

Article

A Deep Learning and Computer Vision Based Multi-Player Tracker for Squash

Maria Martine Baclig ¹, Noah Ergezinger ², Qipei Mei ³, Mustafa Gü³, Samer Adeeb ^{3,*} and Lindsey Westover ^{4,*}

¹ Department of Electrical and Computing Engineering, University of Calgary, Calgary, AB T2N 1N4, Canada; mariamartine.baclig@ucalgary.ca

² Department of Computing Science, University of Alberta, Edmonton, AB T6G 2E1, Canada; ergezinger@ualberta.ca

³ Department of Civil and Environmental Engineering, University of Alberta, Edmonton, AB T6G 2E1, Canada; qipei.mei@ualberta.ca (Q.M.); mustafa.gul@ualberta.ca (M.G.)

⁴ Department of Mechanical Engineering, University of Alberta, Edmonton, AB T6G 2G8, Canada

* Correspondence: adeeb@ualberta.ca (S.A.); lwestove@ualberta.ca (L.W.)

Received: 8 November 2020; Accepted: 7 December 2020; Published: 9 December 2020



Featured Application: Autonomous multi player tracking for kinematic evaluation of elite squash players.

Abstract: Sports pose a unique challenge for high-speed, unobtrusive, uninterrupted motion tracking due to speed of movement and player occlusion, especially in the fast and competitive sport of squash. The objective of this study is to use video tracking techniques to quantify kinematics in elite-level squash. With the increasing availability and quality of elite tournament matches filmed for entertainment purposes, a new methodology of multi-player tracking for squash that only requires broadcast video as an input is proposed. This paper introduces and evaluates a markerless motion capture technique using an autonomous deep learning based human pose estimation algorithm and computer vision to detect and identify players. Inverse perspective mapping is utilized to convert pixel coordinates to court coordinates and distance traveled, court position, 'T' dominance, and average speeds of elite players in squash is determined. The method was validated using results from a previous study using manual tracking where the proposed method (filtered coordinates) displayed an average absolute percent error to the manual approach of 3.73% in total distance traveled, 3.52% and 1.26% in average speeds <9 m/s with and without speeds <1 m/s, respectively. The method has proven to be the most effective in collecting kinematic data of elite players in squash in a timely manner with no special camera setup and limited manual intervention.

Keywords: video tracking; sports broadcast analysis; player identification; kinematics; racquet sports

1. Introduction

Quantitative analysis of human movement has long been an interest within sports biomechanics for its ability to determine performance and strategy [1], as well as its application in rehabilitation to identify injury risk factors [2] and facilitate recovery [3]. The demand for motion analysis to capture more complex environments in sport is pushing for the development of faster, more autonomous, and sophisticated techniques. Biomechanical analysis in applications such as training and competition requires the following unique criteria: provide accurate kinematic data, deliver information in a timely manner, and remove factors which restrict or influence the subject's natural movement [4].

The most widespread and common techniques for kinematic data capture have historically been manual notation on prerecorded videos and marker-based technologies. However, they are not

without their drawbacks. Manual notation involves replay of game film and the manual localization of joints of interest for each sequential frame and from each camera perspective if required [5]. In the implementation by Sanderson and Way, notational analysis was used to describe sequential strokes in squash using symbols and court plans for positional information [6]. This method is affordable and does not require attachment of markers but remains a time-consuming and laborious task prone to subjective error. Marker-based systems utilize multiple cameras and markers placed on the specific joints of the subject to locate the position of the body. Many commercially available systems use automatic optoelectronics which require subjects to place passive reflective markers on their body usually in the form of a suit, as reported by Richards who reviewed passive marker systems including the Ariel system, Motion Analysis' HiRes system, Peak Performance's Motus system, Qualisys' ProReflex system, BTS's ElitePlus system, and Vicon's 370 system [7]. The passive markers work with multiple cameras that emit invisible infrared light and reflect the infrared light back to the cameras. The cluster of markers improves time efficiency as it allows for quick location of the subject. With decreased processing time, limitations remain due to the markers which include long participant preparation time and inevitable variability in placement [8]. There is also a higher possibility of rigid body assumptions being violated [9]. Additionally, markers cannot be used in most competition settings due to the physical and/or psychological effects from having extra attachments. Further, because these methods require specialized equipment, data collection is restricted to local participants, limiting the ability to study elite level athletes from across the world. These limitations have motivated the development of motion analysis systems towards an autonomous, markerless approach using deep learning and computer vision. These applications are mostly on slow movements such as walking or jogging and have remained in laboratory analyses as studies evaluate the accuracy of their systems with the use of multiple cameras [10–13].

More recently, computer vision has been applied to player tracking for indoor sports. Perš and Kovačić [14] focused on tracking handball players using two cameras and performing applied frames subtraction based on motion detection, template tracking, and color-based tracking. Another system applied to handball was discussed by Santiago et al. [15] who proposed a player tracking method based on image and colors processing. Both methods require specific equipment, a special camera setup and present tradeoffs between manual intervention and analysis times. Specifically for squash, a computer vision driven tracking system was developed using the HOG (Histogram of Oriented Gradients) algorithm implemented within the OpenCV library for player detection [16]. In this method, detection is of the general player rather than definitive tracking of the feet causing valuable kinetic information to be lost. In addition, it requires a special camera setup with prior court calibration and it is noted that the algorithm suffers from poor response time.

Few papers study player kinematics and kinetics of squash. In past studies, Hughes and Franks [17] investigated the correlation of velocity and acceleration in the last ten seconds of a rally for winners and losers. To collect positional data of players, observers manually annotated video images of squash matches using a digitizing pad and stylus [18,19]. Vučković et al. [20] built on previous work to analyze entire match play as well as at the rally level, studying the distinction between winner and losers of the rally in terms of court position, total distance, average velocity, and acceleration. In a following study, a correlation was drawn to the use of the 'T' area to player ability [21]. These studies utilized the SAGIT/Squash tracking system, a real-time data acquisition tracking system which requires colored images from a birds-eye camera view positioned above the court and compares to an empty court image to determine player position [22,23]. Correlations have further been studied between the game outcome and rank of elite squash player by quantifying distance traveled, position relative to the T, dominance of the T, average velocities, and frequency distribution of velocities of different ranking players [24]. Buote et al. [24] concluded that total distance and average velocity of the player is not suggestive of the rank of the player or the outcome of the game, however, the player's rank can indicate their ability to dominate the T and control the court. The data were collected using broadcast videos provided by the Professional Squash Association (PSA), analyzing only active match play of the full

court view provided by the main camera. Using video analysis software, Dartfish Team Pro version 8 (2015), markers were manually placed on each foot for every eligible frame to determine player position in the video coordinate system. Ten reference points on the court were recorded to determine a coordinate transformation converting the video image coordinate system to the coordinate system of the plane of the court [24].

