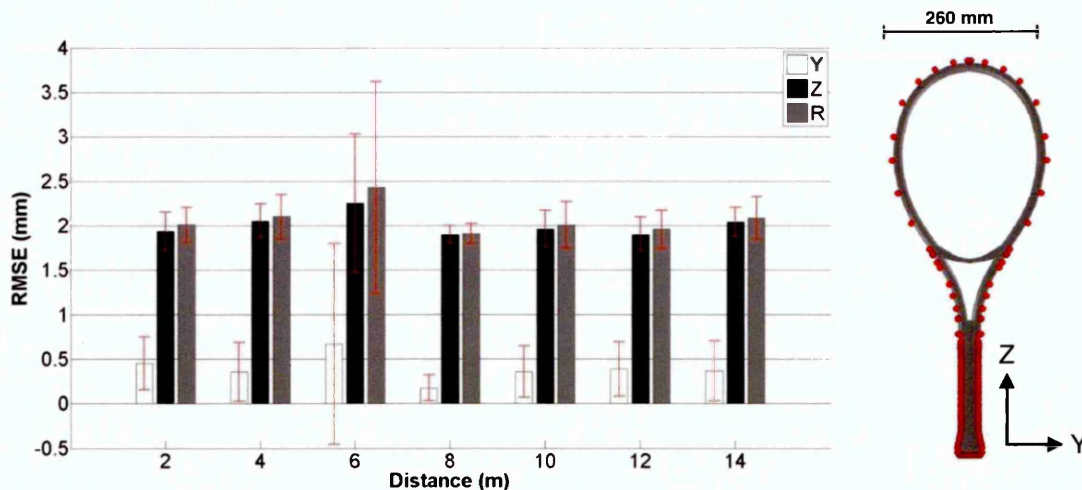


Figure 9.10 Mean RMSE ± standard deviation (mm) for reconstruction of coordinates in the Y, Z and R directions on the racket face plane from camera distances ranging between 2 to 14 m.

Figure 9.11 (a), (b) and (c) show RMSE for reconstruction on the racket face plane for all camera distances tested in the Y, Z and (resultant) R directions respectively. In general, RMSE for reconstruction was similar across all camera distances and angles (relative to the racket plane normal). Mean reconstruction error was 0.39 ± 0.49 , 2.00 ± 0.34 and 2.07 ± 0.49 in the Y, Z and R directions respectively. The slightly larger mean RMSE in all directions at 6 m in Figure 9.10 is mainly a result of the camera pose located at 60° relative to the racket face plane normal in Figure 9.11. The outlier at 6 m occurred because its RMS ETE did not reduce below the threshold of 0.5 of a pixel

following adjustment of the 100 initial candidate relative poses. Therefore, the solution associated with the lowest RMS ETE was selected, which was 0.82 of a pixel. This could have occurred because the silhouette view associated with the camera position located at 60° with respect to the racket face plane contains less information about the true shape of the racket. Chapter 8 showed that in general, racket position accuracy decreases as the angle of a camera pose estimate with respect to the racket face plane increases. Nevertheless, resultant reconstruction for the outlier differs on average by only 3 mm compared to all the other estimates. The black boxes in Figure 9.11 show the relative position of the camera poses with respect to the court lines.

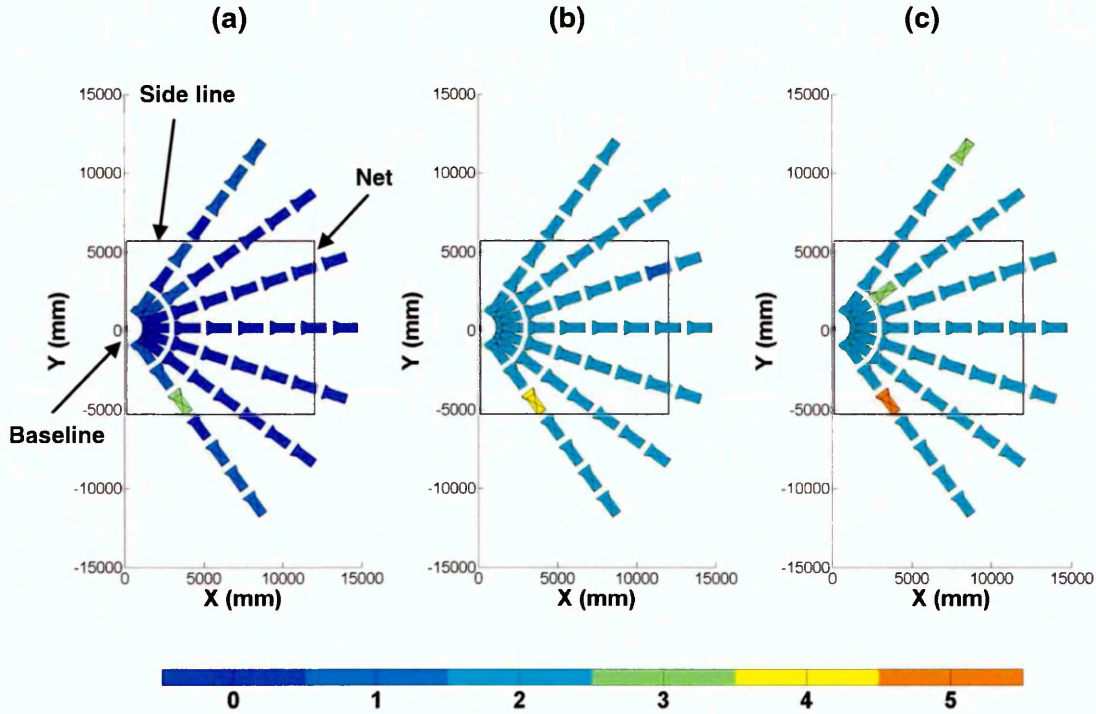


Figure 9.11 RMSE for reconstruction on the racket face plane in the (a) Y, (b) Z and (c) R directions for all the camera distances tested.

9.5.3 Serve simulation

Figure 9.12 (a) and (b) show a 3D view and a side view respectively of the camera pose estimates obtained for each of the racket angles in Figure 9.6, during the simulated serve. The increase in racket angle produced a curve of camera pose estimates increasing in height (Z axis) with respect to the racket. Each camera pose estimate was then used to reconstruct coordinates on the racket face plane.

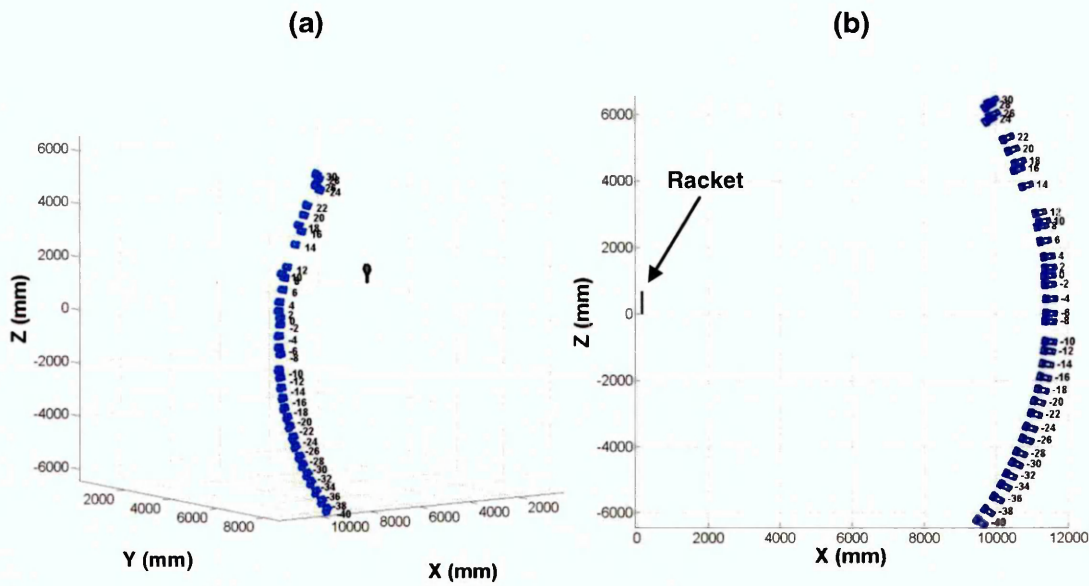


Figure 9.12 Camera pose estimates obtained during the simulated serve showing (a) a 3D view and (b) a side view.

Figure 9.13 (a), (b) and (c) show reconstruction errors in the Y, Z and resultant (R) directions respectively for coordinates on the racket face plane obtained using the upright orientated fully calibrated set, during the simulated serve. Mean reconstruction error was 0.31 ± 0.22 , 2.05 ± 0.16 and 2.08 ± 0.19 mm in the Y, Z and R directions respectively. Figure 9.13 (d), (e) and (f) show reconstruction errors in the Y, Z and R directions respectively for coordinates on the racket face plane obtained using the randomly orientated fully calibrated set, during the same simulated serve. Using this set, mean reconstruction error decreased to 0.26 ± 0.17 , 1.93 ± 0.13 and 1.96 ± 0.14 mm in the Y, Z and R directions respectively. Overall, 81% of racket pose estimations during the simulated serve were more accurate when the fully calibrated containing randomly orientated camera poses was used. For both fully calibrated sets, reconstruction error in the Z direction contributed a larger component of error during the simulated serve. However, the randomly orientated fully calibrated set decreased the number of racket pose estimates with errors above 2 mm in the Z direction by 50% (Figure 9.13 (e)), compared with the upright fully calibrated set (Figure 9.13 (b)). The largest reconstruction errors tended to occur when the racket was located at angles away from 0° i.e. toward and -40 and 30° . This is because out-of-plane 3D position estimation is difficult when only a single camera is used in a view fitting method. Since silhouette consistency depends on the extent to which silhouette scale changes with racket position (Price and Morrison 2007). Silhouettes associated with larger racket angles e.g. -40 and 30° , will be smaller in the image compared to when the racket is angled upright at 0° . Thus, making 3D racket position estimation marginally less accurate at larger racket angles.

