Sheffield Hallam University | Centre for Health and Social Care Research

Sheffield Hallam University | Centre for Sports Engineering Research

Sheffield Hallam University | Materials and Engineering Research Institute

Establishing the accuracy and feasibility of Microsoft Kinect in various multi-disciplinary contexts

Project Report

Jon Wheat[1], Reuben Fleming[2], Maria Burton[3], Jacques Penders[2], Simon Choppin[1], Ben Heller[1]

[1] Centre for Sports Engineering Research, Sheffield Hallam University, UK
[2] Materials and Engineering Research Institute, Sheffield Hallam University, UK
[3] Centre for Health and Social Care Research, Sheffield Hallam University, UK

EPSRC

ENGINEERING FOR LIFE
Enhancing people's lives

Contents

1 Introduction **2**

2 Stage 1 - Identifying potential application areas of the Kinect **3**

2.1 Application areas . . . 3

3 Stage 2 **6**

3.1 Introduction . . . 6

3.2 Investigation 1: Segment tracking . . . 6

3.2.1 Introduction . . . 6

3.2.2 Methods . . . 8

3.2.3 Results . . . 9

3.2.4 Discussion . . . 10

3.3 Investigation 2: Three-dimensional scanning . . . 14

3.3.1 Introduction . . . 14

3.3.2 Methods . . . 15

3.3.3 Results . . . 15

3.3.4 Discussion . . . 15

3.4 Investigation 3: Whole body tracking . . . 18

3.4.1 Introduction . . . 18

3.4.2 Experiment 1: Laboratory based testing . . . 19

3.4.3 Experiment 2: On-court testing . . . 25

3.4.4 Overall Discussion . . . 30

4 Stage 3 **35**

4.1 Investigation 1: Segment tracking . . . 36

4.2 Investigation 2: Three-dimensional scanning . . . 36

4.3 Investigation 3: Whole body tracking . . . 36

5 Stage 4 **37**

6 Outcomes **37**

1 Introduction

Engineering for Life recently agreed to fund seed corn work exploring the research applications of the Microsoft Kinect. The aim of the project was to establish the feasibility and accuracy of the Microsoft Kinect in various contexts related to potential funding applications and explore potential ways in which the accuracy and applicability of the data can be improved. Further, the objectives were to:

- Confirm potential applications of the Kinect - aligned with multi-disciplinary/cross research centre funding applications - on which to focus during the project;
- Establish the accuracy of the data from the Kinect in the identified contexts;
- If the accuracy in some contexts is not appropriate, explore ways in which further/post processing using discipline-specific analysis techniques can be used to improve the accuracy;
- Explore the practicalities and feasibility of using the Kinect in the identified contexts;
- Publish the results of the accuracy analyses;
- Generate data with which to inform future multi-disciplinary, cross research centre funding applications.

The project was organised into four main stages. Stage one involved steering group meetings in which all members of the project team explored and highlighted potential application areas of the Kinect. During stage two, issues of accuracy and feasibility related to the identified application areas were explored via several accuracy investigations. Stage three involved a further steering group meeting. The purpose of this meeting was to - in the context of the applications identified for the use of Kinect - interpret the results of the accuracy studies conducted in stage two. Interdisciplinary discussions at these meetings also considered how the data analysis techniques/algorithms used during stage two might be improved. Stage four involved preliminary work exploring the feasibility of these improvements.

2 Stage 1 - Identifying potential application areas of the Kinect

Work in stage 1 involved two steering group meetings at which all members of the project team were present. The first meeting (held on Monday 5^{th} September 2011) involved general introductory discussion - to the project, each other and our research interests. All members of the project team left the meeting with an understanding of the project aims, objectives and stages. Furthermore, initial discussions of potential multidisciplinary projects at the meeting generated ideas which the project team agreed to develop further before the second steering group.

The purpose of the second steering group meeting held on Tuesday 27^{th} September 2011, was to confirm 2-4 project areas that might lead to future multi-disciplinary grant applications on which to focus during the remainder of the project.