The contribution of the current paper is to advance the development of accurate and reliable markerless motion tracking for squash by removing the need for a special setup, reducing processing times, and limiting user intervention used by previous squash studies. The proposed methodology improves on the previous work done by our research group [24] by replacing the time-consuming and laborious task of player tracking with an autonomous deep learning based human pose estimation to detect individuals in the frame and computer vision to identify the players. Removing the need for specific equipment and limiting significant user intervention increases the number of eligible matches we can analyze in a timely manner. Matches that are filmed by the PSA or filmed similarly are available to be analyzed by our methodology. This study outlines and validates our proposed method with the results of the previous study completed by Buote et al. [24] that quantify the players' distance traveled, position relative to the T, and average velocities. This is the first study to apply deep learning and computer vision motion tracking techniques to study elite squash players in competition.

2. Materials and Methods

The method was tested on a quarterfinals match of the 2013 Canary Wharf Classic Tournament collected from the PSA video collection. The broadcast video was obtained at a frame rate of 25 fps and a resolution of 720×576 . The match was between professional players El Shorbagy (dark grey shirt) and Mustonen (white shirt), with PSA Rank 5th and 53rd respectively at the time of play and consisted of five games. Automated tracking was performed on the match only on the frames which had been previously manually tracked [24]. This allowed direct quantitative comparison to validate the automated tracking's effectiveness. The procedure involves three main steps as described below.

2.1. Preprocessing

The full match broadcast is split up into five games, each as a separate file. Each of the subsequent methods are applied to each game. The games are played back, and a user manually identifies when the game is in play. Only video frames involving gameplay are kept so that analysis does not include moments when players are walking around between rallies, for example. The games are then filtered to only include the main camera angle. Throughout the broadcast, the camera angle is changed to give different views of the court. Only the main camera angle is used because it contains all of the court reference points used in the coordinate transformation. The program starts by generating a histogram for the first frame of video and adds this to a list of reference histograms (no references exist at this point, so the first frame is always the first reference). The histograms measure the frequency of color intensities, with one histogram for each color channel, where the bins are brightness levels and the y-axis measures the number of pixels within that range of brightness. The program then iterates through the rest of the video, for each frame generating a histogram and comparing it to the list of references. If the histogram is similar to one of the references within a defined threshold, it is grouped with those frames. If the frame is not similar to any of the references, it is added as a new reference. The result is separate video clips that each contain a separate camera angle. The longest video will be the main camera angle, as this is the angle that is used most often and is what will be used for player tracking. A timestamp in the top left corner of the frame is also added in this step. An example pre-processed frame is shown in Figure 1.



Figure 1. An example pre-processed frame.

2.2. Player Tracking and Identification

A general-purpose multi-person pose estimation neural network [25] was used for player tracking. The network consists of a two-branch multi-stage convolutional neural network with five initial convolutional layers followed by seven convolutional layers for each joint and limb. It outputs heatmaps for each joint and limb where further code detects peaks in the heatmaps to locate a maximum of 17 key joints and the limbs that connect them as shown in Figure 2. The neural network was trained on the MPII human multi-person dataset [26] and the COCO 2016 keypoints challenge dataset [27]. A tracking confidence value is also provided for each player that was tracked.

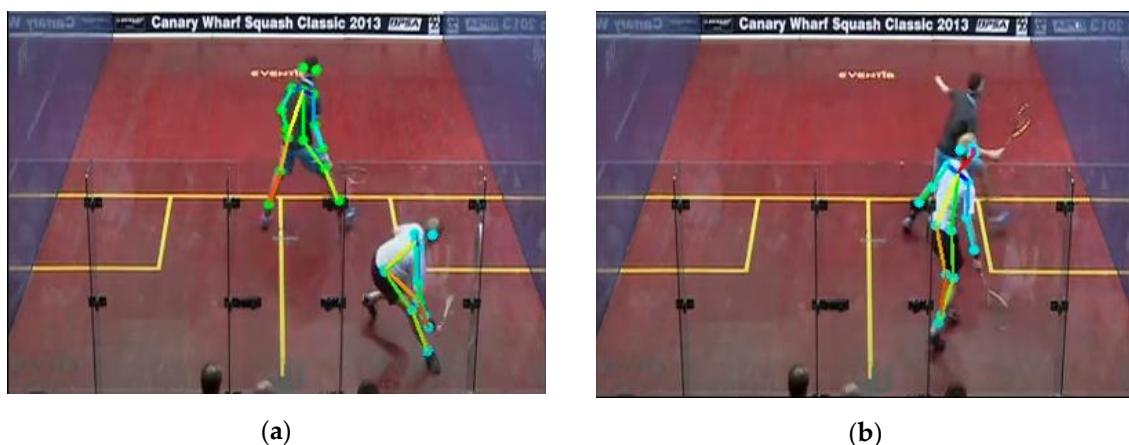


Figure 2. Difficult tracking scenarios. (a) As the player lunges, the arm reaching down is identified as a leg, and the player's left leg is not tracked. (b) The player in behind is not tracked at all, as there is too much occlusion from the player in front.

First, the video is cropped to only include the court as to avoid tracking spectators that might be visible surrounding the court. Before tracking begins, the user is asked to draw a box around each player. A reference histogram of each player's torso is generated for later identification. The tracking process then begins. For each frame, depending on the game scenario, either both, one, or neither of

the players will be tracked by the neural network. Examples of difficult tracking scenarios such as unnatural limb positioning and player occlusion are shown in Figure 2.

A minimum tracking confidence value (2.0) and a minimum number of joints (10) are used to remove tracking errors. Frames with values below the minimum thresholds are removed from the analysis. In the remaining frames, a histogram of each tracked player's torso is generated and compared to the references generated earlier to identify the players. The coordinates of each player's joints are recorded in a spreadsheet.

Timestamps associated with each frame are read by cropping the left top corner. Before tracking, a set of pictures of the numbers 0–9 are provided in the same font and black background they appear in the video. Using Python's computer vision library, OpenCV, contours are detected to separate individual numbers, and a machine learning based classification algorithm k Nearest Neighbors (kNN) is used to associate numbers detected to the set provided. This process repeats for each frame until the video is completed.

2.3. Postprocessing

To convert the screen-space coordinates to court-space coordinates, a rotation matrix and translation vector are generated using reference points of the court as they appear in the video, and their corresponding coordinates in the plane of the court (known based on standard court dimensions). The method to generate the equations was the same as used by Buote et al. [24]. The reference points are shown in Figure 3.