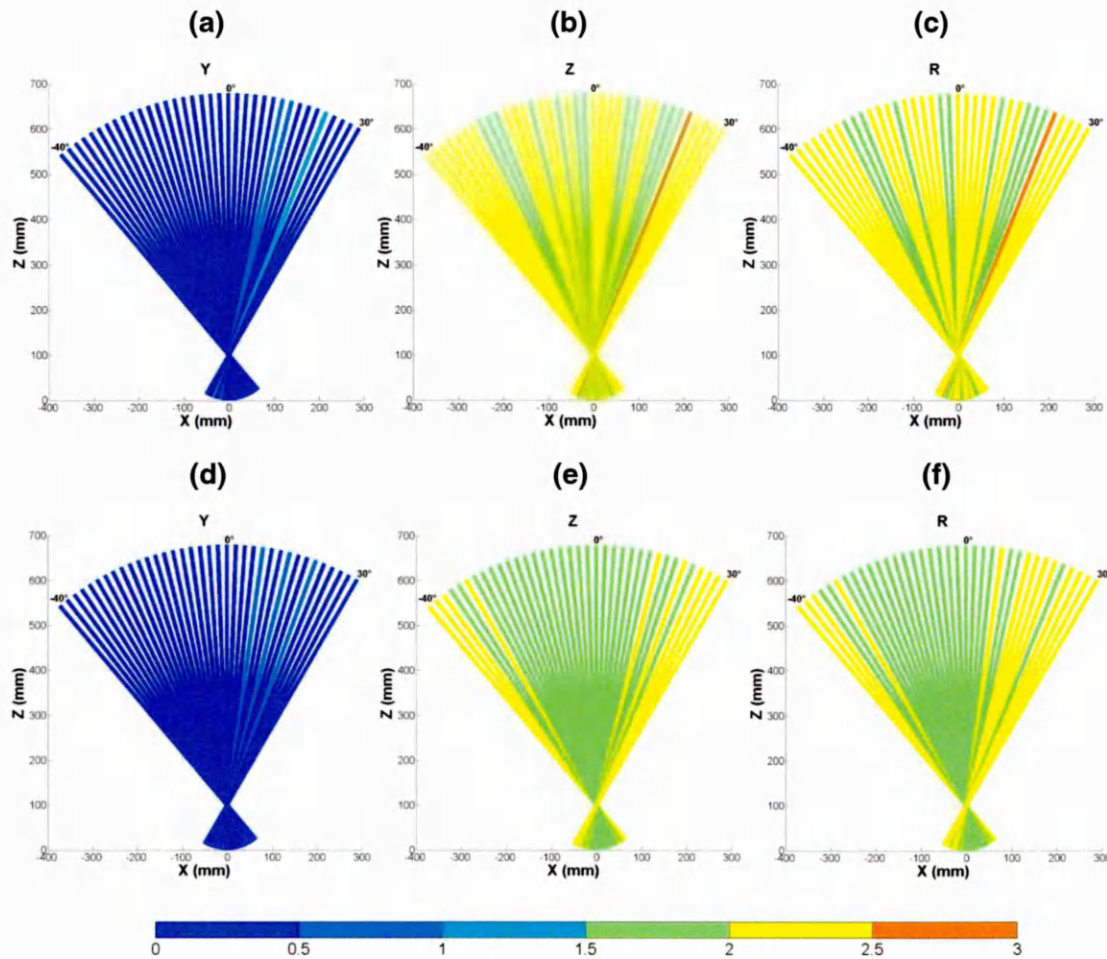


Figure 9.13 Reconstruction error obtained using a fully calibrated set orientated upright in the (a) Y, (b) Z and (c) R directions and reconstruction error obtained using a fully calibrated set orientated randomly in the (d) Y, (e) Z and (f) R directions, during a simulated serve.

9.6 Discussion

This chapter used synthetic data to demonstrate the potential of the method to be applied in real-play conditions. The 3D racket position was estimated from a range of camera angles and distances. Finally, the method was used to estimate racket position over a series of silhouettes during a simulated serve.

9.6.1 Orientation

Racket orientations that were expected to occur during typical tennis shots were simulated. For the simulations, the racket model remained stationary (upright) and the camera poses were rotated about the optical axis to provide racket silhouette views over a range of orientations (Figure 9.2). Camera orientations between 0 and 90° at 10° intervals were tested. Figure 9.7 shows the method was capable of estimating 3D racket position from camera orientations between 0 and 90° to a high accuracy. A fully calibrated set with the camera poses oriented upright produced a mean resultant reconstruction error of 1.43 ± 1.76 mm. This was compared with a set containing

randomly orientated camera poses which produced a marginally larger mean resultant reconstruction error of 1.50 ± 2.40 mm, over the ten unknown view orientations tested.

However, the synthetic fully calibrated set that consisted of camera poses that were randomly orientated provided marginally lower reconstruction errors for unknown poses with orientations above 20° . Therefore, if the method were to be applied in real-play conditions; two fully calibrated sets should be created. A set consisting of camera poses orientated upright would be used to estimate racket position expected to occur during a serve. The serve is played above the head with the upper limb outstretched (Elliott, Marsh and Blanksby 1986, Kotze, Mitchell and Rothberg 2000, Abrams et al. 2011) i.e. ball impact can be assumed to occur when the racket has an orientation of 0° (Figure 9.6). A set consisting of camera poses orientated randomly (in this study mean orientation was $61.19 \pm 23.94^\circ$) should be used to estimate racket position expected to occur during groundstroke's which are played at larger angles from the side of the body (Elliott, Marsh and Overheu 1989a), Elliott, Marsh and Overheu 1989b).

The influence of the orientation of the fully calibrated set on the accuracy of single view fitting methods has not been explored. This is because previous applications have optimised the pose associated with irregular shaped objects (Wong 2001, Price and Morrison 2007). There is less ambiguity in the estimation of poses associated with irregular objects (Price and Morrison 2007), this was demonstrated in chapter 8. However, this project has developed a view fitting method so that accurate results can be achieved for symmetrical objects i.e. a racket.

9.6.2 Distance

Figure 9.10 demonstrated that the method can be used to estimate racket position from camera distances that correspond to realistic positions on a tennis court. Overall, there was little difference in RMSE for reconstruction of coordinates on the racket face plane between the distances tested. For all distances, RMSE in the Z direction contributed a larger component of error to the camera pose estimates which translated to the racket pose estimates.

Firstly, the larger error in the Z direction is likely a result of there being too few constraints in terms of camera height with respect to the racket, imposed by the fully calibrated set. The maximum and minimum camera pose height in the fully calibrated set was 1390.5 and 866.5 mm respectively, providing a range of 523.9 mm over the three tiers of poses. This limits the silhouette views associated with the racket's pose to a smaller area than that of the height of the racket itself (686 mm). A larger range of

camera pose heights in the fully calibrated set could improve 3D racket position estimation. Price and Morrison (2007) found pose estimate accuracy increased with the number of views in the fully calibrated set. A network of fixed cameras in the laboratory that have been calibrated offline as used by Carazza et al. (2010) could provide a fully calibrated set of views with a larger range in height, possibly improving the results. Another reason could be that the camera distances tested were all located at a height of 1200 mm with respect to the local origin, so that the camera centre was 0.5 m higher than the tip of the racket. Testing camera pose distances with a range of heights may help to further understand the sensitivity of the Z component to error. However, strategic camera placement is important because pose pairings at a lower heights with respect to the racket tip could risk the baseline passing through the racket and the tangency points becoming unavailable.

Some of the simulated camera positions in Figure 9.4 could be applied in real play conditions for estimation of 3D racket position using the method. The camera pose circled red in Figure 9.4 was identified as being one of those positions. This was because it was located outside the court and therefore would not interfere during real play conditions. Similar camera positions have been used previously to measure ball and racket trajectories during real play conditions (Elliott et al. 2003, Choppin, Goodwill and Haake 2011). Choppin, Goodwill and Haake (2011) placed two cameras at both ends of the net to capture racket movements at the baseline. The distance from the centre of the baseline to either end of the net is approximately 14 m. From this distance, RMSE for reconstruction of coordinates on the racket face plane was 0.36 ± 0.34 , 2.03 ± 0.17 , and 2.09 ± 0.24 mm in the Y, Z and R directions respectively using the method (Figure 9.11). Choppin, Goodwill and Haake (2011) reconstructed marker position on the frame to within ± 2.5 mm from approximately the same camera distances. However, the results of this study are limited to the simulations used and therefore cannot be compared directly with Choppin, Goodwill and Haake (2011). Nevertheless the results are promising for application of the current method in real play conditions following further development and on-court testing. The smaller camera distances tested are relevant for practice conditions i.e. 2 to 8 m (Figure 9.4). For example, when the aim of a practice session is to focus on improving a particular stroke; the coach may chose to place the camera closer to the player or at a particular viewing angle. This may also be useful for manufacturers who wish to conduct racket testing for product development purposes.