2.1 Application areas

Discussions across the two steering group meetings resulted in *five* potential application areas of the Kinect on which we proposed to focus for the remainder of the project:

1. **Segment Tracking**: The Kinect is capable of tracking the orientation and position of the body segments of a user (e.g. thigh, upper arms etc.). As such, it is possible to acquire data similar to that previously only obtained from measurement systems that are several orders of magnitude more expensive (the motion capture system in the Biomechanics Lab, Collegiate Hall would cost in the order of £100,000). Potentially, this could have a profound effect on motion capture and analysis work across many disciplines. Most relevant would seem to be applications that involve taking measurements out of the traditional laboratory setting into more ecologically valid environments, such as care homes, sports venues or the classroom.

   The use of the Kinect in this way is relevant to many potential grant applications in diverse areas including health, sport, robotics and computer animation. For example, the Kinect could be used to track upper limbs movements of neurological patients during recovery. This type of application would build on previous projects in the University (SMART I and II) in which upper limb movements in patients recovering from stroke were monitored in the home. Indeed, use of the Kinect for this purpose would address several issues with the inertial sensors used in the previous projects for upper limb tracking - e.g. problems donning and doffing the sensors and a lack of engagement with the avatar used for visual feedback. Furthermore, there are many applications - related to, for example, gait analysis that would also be possible when the tracking is extended to the lower body. Projects in these areas would be inherently interdisciplinary as experts from many disciplines (e.g. health, biomechanics, engineering, robotics and computer animation) would be required to monitor, interpret and provide feedback to patients (and athletes in more sport related applications).

2. **Person Tracking**: It is also possible to use the Kinect to track a participant in a slightly less sophisticated way than that described in 1). The point clouds (clusters of three-dimensional points representing a person/object in the field of view) produced by the Kinect can be analysed to track, for example, the centre of mass of the user - without the need for tracking individual body segments. This approach could be used to track people moving through environments or track the location of sports people on courts or pitches. With slightly more sophisticated analyses, the approach could be extended to track specific regions of the body. A potentially very useful example of this would be tracking the feet during gait. This would be relatively simple to achieve and offer considerable benefits over existing technologies. An example of such technology, owned by the Faculty of Health and Wellbeing at SHU, is the GAITRite system. This technology - which takes the form of a long pressure sensing mat - reports various gait analysis parameters (including stride/step length, stride/step time, stride width, walking speed, stance time, swing time, toe-in/toe-out angle) reported to be useful in predicting fall risk in older adults, for example. A system could be developed using the Kinect that could take these measures using hardware costing orders of magnitude less than systems such as GAITRite. Moreover, the added information available from the Kinect would allow additional important gait analysis parameters to be calculated. For example, the foot could be tracked throughout the entire gait cycle - rather than just when the foot is in contact with the ground - allowing foot clearance to be measured and, potentially, identify pathalogical gait adaptations such as ‘vaulting’ and ‘circumduction’. Any such system would be extremely portable and very affordable. Potentially, a more accessible gait analysis system would

allow more clinics, both in and outside hospitals to take important quantitative gait measurements. With data on the accuracy of a suitable measurement system, there are many opportunities for funding applications in this area.

3. **Gesture recognition**: Microsoft refer to the Kinect as a 'Natural User Interface Device', with human computer interaction its core application. As such, the Kinect is suited to gesture recognition making it possible to drive software as well as manipulate, and interact with, virtual worlds though whole-body movements. It has been postulated that this type of interaction might increase physical activity in older and pathological people and, through socialisation in virtual worlds, reduce potential social isolation in these populations. Several gesture recognition 'layers' (e.g. FAAST, University of California, USA) have been developed to sit on top of the application interfaces released for the Kinect. However, there is no indication of the reliability/repeatability with which the gestures can be recognised. Reliable and repeatable gesture recognition would have many applications with close alignment to various health-related funding bids.