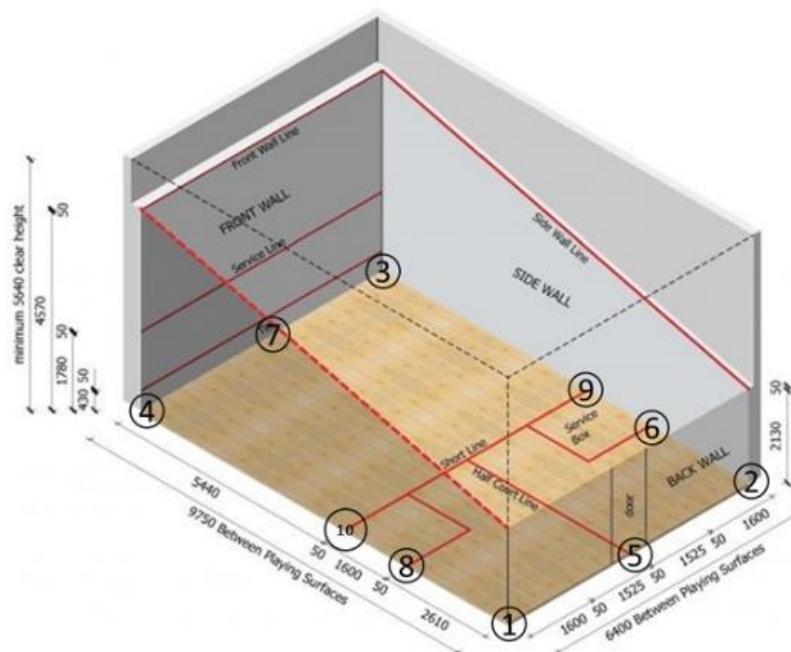


Figure 3. Reference points used for coordinate conversion (permission granted and modified from [28]).

Once court coordinates are obtained, further analysis to attain player data is completed. For positional data, the court floor coordinate system's origin is placed at the T with x increasing to the right and y increasing to the front wall. The x- and y-coordinates of each player's left and right feet are averaged for further analysis. Total distance is determined by calculating the change in coordinates between consecutive frames using the Euclidian norm. A player's average radial distance from the T is calculated by taking the Euclidian norm of their coordinates in each frame [29]. The percentage of time a player spends to the left of the T is determined by the number of coordinates with negative x-values

divided by the total coordinates. The percentage of time spent behind the T is similarly calculated as the number of coordinates with negative y-values divided by the total coordinates.

Velocity components were calculated individually with the change in the x-coordinate and the y-coordinate divided by the time between consecutive frames. The time difference is determined by subtracting the smaller timestamp from the larger timestamp associated with the frames. Average speed over the entirety of a game was determined as the sum of the magnitude of the velocity components divided by the total number of velocity components.

To validate player tracking, coordinates were graphed against the coordinates from the manual tracking method reported by Buote et al. [24] and quantified using the coefficient of determination (R^2). Percent error was calculated for all player data collected as:

$$\% \text{ Error} = \frac{\text{Statistic} - \text{Reference statistic}}{\text{Reference Statistic}} \times 100\%, \quad (1)$$

where the reference statistic is the statistic reported by Buote et al. [24].

A 5th order moving average filter calculated as:

$$\text{Coordinate}_t = \frac{1}{5} \sum_{j=-2}^2 \text{Coordinate}_{t+j}, \quad (2)$$

was applied to the court coordinates prior to analysis to smooth minor fluctuations in foot detection. Further investigation on the unfiltered total distance for 50 consecutive frames (2 s) in each game showed a large discrepancy in total distance traveled compared with the manually tracked data [24] (Table 1) due to characteristic differences in how tracking was managed. For manual tracking, observers were instructed to locate the feet and minimize coordinate changes between frames as to speed up the annotating process and to produce relatively smooth motion tracking. Our proposed method does not hold historic data on the previous frame analyzed, which results in variation in foot detection between frames. From frame to frame, the foot node can be located higher up the ankle or lower down on the floor. Especially for y-coordinates, this can make a small difference during court conversion due to the length of the squash court compared to its width. These minor variations can accumulate causing large differences in summations over time and the final values can be dependent on the number of frames collected. Thus, a filter was applied to the player coordinates to account for variability in foot tracking.

Table 1. Comparison of El Shorbagy's total distance discrepancy calculated from 50 consecutive frames (2 s) from unfiltered coordinates of proposed tracking and manual tracking.

Game	Total Distance Traveled from Manual Tracking (m) [17]	Total Distance Traveled from Proposed Tracking (m)	Average Difference (standard deviation) (m)	Max Absolute Difference ¹ (m)
1	0.820	1.73	0.0179 (± 0.0288)	0.120
2	3.139	5.86	0.0544 (± 0.0831)	0.361
3	0.821	1.84	0.0273 (± 0.0273)	0.127
4	3.511	7.08	0.0959 (± 0.128)	0.508
5	2.532	4.66	0.0531 (± 0.0698)	0.326
Average	2.165	4.23	0.0497 (± 0.0674)	0.288

¹ Difference is calculated as the distance vector statistic minus reference distance vector statistic over the same period, where the reference statistic is the statistic collected from Buote et al. [24].

3. Results

The match spanned five games and 41.4 min with 22.4 min (55.3%) of active match play. An average of 76.5% of active match play was analyzed with the removal of frames using other angles than the

court main camera view. A summary of game length, % of active game play, and % analyzed is reported in Table 2.

Table 2. Game length, active game play, and % of analyzed active match play by Buote et al. [24].

Game	Game Length in Minutes	Minutes (%) of Active Game Play	% of Analyzed Active Match Play
1	6.0	4.5 (74.3%)	78.1%
2	8.1	4.5 (56.1%)	73.6%
3	8.2	4.9 (60.0%)	76.4%
4	9.7	4.8 (49.7%)	74.2%
5	9.4	3.7 (39.4%)	80.2%
Average	8.3	4.5 (55.9%)	76.5%

The match was recorded at 25 fps where manual tracking included frames during active game play taken from the court main camera view [24]. Further analysis by Buote et al. [24] was done only between consecutive frames and did not interpolate between breaks longer than 1/25 s in time. To validate the proposed method, player detection and identification was done on the same frames. Frames were discarded by the proposed method if a player was not identified, which was usually caused by player occlusion or an unnatural pose (i.e., Figure 2). Table 3 presents the number and percentage of frames retained by our methodology compared to the manual tracking.

Table 3. Total # of active match play frames analyzed from manual tracking and inputted into proposed tracking method and # and % of frames outputted from proposed tracking method.

Game	Total # of Frames Collected from Manual Tracking ¹	# of Frames Retained	% of Frames Retained
1	5249	4755	90.6%
2	5000	4223	84.5%
3	5570	4262	76.5%
4	5364	4507	84.0%
5	4483	3744	83.5%
Average	5133	4298	83.82%

¹ Data from Buote et al. [24].

Figure 4 compares the unfiltered and filtered (green points) (a) x and (b) y coordinates over time collected by our proposed tracking method and the manual tracking method over 2000 frames. The red circles in Figure 4 highlight areas in which the filtered coordinates smoothed by a 5th order moving average filter can improve tracking of the unfiltered coordinates.

Unfiltered and filtered coordinates are plotted against the manual tracking coordinates in Figure 5 with the coefficient of determination. Table 4 outlines the R^2 values per game for both players.

Table 5 compares both players positional statistics obtained from unfiltered and filtered coordinates and Buote et al. [24] results of the same match. These parameters are compared to the results of Buote et al. [24] in Table 6.

Velocity statistics including the average speed data of both players calculated from unfiltered, filtered coordinates and Buote et al. [24] are presented in Table 7. Similar to the positional data, average speeds were compared to Buote et al. [24] and differences were quantified in Table 8.