Finally, a serve was simulated using the camera pose circled red in Figure 9.4 and the method was used to estimate 3D racket position over a series of silhouettes. Between

each silhouette frame the racket model was rotated about an axis 10.16 cm from the butt by 2°. Figure 9.13 showed that using a fully calibrated set in which the camera poses were randomly orientated improved the resultant reconstruction error by 76%, compared to when an upright orientated fully calibrated set was used. On average, using the randomly orientated set, 3D racket position during a simulated serve was reconstructed to within 1.96 ± 0.14 mm.

The randomly orientated fully calibrated set provided more accurate results because it contains a range of silhouettes which are more consistent with the orientations of the silhouettes of the simulated serve. This is reflected in the lower average RMS ETE value of 0.19 for pose estimates obtained using the randomly orientated fully calibrated set, compared to an RMS ETE of 0.29 for the upright orientated set. Furthermore, all the reconstruction errors reported in Figure 9.13 are significantly lower than the accuracy criteria of 15 ± 10 mm achieved by the markerless method developed by Corazza et al. (2010), which was used by Sheets et al. (2011) to measure tennis serve kinematics in practice conditions (Chapter 2, page 40). However, the results of the current study regarding application of a view fitting method to estimate 3D racket position, are limited to the simulated conditions expected to occur during a tennis serve.

Figure 9.13 also showed that the largest reconstruction errors occurred when the racket was angled away from 0° i.e. toward -40 and 30°, regardless of which fully calibrated set was used. At these angles the extent to which the racket face plane is out-of-plane with respect to the camera's image plane is largest. Therefore, since only a single camera was used for the view fitting method, the marginally larger resultant reconstruction errors associated with the larger racket angles is not surprising. Future experimentation could explore using a dual camera set-up to solve this problem, which was also recommended by Price and Morrison (2007).

Price and Morrison (2007) also found that using a single view fitting method to estimate the 3D position of particles during out-of-plane trajectories produced inaccuracies. Price and Morrison (2007) state that this occurred because only a single camera was used and silhouette consistency depends on the extent to which silhouette scale changes. Figure 9.14 (a), (b) and (c) show silhouettes associated with the racket's pose at -40, 0 and 30° respectively, during the simulated serve. It can be seen that the silhouettes associated with the racket's pose at -40 (Figure 9.14 (a)) and 30° (Figure 9.14 (c)) appear smaller in the image compared to the silhouette associated with the racket's pose at 0° (Figure 9.14 (b)). Table 9.1 presents the percentage of white pixels

relative to black pixels corresponding to each silhouette in Figure 9.14. The silhouette angled at 0° is comprised of the highest percentage of white pixels because the racket face plane is in full view. The silhouette associated with the racket's pose at -40° has the lowest percentage of white pixels because this is when the racket is located furthest from the camera and the racket face plane is not in full view. While the silhouette associated with the racket's pose at 30° has slightly more white pixels (compared to the silhouette in Figure 9.14 (a)) because the racket is closer to the camera but again the racket face plane is not in full view.

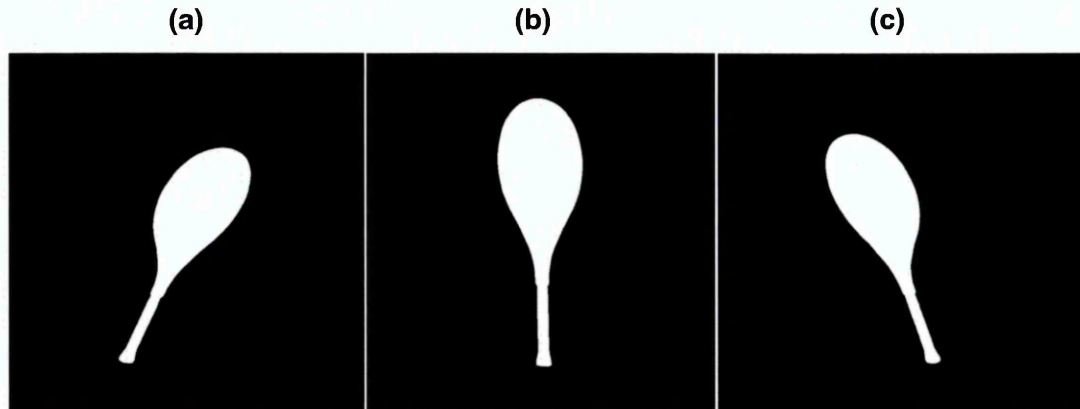


Figure 9.14 Silhouette views associated with racket's pose at (a) -40° , (b) 0° and (c) 30° .

Table 9.1 also shows the RMS ETE for each silhouette in Figure 9.14, measured using the randomly orientated fully calibrated set. The larger RMS ETE values indicate that the silhouettes in Figure 9.14 (a) and (c) are less consistent with the randomly orientated fully calibrated set, compared to the silhouette in Figure 9.14 (b). This shows that silhouette scale does indeed influence consistency and in turn the accuracy of a view fitting method, agreeing with Price and Morrison (2007). Thus, further supporting the findings in chapter 8 (page 124-125) that silhouettes comprised of more pixels produce more accurate pose estimations and that the reflective symmetry associated with frontal racket silhouette views does not decrease the accuracy of the method.

Table 9.1 Percentage of white pixels with respect to black pixels and RMS ETE corresponding to the racket silhouettes in Figure 9.12

	(a)	(b)	(c)
Percent of white pixels	4.7 %	5.8 %	4.9 %
RMS ETE	0.23	0.14	0.22

Since the method provides the best result when the racket is in full view, the most appropriate place to start an analysis is at the point of ball impact i.e. when the racket silhouette is assumed to be oriented upright at 0° during a serve (Figure 9.15). This

silhouette could be manually selected by the user and will provide a reliable initial candidate relative pose for the following optimisation. The method then works backwards to provide 3D racket position associated with silhouette frames before ball impact. During a real serve the racket will vibrate after ball impact which is likely to affect the shape of the silhouette and therefore accuracy of the method to estimate racket position. However, it is sufficient to analyse only a few frames before impact in order to determine racket head speed imparted upon the ball during a serve.

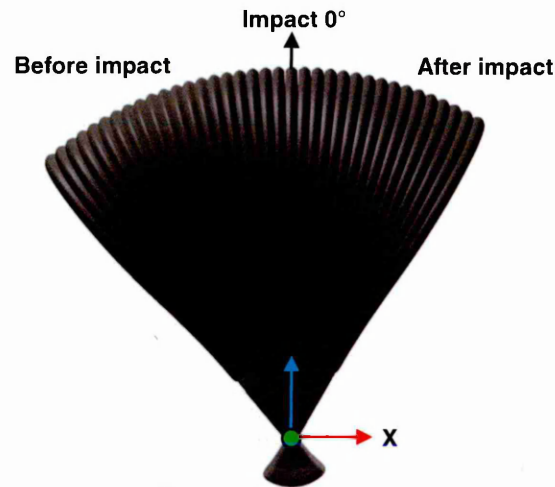


Figure 9.15 The silhouette associated with ball impact when the racket is at 0° should be manually selected to be the first analysis frame. The method then works backwards or forwards to estimate racket position before and after impact respectively.

9.6.3 Application to tennis

Previous studies have reported racket head velocities of 34.8 m s^{-1} (Elliott, Marsh and Blanksby 1989), 38.6 m s^{-1} (Chow et al. 2003), 43.2 m s^{-1} (Reid, Elliott and Alderson 2007) and 26.1 m s^{-1} (Sheets et al. 2011), during a flat serve. All these 3D studies used cameras operating at 200 frames per second (fps), apart from Reid et al. (2007) who used 250 fps. The lower velocity reported by Sheets et al. (2011) is due to the fact that the centre of volume rather than the tip of the racket was tracked.

Elliott, Marsh and Blanksby (1989) used a spacing of 0.005 seconds between frames in order to track racket velocity around impact. Players swung the racket toward the ball with a mean upward trajectory of the racket tip of $4 \pm 2^\circ$ with respect to the horizontal, 0.005 seconds prior to and at impact. In the current study, the tip of the racket model was displaced by a resultant distance of 4 mm between each 2° rotation about the axis 10.16 mm from the butt (Figure 9.6). The method developed in this project accurately estimated racket position when the racket was orientated between -10 and 10° (Figure 9.13). Therefore, the accuracy reported for the simulated serve may translate to a real

serve captured with a camera operating at 200 fps, providing the camera parameters and extracted silhouettes are accurate. However, the findings of this chapter are limited to the simulated conditions used; movement of the racket was not captured using a high-speed camera operating at a particular frame rate.

9.7 Conclusion

The method is capable of estimating the 3D racket position from orientations between 0 and 90°. The orientation of camera poses in the fully calibrated set has a marginal influence on 3D racket position estimation accuracy. A set containing camera poses orientated upright should be used to estimate racket orientations expected to occur during a serve. While a set containing camera poses orientated randomly should be used to estimate racket orientations expected to occur during groundstroke's and volleys.