4. **Virtual presence**: By processing the data from the Kinect, it is possible to combine the information from the depth and rgb cameras to produce a '3D video rendering' of people/objects in the field of view. This has the advantage of allowing an object/person in the scene to be viewed from many different points, which might lead to a richer and more immersive experience when viewing the animation. Commercial companies are already using this technology (not specifically the Kinect but similar depth camera technology) to develop golf coaching software tools (see, for example, http://www.gurutrainingsystems.com).

5. **Scanning**: The depth camera in the Kinect - and the point cloud data it can provide - make it potentially useful as a three-dimensional (3D) scanner. As such, the Kinect could be used to obtain 3D geometries of objects. Such geometry data can be used in a variety of ways related to many health-related funding bids. For example, 3D scans of the breast have been used to inform breast reconstruction surgery and monitor outcomes. Three-dimensional scans have also been used to define body morphology and health risk by enabling the calculation of a volume and circumference measure-

ments to supplement traditional - but sometimes misleading - body mass index measures. Furthermore, 3D geometries obtained from 3D scanners have also been used to aid the identification of structural conditions such as scoliosis as well as the calculation of body segment inertia parameters - used in many biomechanical analyses such as inverse dynamics which can be employed to estimate loads acting on bones and soft tissue in the body. The 3D scanners used in all of these applications are generally very expensive, limiting their wide spread use. A Kinect-based scanner could potentially offer a cheap scanning system with - reduced but - acceptable accuracy for many applications. Such a system could offer the possibility of more widespread use of 3D scanning methods, with many associated funding applications.

3 Stage 2

3.1 Introduction

During the steering groups in stage 1 of the project, five application areas for the Kinect were identified: segment tracking, person tracking, gesture recognition, virtual presence and three-dimensional scanning. It was not possible, in the time-scale of the project, to explore the accuracy and feasibility issues related to the use of the Kinect in all of these contexts. As such, work during stage 2 focussed primarily on three areas: segment tracking, three-dimensional scanning and whole-body person tracking applications. Details of the accuracy/feasibility investigations completed related to these areas are provided in this section.

3.2 Investigation 1: Segment tracking

3.2.1 Introduction

The Kinect is capable of tracking the position and orientation of the body segments of a user (e.g. thigh, upper arms etc.). As such, it is possible to acquire data similar to that previously only obtained from measurement systems that are several orders of magnitude more expensive (the motion capture system in the Biomechanics Lab at Sheffield Hallam University would cost in the order of £125,000). Potentially, this could have a profound effect on motion capture and analysis work across many disciplines. Most relevant would seem applications that involve taking measurements out of the traditional laboratory setting into more ecologically valid environments, such as care homes, sports venues or the classroom.

The use of the Kinect in this way is relevant to many potential grant applications in diverse areas including health, sport, robotics and computer animation. For example, the Kinect could be used to track the upper limb movements of neurological patients during recovery. This type of application would build on previous projects in the University (SMART I: Heath and Social Care Research Centre) in which upper limb movements in patients recovering from stroke were monitored in the home. Indeed, use of the Kinect for this purpose would address several issues with the inertial sensors used in the previous projects for upper limb tracking - e.g. problems donning and doffing the sensors and a lack of engagement with the avatar used for visual feedback. Furthermore, there are many applications - related to, for example, gait analysis that would also be possible when the tracking is extended to the lower body. Projects in these areas would be inherently interdisciplinary as experts from many disciplines (e.g. health, biomechanics, engineering, robotics and computer animation) would be required to monitor, interpret and provide feedback to patients (and athletes in more sport related applications).

Currently, there are three main approaches to obtaining kinematic data from the Kinect, which differ based on the software and algorithms used. The developers of the depth camera within the Kinect - Primesense - make freely available their NITE middleware which provides position and orientation data for each segment in a 15 segment human body model. Alternative third-party commercial software (IPIsoft: cost approximately £500), offers the possibility of tracking the position and orientation of up to 19 body segments. Furthermore, IPIsoft allows two Kinects to be used in the collection of the raw point cloud data which could, ostensibly, serve to increase the accuracy of body segment tracking. Finally, Microsoft offer their own body segment tracking algorithms but, at the time of testing, only the position of the joints of the body - rather than the 6 degrees of freedom, position and orientation of the segments of the model - are returned. As such, the algorithms provided by Microsoft were not included in this study. After the development of appropriate software tools for the collection and analysis of the data from the Kinect, the purpose of this study was to explore the accuracy of body segment kinematic estimates obtained using both freely available (NITE) and commercial (IPIsoft) Kinect tracking algorithms.