The average differences and percent error of the player data collected by the filtered coordinates is summarized in Table 9. With consideration that the error of estimation of the position is recommended to be less than the natural balance of the center of gravity of the human body (between 15 and 20 cm) in an observed movement, the average difference for positional data of 17.6 cm (Table 9) is acceptable but can be improved [30].

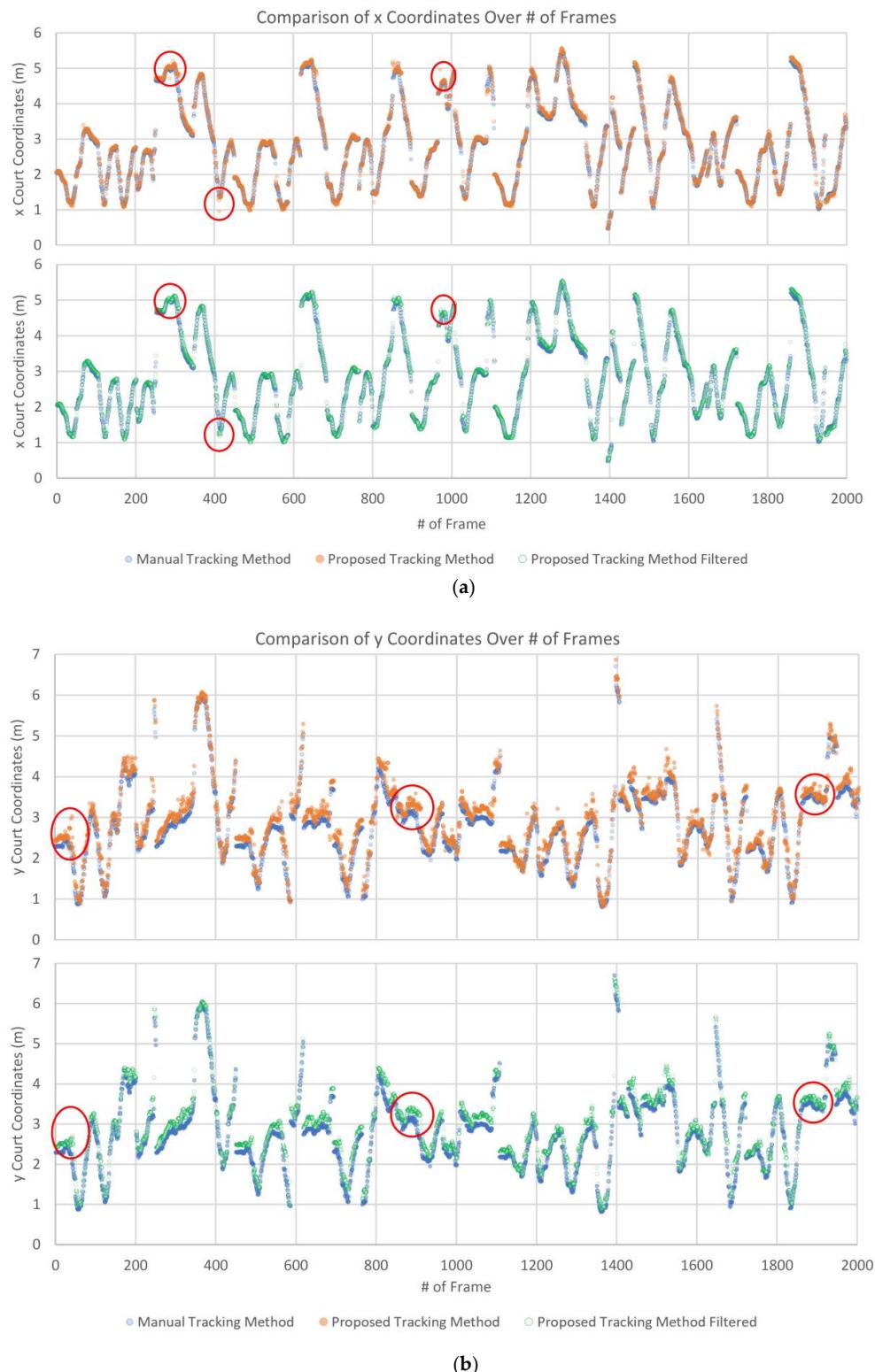


Figure 4. Comparison of Mustonen's Game 5 average feet location on court of unfiltered and filtered (a) x and (b) y coordinates to the manual tracking coordinates. Blue points represent the coordinates collected from the manual tracking method, orange and green points are the unfiltered and filtered coordinates collected from the proposed tracking method, respectively. Red circles highlight significant areas where filtering has improved tracking. Coordinates were taken over the first 2000 frames from 50:09.04–54:47.16 (global timestamp of the broadcasted match).

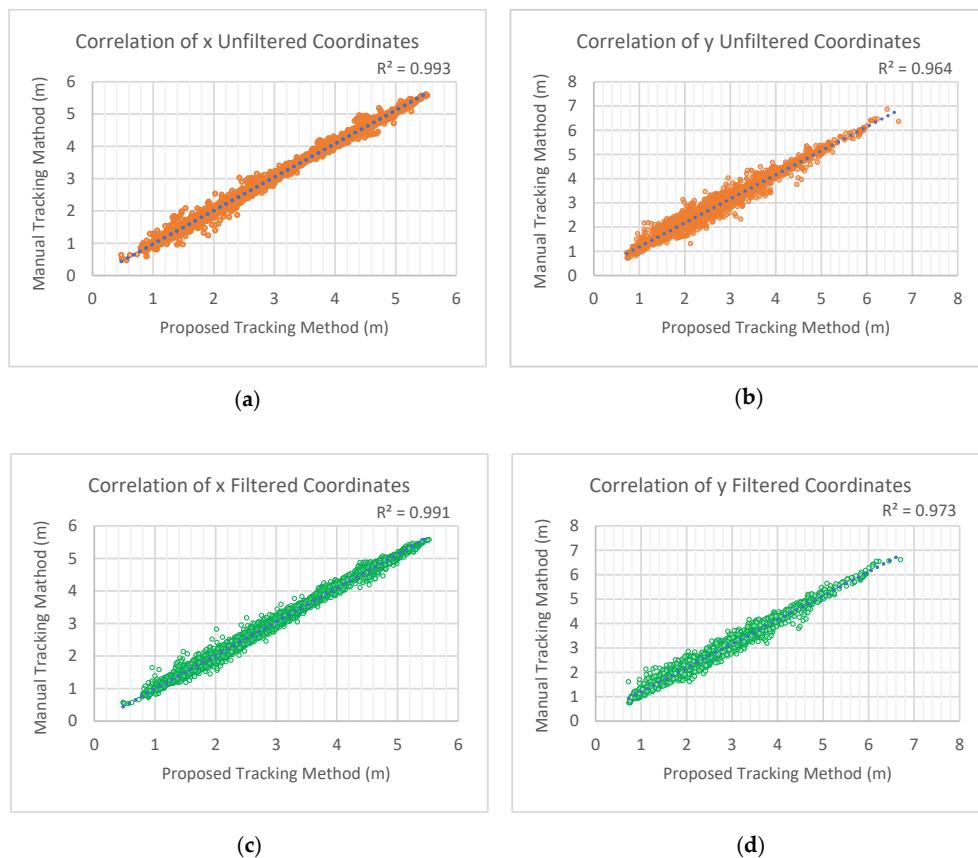


Figure 5. Correlation of Mustonen's Game 5 center of mass on court **(a)** x unfiltered coordinates, **(b)** y unfiltered coordinates, **(c)** y filtered coordinates, and **(d)** y filtered coordinates with calculated R^2 values. For each data point, the x value is the coordinate from the proposed method and the y value is the corresponding coordinate collected from the manual tracking method [24].