The method estimated 3D racket position from camera distances up to 14 m from a range of positions. Some of the positions were located outside the court lines and could be suitable for use in real play conditions. A simulation was used to demonstrate whether the method could estimate racket position expected to occur during a serve. It was found that a fully calibrated set consisting of randomly orientated camera poses provided the best racket position estimate accuracy during the simulated serve. Furthermore, racket position estimate accuracy marginally decreased with the degree to which the racket face plane was out-of-plane with respect the camera's image plane. A number of recommendations were made based on these results for application of the method in real-play conditions. These included the most appropriate camera position, first silhouette analysis frame and camera frame rate. Forming a novel guide for the application of a single view fitting method to estimate 3D racket position in real play conditions.

10 Conclusions

10.1 Introduction

This final chapter summarises the research that has been undertaken in this project. The chapter includes a summary of the camera calibration method used to obtain intrinsic and extrinsic parameters and a common local reference frame. The validation of the method and recommendations for its application in real tennis-play conditions, are also discussed. Finally, there is a section on areas of possible future research.

10.2 Summary of research

The errors induced by camera calibration and silhouette extraction using a view fitting method to estimate 3D racket position have been investigated and are illustrated in Figure 10.1.

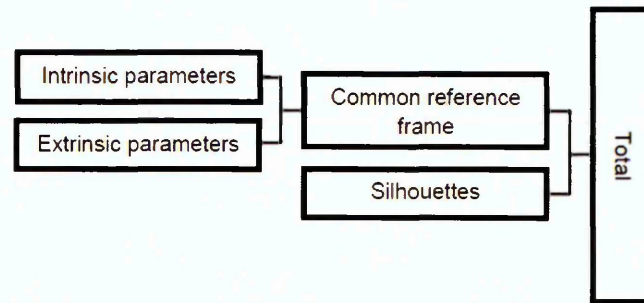


Figure 10.1 Errors that contributed toward 3D racket position estimation accuracy using the method.

10.2.1 Camera calibration

Camera calibration was identified to have an important role in the development of single view fitting methods. Chapter 3 investigated the repeatability and accuracy of intrinsic camera parameters obtained using Zhang's (1999) planar calibration method. By taking into account a number of methodological considerations the calibration parameters were repeatable and as a result calibration error was small. The intrinsic model identified in chapter 3 was taken forward in order to provide accurate extrinsic parameters for two high-speed Phantom Miro M110 cameras. The relative pose (rotation and translation) between the cameras was estimated using a stereo calibration. To produce the best results, it was found that the intrinsic parameters should not be recomputed for the proposed calibration method. This was the preliminary step towards determining the relative pose of a camera with respect to a tennis racket i.e. a common local reference frame.

10.2.2 Common local reference frame

To determine a common local reference frame for the cameras and racket, a rig was constructed on which a racket could be mounted. The racket was laser scanned on the rig in order to measure the angle between the racket tip and the handle. Angle offsets in the plane parallel and perpendicular to the racket face plane were 0.2 and 0.3° respectively, and thus deemed negligible. An orthogonal calibration object consisting of seven control points (upper and lower calibration planes) provided the most accurate relative pose of a camera with respect to the racket. This method was used to create a lab-based fully calibrated set (LFCS) in chapter 8.

10.2.3 Measuring silhouette consistency

The inconsistency of synthetic fully calibrated racket silhouette sets was measured using the epipolar tangency constraint. It revealed that discrete pixel rendering and silhouette extraction contributes to silhouette inconsistency. This is regardless of silhouette shape and the number of pixels in a silhouette. It was found that the consistency of a fully calibrated set can be improved by rendering silhouettes at the highest available resolution. The number of racket silhouette views in a synthetic fully calibrated set had a negligible effect on the overall consistency.

10.2.4 Determining the configuration of a fully calibrated set

A take-one-out validation process was used to determine the most practicable configuration for a lab-based fully calibrated set (LFCS). This involved removing each silhouette from a fully calibrated set in turn and estimating its pose using the method. Candidate relative poses were initialised using spherical coordinates before being adjusted by the Levenberg-Marquardt method in order to minimise the epipolar tangency error (ETE). It was found that the method was unable to estimate the camera pose associated with side silhouette views of the racket to the same accuracy as silhouette views that included the entire racket face plane. Removal of side views from a fully calibrated set greatly improved camera pose translation error in all directions. Therefore, side racket views were removed from further fully calibrated sets.

Since a racket has a symmetrical shape, semi-circular fully calibrated set configurations were tested i.e. sets consisting of silhouette views of only one side of the racket face plane. Sets containing 36, 27, 18 and 9 racket views were tested. It was found that camera pose estimate accuracy decreased with the number of views in a fully calibrated set. However, on average the sets containing 36 and 27 poses produced camera pose estimates that were only 1 and 5 mm and 0.02 and 0.1° less accurate respectively, than that achieved with the 72 pose set. Therefore, only a

relatively small amount of accuracy was sacrificed for a 50% reduction in solution time. Furthermore, a fully calibrated set with views at only one side of the racket face plane was practically more easy to reproduce in the laboratory.

10.2.5 Laboratory-based fully calibrated set (LFCS)

A LFCS with 21 racket views was created based on findings of the synthetic testing. The configuration of the LFCS consisted of 3 tiers of 7 camera poses. The LFCS did not have 27 views, (as with the synthetic testing) since the six camera poses with angles of 80° with respect to the racket face plane normal were not included. This was because it was not possible to obtain adequate image coverage at these larger camera angles to provide accurate calibration parameters. Use of a second master camera could help improve image coverage.

The total error associated with the LFCS was assessed i.e. systematic error arising from the calibration parameters and error from silhouette extraction. This was done by using the computed parameters to calculate the RMS ETE across the set. The RMS ETE was 0.41 of a pixel which was similar to that calculated in previous silhouette-based pose optimisation studies. It was found that silhouette extraction was the main contributor to the inconsistency of the LFCS. This highlighted the accuracy of the calibration method and the importance of accurate silhouette extraction.

The take-one-out validation process was used to provide camera pose estimates obtained using the LFCS. These camera pose estimates were used to reconstruct 3D coordinates on the racket frame surface to within ± 2 mm. This error is comparable to reconstruction errors reported in previous marker-based racket tracking studies. However, the environmental conditions were relatively constrained i.e. the study was performed in a laboratory and the racket remained stationary; stood upright on its butt. Furthermore, the racket was located close to the local origin; if moved further away reconstruction error would likely increase. However, reconstruction error could be improved by modifying the racket silhouette so it appears non-symmetrical. This could be achieved by painting sections of the frame with a particular contrasting colour or pattern however, this would impede silhouette extraction.

10.2.6 Simulation of real play conditions

Further synthetic testing was used to demonstrate the potential of the method in real play conditions. This involved simulating a range of racket orientations and camera distances typical in real play conditions. The silhouette views whose associated pose was estimated, were not part of the original synthetic calibrated set i.e. unknown views.

The 3D racket position associated with unknown views orientated between 0 and 90° was estimated to within 1.5 ± 1.8 mm. Unknown view orientation was the angle of rotation about the optical axis of the associated camera pose.

Furthermore, it was found that the orientation of the camera poses in the fully calibrated set marginally influenced 3D racket position accuracy. Therefore, it is beneficial to use a fully calibrated set containing camera poses which are orientated upright to estimate 3D racket position associated with unknown views that have orientations less than 20° i.e. racket position expected to occur during a serve. For unknown views that have orientations that exceed 20° i.e. racket position expected to occur during a groundstroke or a volley, then the fully calibrated set should consist of camera poses with a range of random orientations.

Camera positions appropriate to be used during real play conditions were also simulated. The camera positions tested were located both inside and outside the court and ranged from 2 to 14 m with respect to the centre mark of the baseline. Over all the positions tested, mean reconstruction error was 0.39 ± 0.49 , 2.00 ± 0.34 and 2.07 ± 0.49 in the Y, Z and resultant (R) directions respectively. An appropriate camera position for use in real play conditions was located 14 m and 40° with respect to the baseline centre mark i.e. located at the side of the net where it would not interfere with the playing environment. This camera position was used to estimate racket position during a simulated serve to within 1.96 ± 0.14 mm using a fully calibrated set containing randomly orientated camera poses. Which was significantly lower than the accuracy criteria of 15 ± 10 mm achieved by the markerless method developed by Corazza et al. (2010), which was used by Sheets et al. (2011) to measure tennis serve kinematics in practice conditions (Chapter 2, page 40).

Furthermore, racket position estimate accuracy during the simulated serve marginally decreased with the degree to which the racket face plane was out-of-plane with respect to the camera's image plane. Demonstrating that the method works best when the majority of the racket face plane is in view and the scale of the unknown silhouette matches the silhouettes in the fully calibrated set. Therefore, the silhouette associated with the racket when it is orientated at 0° should be used as the first analysis frame. The method could then work backwards to provide 3D racket position associated with silhouette frames before ball impact during a serve. Since a racket will vibrate after ball impact, this may affect the shape of the silhouette therefore, the method is not suitable to estimate racket position after impact.