Table 4. El Shorbagy and Mustonen's R^2 values correlating average foot position from manual tracking and unfiltered (UNF) and filtered (FIL) coordinates from our proposed tracking method.

Game	x UNF	y UNF	x FIL	y FIL
EL SHORBAGY				
1	0.986	0.965	0.984	0.967
2	0.993	0.980	0.990	0.983
3	0.993	0.980	0.990	0.983
4	0.992	0.963	0.989	0.968
5	0.982	0.954	0.980	0.962
Average	0.989	0.968	0.987	0.973
MUSTONEN				
1	0.982	0.954	0.980	0.961
2	0.992	0.971	0.989	0.976
3	0.992	0.966	0.989	0.971
4	0.993	0.962	0.990	0.965
5	0.993	0.964	0.991	0.973
Average	0.990	0.963	0.988	0.969
OVERALL				
Average	0.990	0.966	0.988	0.971

Table 5. El Shorbagy and Mustonen’s positional statistics from the 2013 Canary Wharf match. Includes player data from unfiltered (UNF), filtered (FIL) coordinates and Buote et al. (REF).

Game	Distance Traveled (m)			Average Radial Distance to T (m)			% Left of T			% Behind T		
	UNF	FIL	REF	UNF	FIL	REF	UNF	FIL	REF	UNF	FIL	REF
EL SHORBAGY												
1	448	380	367	1.70	1.71	1.85	51.8	51.7	53.2	90.3	90.2	92.5
2	469	396	390	1.56	1.66	1.82	51.7	51.8	56.2	81.7	81.6	85.4
3	397	367	395	1.46	1.58	1.69	47.5	47.4	48.1	80.8	80.8	86.9
4	490	391	392	1.42	1.42	1.53	55.8	55.6	59.0	74.1	73.8	80.7
5	437	377	338	1.32	1.48	1.65	61.4	61.3	64.9	80.2	80.3	87.8
Average	448.2	382.2	376.4	1.49	1.57	1.71	53.6	53.6	56.3	81.4	81.3	86.7
MUSTONEN												
1	456	383	383	1.76	1.75	1.90	49.6	49.9	52.8	94.0	94.0	94.8
2	479	383	369	1.65	1.71	1.85	50.8	50.8	53.7	89.8	90.0	91.9
3	439	393	410	1.68	1.78	1.91	47.6	47.7	50.0	92.0	92.2	94.1
4	510	421	410	1.87	1.86	1.98	55.1	55.2	58.7	88.6	88.8	90.0
5	414	349	341	1.57	1.89	2.02	65.8	65.7	68.7	92.2	92.5	92.6
Average	459.6	385.8	382.6	1.71	1.80	1.93	53.8	53.9	56.8	91.3	91.5	92.7

Table 6. Quantifying the percent error between El Shorbagy and Mustonen’s unfiltered (UNF) and filtered (FIL) positional statistics to the statistics collected from Buote et al. [24].

Game	Distance Traveled % Error (%)		Average Radial Distance to T % Error ¹ (%)		% Left of T % Error ¹ (%)		% Behind T % Error ¹ (%)	
	UNF	FIL	UNF	FIL	UNF	FIL	UNF	FIL
EL SHORBAGY								
1	22.13	3.51	-8.11	-7.41	-0.026	-0.028	-0.024	-0.025
2	20.28	1.38	-14.29	-8.74	-0.082	-0.080	-0.043	-0.044
3	0.43	-7.29	-13.61	-6.21	-0.012	-0.015	-0.070	-0.070
4	25.13	-0.26	-7.19	-7.39	-0.054	-0.058	-0.082	-0.086
5	29.38	11.72	-20.00	-10.06	-0.054	-0.054	-0.087	-0.085
MUSTONEN								
1	18.85	-0.10	-7.37	-7.74	-0.061	-0.055	-0.009	-0.009
2	29.78	3.90	-10.81	-7.78	-0.054	-0.054	-0.023	-0.021
3	6.98	-4.34	-12.04	-7.07	-0.048	-0.046	-0.022	-0.020
4	24.29	2.59	-5.56	-6.06	-0.061	-0.058	-0.016	-0.013
5	21.26	2.23	-22.28	-6.24	-0.042	-0.044	-0.003	-0.001
OVERALL								
Max Absolute% Error¹ (%)	29.78	11.72	22.28	10.06	0.082	0.080	0.087	0.086
Average Absolute% Error¹ (%)	19.85	3.73	12.12	7.47	0.049	0.049	0.038	0.038

¹ % Error is calculated as [(unfiltered/filtered statistic–reference statistic)/reference statistic] · 100% where the reference statistic is the statistic collected from Buote et al. [24].

Table 7. El Shorbagy and Mustonen's velocity statistics from the 2013 Canary Wharf match and % of points excluded. Includes unfiltered (UNF), filtered (FIL) coordinates and Buote et al. (REF) data.

Game	Average Speed excluding >9 m/s (m/s)			% of Points >9 m/s			Average Speed excluding >9 m/s and <1 m/s (m/s)			% of Points >9 m/s and <1 m/s		
	UNF	FIL	REF	UNF	FIL	REF	UNF	FIL	REF	UNF	FIL	REF
EL SHORBAGY												
1	2.11	1.77	1.73	2.19	1.03	0.33	2.83	2.36	2.40	32.1	33.3	34.1
2	2.33	1.97	1.93	3.58	1.52	0.51	2.98	2.47	2.50	28.8	27.5	28.7
3	2.09	1.83	1.77	1.75	1.35	0.11	2.83	2.41	2.38	32.3	32.2	31.5
4	2.36	1.88	1.79	2.89	1.02	0.43	3.00	2.42	2.41	27.5	30.0	32.0
5	2.22	1.99	1.88	3.39	1.63	0.27	3.02	2.52	2.52	32.7	28.6	31.0
Average	2.22	1.89	1.82	2.76	1.31	0.33	2.93	2.44	2.44	30.7	30.3	31.5
MUSTONEN												
1	2.13	1.81	1.82	2.17	1.24	0.23	2.88	2.34	2.41	32.1	30.6	30.8
2	2.35	1.91	1.93	3.67	1.01	0.36	3.01	2.46	2.50	28.4	29.8	31.7
3	2.22	1.93	1.85	2.96	1.62	0.09	2.90	2.47	2.41	30.2	29.5	29.5
4	2.39	1.96	1.88	3.65	1.39	0.47	3.00	2.45	2.48	26.8	27.5	30.4
5	2.12	1.98	1.91	2.30	1.07	0.13	3.12	2.54	2.51	37.0	29.2	29.6
Average	2.24	1.92	1.88	2.95	1.27	0.26	2.98	2.45	2.46	30.9	29.3	30.4

Table 8. Quantifying the percent error between El Shorbagy and Mustonen's unfiltered (UNF) and filtered (FIL) velocity statistics to the statistics collected from the manual tracking method.

Game	Average Speed excluding >9 m/s % Error ¹ (%)		Average Speed excluding >9 m/s and <1 m/s % Error ¹ (%)	
	UNF	FIL	UNF	FIL
EL SHORBAGY				
1	21.85	1.85	17.67	-1.75
2	21.19	2.18	19.04	-1.48
3	17.80	3.33	18.74	1.09
4	32.01	5.20	24.61	0.66
5	18.24	6.01	20.00	0.04
MUSTONEN				
1	17.03	-0.11	19.71	-2.74
2	26.63	4.15	21.96	-0.04
3	20.27	4.54	20.25	2.12
4	26.81	3.88	20.69	-1.41
5	11.10	3.98	24.50	1.31
OVERALL				
Max Absolute% Error¹ (%)	32.01	6.01	24.61	2.74
Average Absolute% Error¹ (%)	21.29	3.52	20.72	1.26

¹ % Error is calculated as [(unfiltered/filtered statistic–reference statistic)/reference statistic] ·100% where the reference statistic is the statistic collected from Buote et al. [24].

Table 9. Summary of average absolute percent error comparing player data collected from filtered coordinates (smoothed using a 5th order moving average filter) and results of Buote et al. from the 2013 Canary Wharf match [24].

Player Data	Average Difference ^{1,3} of Filtered Coordinates (% Error ^{2,3})
Position (m)	0.176 (6.75%)
Distance Traveled (m)	4.3 (3.73%)
Average Radial Distance to the T (m)	-0.135 (7.47%)
% Left of T	-0.028 (0.049%)
% Behind of T	-0.033 (0.038%)
Average Speed excluding > 9 m/s (m/s)	0.065 (3.52%)
Average Speed excluding > 9 m/s and <1 m/s (m/s)	-0.005 (1.26%)

¹ Difference is calculated as unfiltered/filtered statistic–reference statistic. ² % Error is calculated as [(unfiltered/filtered statistic–reference statistic)/reference statistic] ·100%. ³ Reference statistic is the statistic collected from Buote et al. [24].

4. Discussion

This study aims to apply deep learning and computer vision processes to evaluate kinematics of elite squash players for the first time. The method was validated when compared with previous results from a manual tracking study [24]. Our method presents many advantages to prior data acquisition methods. The ability to analyze any matches filmed by the PSA or suitable matches filmed from a similar angle, requiring no special camera setup or wearable markers that could impede player movement, significantly increases the number of elite matches eligible for analysis.

A notable advancement in the present study is the speed of player tracking, which has been considerably accelerated to 0.3 s per frame. Player statistics are rapidly generated using Python code for easy computation. Thus, an ideal full match analysis takes approximately 3 h including tracking and analysis, where the majority of the process is autonomous. Presently, manual intervention is only required during pre-processing to identify active play. Broadcasts of professional squash matches do not have a definitive visual or auditory indicator of when a rally begins or ends, unlike other racquet sports. Based on our preliminary investigation, some strategies that could be implemented in the future to address this include tracking when the scoreboard changes, noting a change of the camera angle or pan away from the court (note that these implementations will not be completely instantaneous).

For the match analyzed, active play was slightly higher than half the total time of each game (55.9% on average). This supports the interpretation of squash being a sport demanding of short, high intensity bursts rather than endurance and constant intensity [31]. Other camera angles such as the sidewall and close up secondary cameras do not display both players and are typically used for repetitive shots, usually drop shots or backhands down the wall from the left back corner. However, the movement of the players were cyclical between the T and the corner and deemed to be relatively equal, providing valid results for comparison and aligning with previous studies [17,20,21,24]. Current work is being done to implement autonomous collection of other frame angles of the match. Future work can be done establishing court conversion matrices using the inverse perspective mapping method used on the main camera angle and other camera angles to analyze the full length of active match play [32,33].

Frames analyzed by the manual tracking method were used with the proposed tracking method with an average of 83.82% of the input frames per game, resulting in detection and identification of both players. Frames where the system was unable to successfully detect players were due to player occlusion or unnatural pose as mentioned previously in the methods section. A global timestamp was assigned and detected in each frame to account for difference in time when calculating for velocity between missing frames. Comparison of time series dependent results such as player velocities provide support for the effectiveness of this approach.

Court conversions were determined using reference points in the frame specific to the court and camera angle. The equations have been noted to be more accurate in predicting a player's position

near the T than the top corners, likely due to the distribution of reference points having a higher concentration in the center (service lines) compared to the corners [24]. The raw position coordinates were smoothed due to the variation of foot detection (described in the methods section) using a 5th order moving average filter. The R^2 values calculated displayed a slightly higher accuracy for the x unfiltered coordinates compared to the x filtered coordinates (0.990 and 0.988, respectively) (Table 4). Further, the y filtered coordinates were noted to be considerably higher than the y unfiltered coordinates (0.971 and 0.966, respectively) (Table 4). This indicates that the accuracy of the system depends on the margin of error of the y coordinates. This is due to the dimensions of a squash court where the length is longer than the width. Because of the camera angle perspective, the video image produces a court that is compressed lengthwise and is wider at the bottom (back wall) compared to the top (front wall), causing y coordinates to have a larger margin for error during detection.

The filtered coordinates displayed more reliability for cumulative statistics such as total distance with an average percent error of 3.73% compared to the unfiltered coordinates with an average percent error of 19.85% (Table 6). The variation in foot detection with the proposed method resulted in larger changes in coordinate position between frames compared to the manual tracking method, which resulted in consistently higher values for total distance traveled. Filtering was able to remove the problematic fluctuations, resulting in total distance traveled values that were closer to the manually measured values. This is especially evident in Game 3, where both players have the lowest total distance percent error of 0.43% (El Shorbagy) and 6.98% (Mustonen) (Table 6) when compared to the rest of the games. Game 3 also has the lowest number of frames collected at 76.50% as opposed to the average number of frames collected at 85.65% (without Game 3), supporting the need to filter the raw coordinates. Like previous studies, it appears players travel similar distances as their opponent in each individual game and distances traveled can be correlated to the length of game [20].

Vučković et al. [21] suggested that the dominance of a rally can be indicated by the time spent near the T, except for closely contested games. This is in agreement with our results as the winner and higher ranked player of the match, El Shorbagy (1.49 m for unfiltered coordinates, 1.57 m for filtered coordinates, and 1.71 m according to Buote et al.) maintained a smaller average radius to the T than Mustonen (1.71 m for unfiltered coordinates, 1.80 m for filtered coordinates, and 1.93 m according to [24]). This is reflective of common squash tactics where skilled players play accurate shots to force their opponent to leave the T area, while less skilled players play a greater number of shots closer to the center of the court [21,24].