10.3 Conclusion

Previous methods that have been used to track tennis racket movements require manual digitisation of markers in the image plane, which can be time consuming and inaccurate. A markerless single view fitting method has been developed to accurately estimate the 3D position of a racket. The method relies on a racket silhouette obtained using a camera. No previous research has developed a markerless method to estimate 3D racket position that does not require on-court calibration of multiple cameras in order to obtain a subject specific visual hull. The main conclusions of this project are:

- Existing marker-based systems and markerless motion capture multi-camera systems could be replaced with a single view fitting method to estimate racket position in real tennis-play conditions.
- Accurate intrinsic and extrinsic calibration parameters are a vital prerequisite for a view fitting method to be developed to estimate 3D racket position. The calibration methods outlined in this thesis allowed a laboratory-based fully calibrated set to be created with an overall inconsistency of less than 0.5 pixels.
- Discrete pixel rendering and extraction of racket silhouettes was the main contributor to the inconsistency of a fully calibrated set. Therefore, racket silhouettes should be captured at the highest available resolution.
- The method can estimate 3D racket position associated with a range of silhouette view orientations. These silhouette view orientations covered a range of racket orientations expected to occur in tennis.
- Fully calibrated sets containing silhouette views orientated upright should be used to estimate racket position expected to occur during a serve.
- Fully calibrated sets containing silhouette views orientated randomly should be used to estimate racket position expected to occur during groundstroke's and volleys.
- For the most accurate estimation of racket position during a stroke, only a few silhouette frames before impact should be analysed and these should contain a full view of racket face plane. Racket position estimation accuracy during a stroke is influenced by silhouette scale i.e. the degree to which silhouette size in the image changes as the racket moves further away or closer to the camera.

The main advantages of the method developed in this project over current published single view fitting methods are:

- Previously, single view fitting methods have received limited attention. Their main applications have been to improve the quality of visual hulls and to estimate trajectories associated with irregular shaped objects. The study by Price and Morrison (2007) was published in a brief conference paper and predominantly used synthetic data to estimate the trajectories.
- The method developed in this project has been extensively validated against both experimental and synthetic data using a symmetrical shaped object (the racket). Single view fitting methods are notoriously more ambiguous when estimating the pose associated with symmetrical objects. This highlights the accuracy of the view fitting method developed in this project.
- Compared to previous markerless motion capture methods applied in tennis, the method does not require a visual hull; it relies on an accurate fully calibrated silhouette set.
- Single view fitting methods have never been applied in a sporting context. The simulations conducted in this project allowed recommendations to be made for when the method is applied in real play conditions i.e. appropriate camera position and identification of racket silhouette views which provide the most accurate results.

The limitations of this project are:

- The generation of a laboratory-based fully calibrated set (LFCS) was limited to the resources available; a bespoke capture volume containing multiple calibrated cameras is likely to provide more accurate results.
- Racket position estimation using the LFCS was carried out using silhouettes associated with the racket's pose which were orientated upright only.
- Racket position associated with unknown silhouette views were not estimated using the LFCS.
- Application of the method to real-play conditions was limited to synthetic data.

The steps required so that the method can be applied in real-play conditions are:

- To develop a method for extraction of racket silhouettes from noisy backgrounds i.e. in real-play conditions spectators and the media are present.
- To develop the silhouette extraction method to remove the section of the racket silhouette which is occluded by the players hand.
- To provide the method with a user-friendly interface in the form of a tool. The tool should allow the user to specify camera distance and height with respect to

the local origin so that candidate relative poses can be generated using the method developed in this thesis. The tool should also allow the user to identify the first silhouette frames for analysis.

- To build an error check into the tool i.e. to ensure that only physically possible solutions are found.
- To test the method's ability to estimate racket position for a range of tennis strokes from non-invasive camera positions and to validate the results against the marker-based method developed by Choppin et al. (2007a).

10.4 Future research

This investigation has highlighted a number of possibilities for future research. This final section discusses possible areas of further study.

10.4.1 Future development in the laboratory

There are two main areas for future development in the laboratory that should be considered. The first being to investigate whether the distance of the racket from the local origin influences 3D racket position estimation accuracy using the method. This project demonstrated that 3D racket position estimation accuracy decreased with camera pose estimate accuracy. This was because the racket was located closer to the local origin than the camera pose estimates. Previously, the local common origin has been placed at the centre of the baseline to track racket movements using stereo camera systems during serves (Elliott et al. 2003, Sheets et al. 2011) and groundstrokes (Bahamonde and Knudson 2003, Choppin, Goodwill and Haake 2011). Therefore, racket distances (with respect to the local origin) that are suitable for real play conditions should be replicated in the laboratory. This could involve using a larger calibration plane and capturing silhouettes from larger camera distances.

Secondly, racket position should be estimated during a simulated stroke using the method. This could be achieved by attaching the racket to a pendulum and allowing it to swing (at a known velocity) about an axis in the handle. The method could then be used to estimate 3D racket position associated with each silhouette frame, allowing for racket velocity and angle to be calculated. Estimations could be validated against results obtained simultaneously using the marker-based method developed by Choppin et al. (2007a).

10.4.2 Application in real-play conditions

Three main areas have been identified for development so that the method can be applied in real-play conditions:

1. Silhouette extraction.
2. Candidate relative pose initialisation.
3. Detecting when the method fails.

Silhouette extraction was identified as the main contributor to inconsistency across a fully calibrated silhouette set. In this project silhouette extraction was facilitated by using a black racket in front of a back light. However, in real-play conditions the background will be noisy because spectators and the media will be present. Furthermore, players rackets are often decorated with a variety of colourful cosmetics, making racket silhouette extraction difficult. Image differencing has previously been used in tennis to identify the ball (Kelley 2011) and the player (Dunn 2014), therefore could be used to segment both the player and the racket. This would involve subtracting images that contain the player and racket from a reference image. However, separating the player from the racket is more challenging; no research currently exists into the application of a view fitting method to estimate the position of a non-rigid body i.e. the human body. Therefore, a workaround would be to remove the section of the handle from the silhouette occluded by the player's hand. Players' tend to hold their racket toward to the bottom of the handle therefore, only a small section of the silhouette would need to be removed. If this was the case, the fully calibrated set would also need to be modified by removal of pixels at the bottom of the handle in all the silhouette images. These pixels would need to be removed in a controlled manner so that it was not detrimental to the accuracy of the method. It is likely that the length of the racket handle in the silhouette image contributes to the accuracy of the racket's orientation about its long axis.

Providing the Levenberg-Marquardt method with an initial candidate relative pose that is close to the solution is important in order to prevent a local minimum (that is far from the solution) being found. In order to be applied in real-play conditions, the method would require some user input to provide the algorithm with knowledge of the camera's position with respect to the local origin. It is anticipated that 3D racket position will be estimated during strokes that occur relatively close to the local origin i.e. near the centre mark of the baseline. Therefore, the distance and height of the camera with respect to the local origin would need to be input to the algorithm. From this information a resultant camera distance could be determined and used to create a spherical coordinate configuration, using methods described in section 7.4 (page 94-98). The rotation component of the initial candidate relative pose could be determined using a method similar to that developed by Wong and Cipolla (2004). Whereby, the boundary

of the unknown silhouette view (whose relative pose is being estimated) is superimposed over its closest match in the fully calibrated set. This would rely on the users visual judgement. The rotation component of the closest match in the fully calibrated set would then be assigned to the initial candidate relative pose. This process would only need to be performed once (for the initial candidate relative pose), since the previous solution can be used as a candidate relative pose for the next optimisation.

Predicting when the method could fail is important for its application in real-play conditions. As suggested in chapter 9, a threshold based on the number of pixels in a silhouette could be used to decipher which silhouettes should be taken forward for analysis. It was found that the accuracy of the method to estimate 3D racket position during a stroke decreases with the number of pixels in a silhouette. This process could be facilitated by the user manually selecting racket silhouette frames for analysis before any are passed to the algorithm. The algorithm would then perform more stringent checks.

Despite the fact the method can provide accurate results, there could be occasions when it gives a poor result. Furthermore, this ambiguity could be exaggerated during real-play conditions because the scenario will be less constrained. Therefore, a racket position error check should be incorporated into the method. This could involve fitting a trend line through a series of estimated racket positions in order to check the expected position relative to the previous. Since camera frame rate will be known, it should be relatively simple to determine whether the current 3D racket position is a realistic estimation.

10.4.3 Programming language

The single view fitting method in this project was developed using the MATLAB (MATHWORKS 2015) programming language. However, programming language can have a significant effect on analysis time. OpenCV (Open Source Computer Library written in C++) has been demonstrated to perform image analysis operations 4 - 30 times faster than MATLAB (Matuska, Hudec and Benco 2012). Therefore, the efficiency of the method could be further improved to handle more silhouette images.

10.4.4 Markerless 3D position estimation tool

The single view fitting method in this project was developed purposefully for application in tennis. However, the method is not limited to racket shaped silhouettes. Other rigid striking implements used in sport could also be applied to the method i.e. golf clubs and hockey sticks. Furthermore, since these implements are not symmetrically shaped,

the method could potentially provide more accurate results, as shown section 8.6 (page 121-125).

10.4.5 Incorporation with existing tennis software

Finally, once the single view fitting method developed in this project has been validated for use in real-play conditions, it would be interesting to use it to measure racket movements alongside existing tennis analysis software. Single camera tools to measure ball trajectories Kelley (2011) and player movement Dunn (2014) in tennis have already been developed. Such a combination of information about player performance during real play conditions is unprecedented and could be used to inform future equipment fitting studies.