Players spent an average of 53.7% (unfiltered and filtered coordinates) of the time on the left side of the T which concurs with the findings of 56.5% from Buote et al. [24]. Since the left side wall camera view was not analyzed, these percentages are expected to be higher. This aligns with Vučković et al. [34] who recorded an average of 64.6% of shots coming from the left side of the court for 10 matches played at the men's World Team Championship in 2003. As both players were right-handed, a higher percentage spent on the left (backhand) side was expected since at the elite level, a common tactic is to play to your opponent's backhand which is considered weaker and more difficult [24]. An overwhelming 86.4% (unfiltered and filtered coordinates) of the time was spent behind the T, agreeing with the manual tracking average of 89.7% from Buote et al. [24]. This is similar to the previous studies of Vučković et al. who found 74.5% of shots coming from behind the T at the same 10 matches recorded at the men's World Team Championship in 2003 as stated above [34]. The tendency to favor the left and situate yourself behind the T typically occurs when a player returns to center to anticipate the next shot. The lower percentages calculated using the proposed method compared to the reference is likely since most missing frames due to player occlusion occur near the T during their return to the ideal position.

The average speeds calculated by the filtered coordinates (overall 1.90 m/s) are much closer to Buote et al.'s results (1.85 m/s) [24] than the unfiltered coordinates (2.23 m/s). This supports the need for filtering of coordinates and is once again likely due to the variation in foot detection, causing increased distance traveled and in turn higher reported speeds between consecutive frames. The results of the

filtered coordinates align with previous studies where Buote et al. [24] recorded a maximum average speed of 2.04 m/s over 5 matches of elite players from 2012–2014 and Hughes and Franks [17] recorded a maximum mean speed of 1.98 m/s, while the maximum average speed was 1.99 m/s using filtered coordinates. As the average walking speed is around 1.4 m/s and the walk-to-run transition speed has noted to occur below 2 m/s, our results of 1.90 m/s as the overall average speed reflects the idea that squash comprises of shifts between walking and running [35–38].

Removing speeds below 1 m/s is argued by Buote et al. [24] to provide a more realistic idea of how fast players move to return shots. Speeds that fall under 1 m/s are primarily identified when a player is at center court waiting for their opponent to play a shot, during the pause for accuracy and power before a player makes their shot, and when players change directions. With this selection, our results show that for 70.2% of the time during active match play, players moved at an average speed of 2.44 m/s and only spent 29.8% of the time moving less than 1 m/s. This is reflective of Buote et al.'s analysis of 5 matches [24] as mentioned above, which found the mean speed of players as 2.52 m/s (excluding speeds less than 1 m/s), 69.6% of the time during active match play. These speeds represent the incredible level of conditioning and endurance elite squash players must possess to compete.

A limitation of this study is the inability to analyze the entirety of active match play (83.32% analyzed on average, Table 3). Another constraint is the assumption that players slide horizontally across the plane of the court when converting video coordinates into court coordinates, meaning that vertical movement of a player due to jumping is considered as distance traveled. In addition, the conversion does not take into account any lens warping. Our future work will focus on continuing to develop the reliability of this method, add the analysis of additional camera angles, refine the model to reduce/handle missing frames, and to gather data on recent PSA matches. Further research opportunities include analysis of upper body and arm kinematics.

5. Conclusions

With the increasing availability and access to broadcasted elite squash matches, our study utilizes recent advancements in human pose estimation and computer vision to quantify squash kinematics and tactics using video analysis. This method offers the ability to analyze any PSA match or any match filmed similarly and suitably, giving access to a large collection of elite player data to be analyzed for the first time. It is also entirely autonomous apart from selecting active match play.

Our results support previously identified elite squash tactics and strategy in former studies. This methodology has proven to be accurate and reliable in comparison to results of a manual tracking method [24]. It is also the most effective in collecting kinematic data with no special camera setup, limited manual intervention, and has a clear advantage in its ability to provide analysis in a timely manner.

Author Contributions: Conceptualization, M.M.B. and Q.M.; Formal analysis, M.M.B., N.E. and Q.M.; Funding acquisition, S.A. and L.W.; Investigation, M.M.B., N.E. and Q.M.; Methodology, M.M.B., N.E. and Q.M.; Project administration, Q.M., M.G., S.A. and L.W.; Resources, M.G., S.A. and L.W.; Software, M.M.B., N.E. and Q.M.; Supervision, Q.M., M.G., S.A. and L.W.; Validation, M.M.B., N.E. and Q.M.; Visualization, M.M.B. and N.E.; Writing – original draft, M.M.B. and N.E.; Writing – review & editing, M.M.B., N.E., Q.M., M.G., S.A. and L.W.. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the Natural Sciences and Engineering Research Council of Canada (NSERC).

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Bezodis, N.E.; Salo, A.I.T.; Trewartha, G. Relationships between lower-limb kinematics and block phase performance in a cross section of sprinters. *Eur. J. Sport Sci.* **2015**, *15*, 118–124. [[CrossRef](#)] [[PubMed](#)]
2. Ford, K.R.; Myer, G.D.; Toms, H.E.; Hewett, T.E. Gender differences in the kinematics of unanticipated cutting in young athletes. *Med. Sci. Sports Exerc.* **2005**, *37*, 124–129. [[CrossRef](#)] [[PubMed](#)]