11 References

- ABDEL-AZIZ, Y. I. and KARARA, H. M. (1971). Direct Linear Transformation from Comparator Coordinates into Object Space Coordinates in Close-Range Photogrammetry. In: *Proceedings of the Symposium on Close-Range Photogrammetry*, Falls Church, VA, USA, 1-18, (American Society of Photogrammetry).
- ABDI, H. (2003). In: LEWIS-BECK, M., BRYMAN, A. and FUTING, T. (eds.). *Encyclopaedia for Research Methods for the Social Sciences*. Sage, Thousand Oaks, 792-795.
- ABRAMS, G. D., SHEETS, A. L., ANDRIACCHI, T. P. and SAFRAN, M. R. (2011). Review of tennis serve motion analysis and the biomechanics of three serve types with implications for injury. *Sports biomechanics*, **10** (4), 378-390.
- ALLEN, T. B., HAAKE, S. J. and GOODWILL, S. R. (2009). Comparison of a finite element model of a tennis racket to experimental data. *Sports Engineering*. **12**, 87-98.
- ALLEN, T. B., HAAKE, S. J. and GOODWILL, S. R. (2011). Effect of tennis racket parameters on a simulated groundstroke. *Journal of sports sciences*. **29**, 311-325.
- ALLEN, T. B., CHOPPIN, S. and KNUDSON, D. (2015). A review of tennis racket performance parameters. *Sports Engineering*. [online]. 22 January 2015.
- ARTENGO. (2015). Artengo personal coach. [online]. Last accessed 7th June 2015 at: <http://www.artengo.co.uk/tennis-idfam2472>
- AZEVEDO, T. C., TAVARES, J. M. R. and VAZ, M. A. (2008). 3D object reconstruction from uncalibrated images using a single off-the-shelf camera. *VIPimage 2007*.
- BABOLAT. (2015). *Babolat Play*. [online]. Last accessed 7th June 2015 at: <http://en.babolatplay.com/>
- BAHAMONDE, R. E. and KNUDSON, D. (2003). Kinetics of the upper extremity in the open and square stance tennis forehand. *Journal of Science and Medicine in Sport*, **6** (1), 88-101.

BLENDER. (2015). [online]. Available from: <https://www.blender.org/> [Accessed 4th March 2015].

BOUGUET, J. Y. (2010). Camera calibration toolbox for MATLAB. [online]. Available from: http://www.vision.caltech.edu/bouguetj/calib_doc/ [Accessed 10th November 2011].

BOYER, E. (2006). On using silhouettes for camera calibration. In: *Proceedings of the 7th Asian conference on computer vision (ACCV)*, Hyderabad, India.

BRADSKI, G. and KAEHLER, A. (2008). *Learning OpenCV: Computer vision with the OpenCV Library*. O'Reilly Media, CA, USA.

BRIGGER, P., HOEG, J. and UNSER, M. (2000). B-spline snakes: a flexible tool for parametric contour detection. *IEEE Transactions on Image Processing*, **9**(9), 1484-1496.

BRODY, H. (2000). Player sensitivity to the moments of inertia of a tennis racket. *Sports Engineering*, **3**, 145-148.

BRODY, H., CROSS, R. and LINDSEY, C. (2002). *The Physics and Technology of Tennis*. Solana Beach, CA: Racquet Tech Publishing.

CAPEL-DAVIES, J. and MILLER, S. (2007). Evaluation of automated line-calling systems. In: MILLER, S. and CAPEL-DAVIES, J. (eds.). *Tennis Science and Technology 3*, International Tennis Federation, London, 387-393.

CESERACCIU, E., SAWACHA, Z., FANTOZZI, S., CORTESI, M., GATTA, G., CORAZZA, S. and COBELLI, C. (2011). Markerless analysis of front crawl swimming. *Journal of biomechanics*, **44** (12), 2236-2242.

CHOPPIN, S., GOODWILL, S. R., HAAKE, S. J. and Miller, S. (2007a). 3D player testing at the Wimbledon Qualifying Tournament. In: MILLER, S. and CAPEL-DAVIES, J. (eds.). *Tennis Science and Technology 3*, International Tennis Federation, London, 333-340.

CHOPPIN, S., GOODWILL, S. R., HAAKE, S. J. and Miller, S. (2007b). 3D player testing results from the Wimbledon Qualifying Tournament. In: MILLER, S. and

CAPEL-DAVIES, J. (eds.). *Tennis Science and Technology 3*, International Tennis Federation, London, 341-348.

CHOPPIN, S. (2008). *Modelling of tennis racket impacts in 3D using elite players*. PhD, University of Sheffield.

CHOPPIN, S., GOODWILL, S, R. and HAAKE, S, J. (2011). Impact characteristics of the ball and racket during play at the Wimbledon qualifying tournament. *Sports Engineering*, **13** (4), 163-170.

CHOW, J., CARLTON, L., LIM, Y. T., CHAE, W. S., SHIM, J. H., KUENSTER, A. and KOKUBUN, K. (2003). Comparing the pre-and post-impact ball and racquet kinematics of elite tennis players' first and second serves: a preliminary study. *Journal of sports sciences*, **21** (7), 529-537.

CIPOLLA, R. and BLAKE, A. (1990). The dynamic analysis of apparent contours. In: , *Proceedings of 3rd International Conference on Computer Vision*, 616-623.

CIPOLLA, R. and BLAKE, A. (1992). Surface shape from the deformation of apparent contours. *International journal of computer vision*, **9** (2), 83-112.

CLERICI, G. (1976). *Tennis*. Octopus Books, London.

COE, A. (2007). The balance between technology and tradition in tennis. In: HAAKE, S. and COE, A. (eds.). *Tennis Science and Technology*, Oxford: Blackwell Sciences, 3-40.

CONAIRE, C. Ó., KELLY, P., CONNAGHAN, D. and CONNOR, N. E. (2009). Tennissense: A Platform For Extracting Semantic Information From Multi-Camera Tennis Data. In: *IEEE Digital Signal Processing, 2009 16th International Conference*, 1-6.

CORAZZA, S., MÜNDERMANN, L., CHAUDHARI, A. M., DEMATTIO, T., COBELLI, C. and ANDRIACCHI, T. P. (2006). A markerless motion capture system to study musculoskeletal biomechanics: visual hull and simulated annealing approach. *Annals of Biomedical Engineering*, **34** (6), 1019-1029.

- CORAZZA, S., GAMBARETTO, E., MÜNDERMANN, L. and ANDRIACCHI, T. P. (2010). Automatic generation of a subject-specific model for accurate markerless motion capture and biomechanical applications. *IEEE Transactions on Biomedical Engineering*, **57** (4), 806-812.
- CORAZZA, S., MÜNDERMANN, L., GAMBARETTO, E., FERRINGO, G. and ANDRIACCHI, T. P. (2010). Markerless motion capture through visual hull, articulated ICP and subject specific model generation. *International journal of computer vision*, **87**, 156-169.
- CROSS, R. (1999). Impact of a ball with a bat or racket. *American Journal of Physics*, **67**, 692-702.
- CROSS, R. (2002). Grip-slip behaviour of a bouncing ball. *American Journal of Physics*. **70** (11), 1093-1102.
- CROSS, R. and BOWER, R. (2006). Effects of swing-weight on swing speed and racket power. *Journal of sports sciences*. **24** (1), 23-30.
- CROSS, R. and NATHAN, A. (2009). Performance versus moment of inertia of sporting implements. *Sports technology*. **2** (1-2), 7-15.
- CROSS, R. (2010). The polar moment of inertia of striking implements. *Sports Technology*, **3** (3), 215-219.
- DIGNALL, R.J., HAAKE, S. J. and CHADWICK, S. G. (2000). Modelling of an oblique tennis ball impact on a court surface. In: SUBIC, A. J. and HAAKE, S. J. (eds.). *The Engineering of sport*, Oxford: Blackwell Sciences, 185-192.
- DI SALVO, V., COLLINS, A., MCNEILL, B. and CARDINALE, M. (2006). Validation of Prozone®: A new video-based performance analysis system. *International Journal of Performance Analysis in Sport*, **6** (1), 108-119.
- DRISCOLL, H., KELLEY, J., KIRK, B., KOERGER, H. and HAAKE, S. (2014). Measurement of studded shoe-surface interaction metrics during in situ performance analysis. *Sports Engineering*. [online]. 14th August 2014.

DUNN, M. (2014). *Video-based step measurement in sport and daily living*. PhD, Sheffield Hallam University.

DUNN, M., WHEAT, J., MILLER, and HAAKE, S. (2012). Reconstructing 2D planar coordinates using linear and non-linear techniques. In: *Proceedings of the 30th International Conference on Biomechanics in Sports*, 381-383.

ELLIOTT, B., MARSH, T. and BLANKSBY, B. (1986). A Three-Dimensional Cinematographic Analysis of the Tennis Serve. *International Journal of Sport Biomechanics*. **2** (4), 260-271.

ELLIOTT, B., MARSH, T. and OVERHEU, P. (1989a). A biomechanical comparison of the multisegment and single unit topspin forehand drives in tennis. *International Journal of Sport Biomechanics*, **5** (3), 350-364.

ELLIOTT, B., MARSH, A. P. and OVERHEU, P. R. (1989b). The topspin backhand drive in tennis-a biomechanical analysis. *Journal of Human Movement Studies*. **16** (1), 1-16.

ELLIOTT, B., MARSHALL, R. N. and NOFFAL, G. J. (1995). Contributions of upper limb segment rotations during the power serve in tennis. *Journal of Applied Biomechanics*. **11** (4), 433-442.

ELLIOTT, B. (2006). Biomechanics and tennis. *British Journal of Sports Medicine*. **40**, 392-396.

FORBES, K., VOIGT, A. and BODIKA, N. (2002). An inexpensive, automatic and accurate camera calibration method. In: *Proceedings of the Thirteenth Annual Symposium of the Pattern Recognition Association of South Africa (PRASA)*.

FORBES, K., VOIGT, A. and BODIKA, N. (2003). Using silhouette consistency constraints to build 3D models. In: *Proceedings of the Fourteenth Annual Symposium of the Pattern Recognition Association of South Africa (PRASA)*.

FORBES, K. (2007). *Calibration, recognition, and shape from silhouettes of stones*. PhD, University of Cape Town.

- GILLET, E., LEROY, D. and GERMAINE, N. (2003). Kinematic analysis of balance at tennis players according to the level of expertise. In: MILLER, S. (ed.). *Tennis Science and Technology 2*, International Tennis Federation, London, 217-220.
- GOODWILL, S. R. (2002). *The dynamics of tennis ball impacts on tennis rackets*. PhD, The University of Sheffield.
- GOODWILL, S. R., CAPEL-DAVIES, J., HAAKE, S. J. and MILLER, S. (2007). Ball spin generation by elite players during match play. In: MILLER, S. and CAPEL-DAVIES, J. (eds.). *Tennis Science and Technology 3*, International Tennis Federation, London, 349-356.
- GRATTAROLA, A. A. (1992). Volumetric reconstruction from object silhouettes: A regularization procedure. *Signal Processing*, **27** (1), 27-35.
- HAAKE, S. J., ALLEN, T. B., CHOPPIN, S. and GOODWILL, S. R. (2007). The evolution of the tennis racket and its effect on serve speed. In: MILLER, S. and CAPEL-DAVIES, J. (eds.). *Tennis Science and Technology 3*, International Tennis Federation, London, 257-271.
- HARTLEY, R. I. and ZISSERMAN, A. (2004). *Multiple view geometry in computer vision*. 2nd ed., Cambridge University Press.
- HAWK-EYE. (2015). Last accessed 30th March 2015 at: <http://www.hawkeyeinnovations.co.uk/>
- HEIKKILA, J. and SILVEN, O. (1997). A four-step camera calibration procedure with implicit image correction. In: *Computer Vision and Pattern Recognition*, 1106-1112.
- HERNÁNDEZ, C. (2004). *Stereo and silhouette fusion for 3D object modelling from uncalibrated images under circular motion*. PhD, Ecole Nationale Supérieure des Télécommunications.
- HERNÁNDEZ, C., SCHMITT, F. and CIPOLLA, R. (2007). Silhouette coherence for camera calibration under circular motion. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **29** (2), 343-349.

ITF Tennis. (2014). *ITF Tennis -Technical*. [online]. Last accessed 22nd April 2015 at <http://www.itftennis.com/technical/player-analysis/overview.aspx>

ITF Tennis. (2015). *Rules of Tennis*. [online]. Last accessed 29th April 2015 at <http://www.itftennis.com/media/194067/194067.pdf>

KANEMITSU, Y. (2003). The relationship between racket properties and player preference. *Sports Technology*, **2**, 41–48.

KASS, M., WITKIN, A. and TERZOPOULOS, D. (1988). Snakes: Active contour models. *International journal of computer vision*, **1**(4), 321-331.

KELLEY, J. (2011). *Measuring ball spin rates in match play tennis*. PhD, Sheffield Hallam University.

KELLEY, J., CHOPPIN, S. B., GOODWILL, S. R. and HAAKE, S. J. (2010). Validation of a live, automatic ball velocity and spin rate finder in tennis. In: SABO, A., KAFKA, P., LITZENBERGER, S. and SABO, C. (eds.). *The Engineering of Sport 8*. Procedia Engineering. **2** (2), 2967-2972.

KELLEY, J. (2014). A camera calibration method for a hammer throw analysis tool. In: JAMES, D., CHOPPIN, S., ALLEN, T., WHEAT, J. and FLEMING, P. (eds.). *The Engineering of Sport 10*. Procedia Engineering. **10** (72), 74-79.

KNUDSON, D. and BAHAMONDE, R. E. (1999). Trunk and racket kinematics at impact in the open and square stance tennis forehand. *Biology of Sport*, **16**(1), 3-10.

KNUDSON, D. V. and BLACKWELL J. R. (2005). Variability of impact kinematics and margin for error in the tennis forehand of advanced players. *Sports Engineering*, **8**, 75-80.

KOTZE, J., MITCHELL, S. R. and ROTHBERG, S. J. (2000). The role of the racket in high-speed tennis serves. *Sports Engineering*, **3**, 67-84.

LAURENTINI, A. (1994). The visual hull concept for silhouette-based image understanding. *IEEE Transactions on pattern analysis and machine intelligence*, **16** (2), 150-162.

- LAZEBNIK, S., BOYER, E. and PONCE, J. (2001). On computing exact visual hulls of solids bounded by smooth surfaces. In: *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*.
- LAZEBNIK, S. (2002). *Projective visual hulls*. Master's thesis, University of Illinois at Urbana-Champaign.
- LIU, J., ZHENG, N. and XIONG, L. (2009). Silhouette quality quantification for gait sequence analysis and recognition. *Signal Processing*, **89**(7), 1417-1427.
- LONGUET-HIGGINS, H.C. (1981). A computer algorithm for reconstructing a scene from two projections. *Nature*, **293**, 133-135.
- LORENSEN, W. E. and CLINE, H. E. (1987). Marching cubes: A high resolution 3D surface construction algorithm. In: *ACM siggraph computer graphics*, **21**(4), 163-169.
- MARSOUSI, M., EFTEKHARI, A., KOCHARIAN, A. and ALIREZAIE, J. (2010). Endocardial boundary extraction in left ventricular echo-cardiographic images using fast and adaptive B-spline snake algorithm. *International journal of computer assisted radiology and surgery*, **5**(5), 501-513.
- MARTÍNEZ-GALLEGO, R., GUZMÁN, J. F., JAMES, N., Perš, J., RAMÓN-LLIN, J. and VUČKOVIČ, G. (2013). Movement characteristics of elite tennis players on hard courts with respect to the direction of ground strokes. *Journal of Sports Science and Medicine*, **12**, 275-281.
- MATUSIK, W., BUEHLER, C., RASKAR, R., GORTLER, S. J. and MCMILLAN, L. (2000). Image-based visual hulls. In: *Proceedings of the 27th annual conference on Computer graphics and interactive techniques*. ACM Press/Addison-Wesley Publishing Co., 369-374.
- MATUSIK, W., BUEHLER, C. and MCMILLAN, L. (2001). Polyhedral visual hulls for real-time rendering. In: *Proceedings of 12th Eurographics Workshop on Rendering*, June 2001, 115-125.
- MATUSKA, S., HUDEC, R. and BENCO, M. (2012). The comparison of CPU time consumption for image processing algorithm in Matlab and OpenCV. In: *Proceedings of the 9th ELEKTRO International Conference*, 75-78.

- MILLER, S. (2006). Modern tennis rackets, balls, and surfaces. *British journal of sports medicine*, **6** (40), 401-405.
- MITCHELL, S. R., JONES, R. and KING, M. (2000). Head speed vs. racket inertia in the tennis serve. *Sports Engineering*, **3**, 99–110.
- MORÉ, J. J. (1978). The levenberg-marquardt algorithm: Implementation and theory. In: WATSON, G. A. (ed.). *Lecture Notes in Mathematics*, Numerical Analysis, Springer-Verlag, **630**, 105-116.
- MOTION ANALYSIS COOPERATION - MAC. (2012). [online]. Last accessed 11th August 2012 at: <http://www.motionanalysis.com/>
- MÜNDERMANN, L., CORAZZA, S., CHAUDHARI, A. M., ALEXANDER, E. J. and ANDRIACCHI, T. P. (2005). Most favourable camera configuration for a shape-from-silhouette markerless motion capture system for biomechanical analysis. In: *Proceedings of SPIE, Videometrics VIII*, **5665**, 278-287.
- MÜNDERMANN, L., CORAZZA, S. and ANDRIACCHI, T. P. (2006). The evolution of methods for the capture of human movement leading to markerless motion capture for biomechanical applications. *Journal of Neuroengineering and Rehabilitation*, **3** (1).
- Ng, C. C., Yap, M. H., Costen, N. and Li, B. (2015). Wrinkle Detection using Hessian Line Tracking. *IEEE Access*, **3**:1079-1088.
- NIEM, W. and BUSCHMANN, R. (1994). Automatic modelling of 3d natural objects from multiple views. In: PAKER, Y. and WILBUR, S. (eds.). *Image Processing for Broadcast and Video Production*. Workshops in Computing Series, Springer, Hamburg 1994.
- OTSU, N. (1979). A threshold selection method for gray-level histograms. *IEEE Transactions on Systems, Man and Cybernetics*, **9** (1), 62-66.
- PALLIS, J. M. (2004). [online]. Last accessed 14th March 2015 at: http://www.tennisserver.com./set/set_04_01.html
- PANSIOT, J. (2009). *Markerless visual tracking and motion analysis for sports monitoring*. PhD, Imperial College London.