3. Devita, P.; Hortobagyi, T.; Barrier, J. Gait biomechanics are not normal after anterior cruciate ligament reconstruction and accelerated rehabilitation. *Med. Sci. Sports Exerc.* **1998**, *30*, 1481–1488. [[CrossRef](#)] [[PubMed](#)]
4. Atha, J. Current techniques for measuring motion. *Appl. Ergon.* **1984**, *15*, 245–257. [[CrossRef](#)]
5. Sanderson, F. A notational system for analysing squash. *Phys. Educ. Rev.* **1983**, *6*, 19–23.
6. Sanderson, F.; Way, K. The development of an objective method of game analysis in squash rackets. *Br. J. Sports Med.* **1977**, *11*, 188. [[CrossRef](#)]
7. Richards, J.G. The measurement of human motion: A comparison of commercially available systems. *Hum. Mov. Sci.* **1999**, *18*, 589–602. [[CrossRef](#)]
8. Growney, E.; Meglan, D.; Johnson, M.; Cahalan, T.; An, K.N. Repeated measures of adult normal walking using a video tracking system. *Gait Posture* **1997**, *6*, 147–162. [[CrossRef](#)]
9. Cappozzo, A.; Catani, F.; Leardini, A.; Benedetti, M.G.; Della Croce, U. Position and orientation in space of bones during movement: Experimental artefacts. *Clin. Biomech.* **1996**, *11*, 90–100. [[CrossRef](#)]
10. Corazza, S.; Mündermann, L.; Chaudhari, A.M.; Demattio, T.; Cobelli, C.; Andriacchi, T.P. A markerless motion capture system to study musculoskeletal biomechanics: Visual hull and simulated annealing approach. *Ann. Biomed. Eng.* **2006**, *34*, 1019–1029. [[CrossRef](#)]
11. Ceseracciu, E.; Sawacha, Z.; Cobelli, C. Comparison of markerless and marker-based motion capture technologies through simultaneous data collection during gait: Proof of concept. *PLoS ONE* **2014**, *9*, e87640. [[CrossRef](#)] [[PubMed](#)]
12. Sandau, M.; Koblauch, H.; Moeslund, T.B.; Aanæs, H.; Alkjær, T.; Simonsen, E.B. Markerless motion capture can provide reliable 3D gait kinematics in the sagittal and frontal plane. *Med. Eng Phys.* **2014**, *36*, 1168–1175. [[CrossRef](#)] [[PubMed](#)]
13. Ong, A.; Harris, I.S.; Hamill, J. The efficacy of a video-based marker-less tracking system for gait analysis. *Comput. Methods Biomed. Eng.* **2017**, *20*, 1089–1095. [[CrossRef](#)]
14. Perš, J.; Kovačič, S. A system for tracking players in sport games by computer vision. *Electrotech, Rev. J. Electr. Eng. Comput. Sci.* **2000**, *5*, 281–288.
15. Santiago, C.; Sousa, A.; Reis, L.; Estriga, M. Real time colour based player tracking in indoor sports. In *Computational Vision and Medical Image Processing*; CRC Press: Cleveland, OH, USA, 2010; pp. 17–35.
16. Tahan, O.; Rady, M.; Sleiman, N.; Ghantous, M.; Merhi, Z. A computer vision driven squash players tracking system. In Proceedings of the 2018 19th IEEE Mediterranean Electrotechnical Conference (MELECON), Marrakech, Morocco, 2–7 May 2018; pp. 155–159.
17. Hughes, M.; Franks, I. Dynamic patterns of movement of squash players of different standards in winning and losing rallies. *Ergonomics* **1994**, *37*, 23–29. [[CrossRef](#)] [[PubMed](#)]
18. Franks, I.; Sinclair, G.; Thomson, W.; Goodman, D. Analysis of the coaching process. *Sci. Period. Res. Technol. Sport* **1986**, *1*, 1–12.
19. Hughes, M.; Franks, I.M.; Nagelkerke, P. A video-system for the quantitative motion analysis of athletes in competitive sport. *J. Hum. Mov. Stud.* **1989**, *17*, 217–227.
20. Vučković, G.; Dežman, B.; Erčulj, F.; Kovačič, S.; Perš, J. Comparative movement analysis of winning and losing players in men's elite squash. *Kinesiol. Slov.* **2003**, *9*, 74–84.
21. Vučković, G.; Perš, J.; James, N.; Hughes, M. Tactical use of the T area in squash by players of differing standard. *J. Sports Sci.* **2009**, *27*, 863–871. [[CrossRef](#)]
22. Vučković, G.; Perš, J.; James, N.; Hughes, M. Measurement error associated with the SAGIT/Squash computer tracking software. *Eur. J. Sport Sci.* **2010**, *10*, 129–140. [[CrossRef](#)]
23. Perš, J.; Vučković, G.; Kovačič, S.; Dežman, B. A low-cost real-time tracker of live sport events. In Proceedings of the 2nd International Symposium on Image and Signal Processing and Analysis in Conjunction with 23rd International Conference on Information Technology Interfaces, Pula, Croatia, 19–32 June 2001; pp. 362–365.
24. Buote, K.; Jomha, N.; Adeeb, S. Quantifying motion and 'T' dominance of elite squash players. *J. Sport Hum. Perf.* **2016**, *4*, 1–19.
25. Cao, Z.; Simon, T.; Wei, S.; Sheikh, Y. Realtime multi-person 2D pose estimation using part affinity fields. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 22–25 July 2017; pp. 7291–7299.

26. Andriluka, M.; Pishchulin, L.; Gehler, P.; Schiele, B. 2D human pose estimation: New benchmark and state of the art analysis. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA, 23–28 June 2014; pp. 3686–3693.
27. Ling, T.Y.; Maire, M.; Belongie, S.; Bourdev, L.; Girshick, R.; Hays, J.; Perona, P.; Ramanan, D.; Zitnick, C.L.; Dollár, P. Microsoft COCO: Common objects in context. In Proceedings of the European Conference on Computer Vision ECCV 2014: Computer Vision—ECCV 2014, Zurich, Switzerland, 6–12 September 2014; pp. 740–755.
28. World Squash. Court Specifications. 12 March 2019. Available online: <https://www.worldsquash.org/court-construction/> (accessed on 6 November 2020).
29. McGarry, T.I.M.; Khan, M.A.; Franks, I.M. On the presence and absence of behavioural traits in sport: An example from championship squash match-play. *J. Sports Sci.* **1999**, *17*, 297–311. [CrossRef] [PubMed]
30. Leser, R.; Baca, A.; Ogris, G. Local Positioning Systems in (Game) Sports. *Sensors* **2011**, *11*, 9778–9797. [CrossRef] [PubMed]
31. Girard, O.; Chevalier, R.; Habrard, M.; Sciberras, P.; Hot, P.; Millet, G.P. Game analysis and energy requirements of elite squash. *J. Strength Cond. Res.* **2007**, *21*, 909–914. [CrossRef]
32. Heckbert, P. Fundamentals of Texture Mapping and Image Warping. Master’s Thesis, University of California, Berkley, CA, USA, June 1989. Available online: <http://www2.eecs.berkeley.edu/Pubs/TechRpts/1989/CSD-89-516.pdf> (accessed on 2 November 2020).
33. Bloomenthal, J.; Rokne, J. Homogeneous coordinates. *Vis. Comput.* **1994**, *11*, 15–26. [CrossRef]
34. Vučković, G.; James, N.; Hughes, M.; Murray, S.; Sporiš, G.; Perš, J. The effect of court location and available time on the tactical shot selection of elite squash players. *J. Sports Sci. Med.* **2013**, *12*, 66.
35. Mohler, B.J.; Thompson, W.B.; Creem-Regehr, S.H.; Pick, H.L., Jr.; Warren, W.H., Jr. Visual flow influences gait transition speed and preferred walking speed. *Exp. Brain Res.* **2007**, *181*, 221–228. [CrossRef]
36. Raynor, A.J.; Yi, C.J.; Abernethy, B.; Jong, Q.J. Are transitions in human gait determined by mechanical, kinetic or energetic factors? *Hum. Mov. Sci.* **2002**, *21*, 785–805. [CrossRef]
37. Kram, R.; Domingo, A.; Ferris, D.P. Effect of reduced gravity on the preferred walk-run transition speed. *J. Exp. Biol.* **1997**, *200*, 821–826.
38. Colyer, S.L.; Evans, M.; Cosker, D.P.; Salo, A.L. A review of the evolution of vision-based motion analysis and the integration of advanced computer vision methods towards developing a markerless system. *Sports Med. Open* **2018**, *4*, 24. [CrossRef]

Publisher’s Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).