PERŠ, J., BON, M., KVAČIČ, G., ŠIBILA, M. and DEŽMAN, B. (2002). Observation and analysis of large-scale human motion. *Human Movement Science*, **21** (2), 295-311.

PINGALI, G. S., Jean, Y. and Carlbom, I. (1998). Real time tracking for enhanced tennis broadcasts. In: *Computer Vision and Pattern Recognition*. Jun23-28 1998, 260-265.

PLUIM, B., MILLER, S., DINES, D., RENSTRÖM, P., WINDLER, G., NORRIS, B., STROIA, K., DONALDSON, A. and MARTIN, K. (2007). Sport science and medicine in tennis. *British Journal of Sports Medicine*. 41 (11), 703-704.

POLHEMUS. (2012). [online]. Last accessed 11th August 2012 at: <http://www.polhemus.com/>

PRICE, M. and Morrison G. (2007). Validating rigid body simulation of real particle shapes using pose estimation from high speed video. In: *proceedings of the 4th international conference on discrete element methods*, Brisbane, Australia, August 2007.

PROZONE. (2012). [online]. Last accessed 11th August 2012 at: <http://www.prozonesports.com>

RAYLEIGH, F. R. S. (1877). On the irregular flight of a tennis ball. *Messenger of Mathematics*, **7**, 14-16.

REID, M., ELLIOTT, B. and ALDERSON, J. (2007). Shoulder joint loading in the high performance flat and kick tennis serves. *British journal of sports medicine*, **41** (12), 884-889.

SCHORAH, D., CHOPPIN, S. and JAMES, D. (2015). Effects of moment of inertia on restricted motion swing speed. *Sports Biomechanics*. [online]. 21st April 2015.

SHEETS, A. L., ABRAMS, G. D., CORAZZA, S., SAFRAN, M. R. and ANDRIACCHI, T. P. (2011). Kinematics differences between the flat, kick, and slice serves measured using a markerless motion capture method. *Annals of biomedical engineering*, **39** (12), 3011-3020.

SHERRY, D. and HAWKINS, P. (2001). *Video Processor Systems for Ball Tracking in Ball Games*. Roke Manor Research Limited. [online]. U.K. Patent 041884. Patent from WIPO. Last accessed 4th February 2015 at: <http://wipo.int/pctdb/en/wo.jsp?WO=2001041884>.

SINHA, S. N. and POLLEFEYS, M. (2010). Camera network calibration and synchronization from silhouettes in archived video. *International journal of computer vision*, **87** (3), 266-283.

SONY. (2015). *Smart Tennis Sensor*. [online]. Last accessed 7th June 2015 at: <http://www.smarttennissensor.sony.net/NA/>

SRINIVASAN, S. (2004). *The theory of dimensioning: An introduction to parameterizing geometric models*. CRC Press.

STAMMBERGER, T., ECKSTEIN, F., MICHAELIS, M., ENGLMEIER, K. H. and REISER, M. (1999). Interobserver reproducibility of quantitative cartilage measurements: comparison of B-spline snakes and manual segmentation. *Magnetic resonance imaging*, **17**(7), 1033-1042.

STROBL, K., SEPP, W., FUCHS, S., PAREDES, C. and ARBTER, K. (2007). Camera Calibration Toolbox for MATLAB, http://www.vision.caltech.edu/bouguetj/calib_doc/

SVOBODA, T., MARTINEC, D. and PAJDLA, T. (2005). A convenient multi-camera self-calibration for virtual environments. *PRESENCE: Teleoperators and Virtual Environments*, **14**, 407-422.

TAVARES, J. M. R., AZEVEDO, T. C. and VAZ, M. A. (2008). 3D reconstruction of external anatomical structures using computer vision. *CMBBE 2008*.

TENNIS-TECHNOLOGY. (2015). *Babolot Play Pure Drive*. [online]. Last accessed 4th July 2015 at: <http://www.tennis-technology.com/babolatplay.html>

WENG, J., COHEN, P. and HERNIOU, M. (1992). Camera calibration with distortion models and accuracy evaluation. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, **10**, 965-980.

WHITESIDE, D., ELLIOTT, B., LAY, B. and REID, M. (2014). The effect of racquet swing weight on serve kinematics in elite adolescent female tennis players. *Journal of Science and Medicine in Sport*. **17** (1), 124-128.

WOLFRAMMATHWORLD. (2015). [online]. Last accessed 26th April 2015 at: <http://mathworld.wolfram.com/SphericalCoordinates.html>

WONG, K. -Y. K. (2001). *Structure and motion from silhouettes*. PhD, University of Cambridge.

WONG, K. -Y. K. and CIPOLLA, R. (2004). Reconstruction of sculpture from its profiles with unknown camera positions. *IEEE Transactions on Image Processing*, **13** (3), 381-389.

XSENS. (2012). [online]. Last accessed 21st July 2012 at: <http://www.xsens.com/>

YAN, F., CHRISTMAS, W. and KITTLER, J. (2005). A tennis ball tracking algorithm for automatic annotation of a tennis match. *British machine vision conference*. September 5-8 2005, 619-628.

ZEPP. (2015). *Zepp Tennis*. [online]. Last accessed 7th June 2015 at: <http://www.zepp.com/tennis/>

ZHANG, Z. (1999). Flexible camera calibration by viewing a plane from unknown orientations. In: *proceedings of the 17th international Conference on Computer Vision*. Corfu, Greece, September 1999, 666-673.

12 Personal bibliography

ELLIOTT, N., CHOPPIN, S., GOODWILL, S. R. and ALLEN, T. B. (2014). Markerless tracking of tennis racket motion using a camera. In: *proceedings of the 10th conference of the international sports engineering association (ISEA)*, Sheffield, 344 - 349.

ELLIOTT, N., CHOPPIN, S., GOODWILL, S. R. and ALLEN, T. B. (2012). Image-based visual hull of a tennis racket. In: *proceedings of the 9th conference of the international sports engineering association (ISEA)*, Lowell, USA, July 2012, 877.

13 Other publications

WASHIDA, Y., ELLIOTT, N. and ALLEN, T. (2014). Measurement of main string movement and its effect on tennis ball spin. In: *proceedings of the 10th conference of the international sports engineering association (ISEA)*, Sheffield, 557 - 562.

NICOLAIDES, A., ELLIOT, N., KELLEY, J., PINAFFO, M. and ALLEN, T. (2013). Effect of string bed pattern on ball spin generation from a tennis racket, *Journal of Sports Engineering*, **16** (3),181-188.

14 Appendix

14.1.1 Calculation of internal camera parameters from Blender (v2.70)

Consider an image rendered at a resolution of 1280 x 800 pixels using a camera with a focal length of 75 mm. Blender's default horizontal image size is 32 mm. At this resolution 1 pixel equates to 0.025 mm (32/1280 = 0.025). This allows the vertical image size to be calculated;

$$\text{Vertical image size} = 0.025 * 800 = 20 \text{ mm.} \quad (14.1)$$

Using these values it is possible to compute the matrix KK, which encapsulates the intrinsic parameters f_x , f_y , c_x and c_y (equation 13.1).

$$KK = \begin{bmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix} \quad (14.2)$$

Therefore, the matrix KK for a synthetic camera pose in Blender is given by

$$KK = \begin{bmatrix} 0.075(\frac{1280}{0.032}) & 0 & 1280/2 \\ 0 & 0.075(\frac{800}{0.020}) & 800/2 \\ 0 & 0 & 1 \end{bmatrix} \quad (14.3)$$

For this intrinsic model, the principal point is at the centre of the image and lens distortion is not modelled. The following section provides a MATLAB function to compute the matrix KK, given the focal length (mm) and image dimensions (pixels) from Blender.

14.1.2 MATLAB function to calculate the internal parameters for a camera in Blender

```
function KK = getBlenderKKMatrix2(fc, cc)

%INPUT
%fc = 2x1 vector containing focal length (mm) provided by Blender
%cc = 2x1 vector containing the dimensions at which the image was rendered in
Blender

%OUTPUT
% The matrix KK containing the intrinsic parameters fx, fy, cx, and cy.
% [fx    0    Px;
%  0    fy    Py;
%  0    0    1];

f_mm = fc;
f_meters = f_mm/1000;
```



```

Px = cc(1)+1;
Py = cc(2)+1;

sensorSize_Hz_mm = 32;
%sensorSize_Hz_mm = 128;
sensorSize_Hz_meters = sensorSize_Hz_mm/1000;

onePixel2oneMM_X = sensorSize_Hz_mm/Px;

sensorSize_Ver_mm = onePixel2oneMM_X * Py;
sensorSize_Ver_meters = sensorSize_Ver_mm/1000;

fx = f_meters(1) * (Px/sensorSize_Hz_meters); %rearranged =
fx/(Px/sensorSize_Hz_meters)
fy = f_meters(1) * (Py/sensorSize_Ver_meters); %rearranged =
fy/(Py/sensorSize_Ver_meters)

onePixel2oneMM_Y = sensorSize_Ver_mm/Py;
onePixel2oneMM = [onePixel2oneMM_X;onePixel2oneMM_Y];
%fxy = (fx + fy)/2;

KK = [fx, 0, Px/2;...
      0, fy, Py/2;...
      0, 0, 1];

```