3.2.2 Methods

Participants After ethical approval from the Faculty of Health and Wellbeing Ethics Committee, ten participants volunteered to participate in the study. Written informed consent was obtained before data collection began. All participants were free from injury that might limit their performance of the movements required during testing.

Experimental Setup A 12 digital-camera motion capture system (Motion Analysis Corporation - MAC, Santa Rosa, CA, USA) sampling at 60 Hz, with cameras setup in optimal positions around a 2 x 2 x 2.5 m capture volume, was used as ground truth. The MAC system required 28 retro-reflective markers to be attached to the body at relevant anatomical landmarks before a static calibration trial, during which the participant stood in the anatomical position for approximately 2 seconds. After the static trial, 5 markers were removed leaving 23 markers for the dynamic trials. The NITE and IPIsoft tracking algorithms imposed different requirements on the positioning of the Kinects. Therefore, the data capture process was performed twice. For the NITE capture a single Kinect was placed directly in front of the participant (approximately 2 m away). When collecting data for use with the IPIsoft algorithms, a second kinect was positioned approximately 70° to the participant's right of the first - the use of two kinects is possible with IPIsoft but not NITE . To synchronise capture between the Kinect and MAC systems a light box was placed within the field of view. A trigger illuminated LEDs within the box and simultaneously sent a +5 V signal to the MAC system. These events were used to align both recording systems.

Procedures Several movements were performed by each participant, included to cover a variety of segment orientations and velocities - see Table 1. Three repeats of each movement were performed and participants were able to rest in between movements when this was required. The protocol was repeated twice; once with a Kinect positioned optimally for NITE and once with two Kinects positioned optimally for IPIsoft. Data from the MAC system were recorded concurrently during both repeats of the protocol and the order in which collection for NITE and IPIsoft was performed was randomised.

Data Analysis For brevity, data for only the reach and throw motions are presented in this report. These movements were chosen as the reach is a relatively slow activity of daily living and the throw is a relatively fast sporting motion; representing a broad spectrum of movement that the Kinect can be used to analyse. The start and end of both motions were defined using the kinematic data from the MAC system. The start of the motion for the throw was defined as the onset of movement and, for the reach, the onset of forward movement of the hand. The end of the motion for both movements was defined as the point of maximum elbow extension. Both shoulder and elbow flexion-extension angles were calculated. For the data derived from the MAC system as well as the NITE and IPIsoft Kinect data, local segment coordinate systems were defined in which the $x - axis$ pointed anteriorly, the $y - axis$ pointed superiorly and the $z - axis$ pointed laterally. In accordance with International Society of Biomechanics guidelines [15], elbow flexion-extension was defined as the first Euler angle in the ZXY sequence (Reference Segment: Upper arm, Target Segment: Lower Arm) and the shoulder angle was calculated as the second Euler angle in the YXY sequence (Reference Segment: Thorax, Target Segment: Upper arm). These angles were calculated for the NITE, IPIsoft and MAC data and all resulting time series were normalised to 100 data points.

The MAC data were used as ground truth and the accuracy of the NITE and Ipisoft algorithms was assessed by calculated the root mean square error (RMSE) and maximum error.

3.2.3 Results

Table 2 illustrates the RMSE and maximum error between the Kinect and MAC angles at the elbow and shoulder for the reach and throw action as calculated by the NITE and IPIsoft tracking algorithms (participant 7 has been omitted from the throw data due to collection problems). As example representative data, Figure 2 illustrates the mean motion of participant 3 in the reach and throw actions as calculated by NITE, Figure 3 illustrates the mean motion of participant 9 in the reach and throw actions as calculated by IPIsoft.

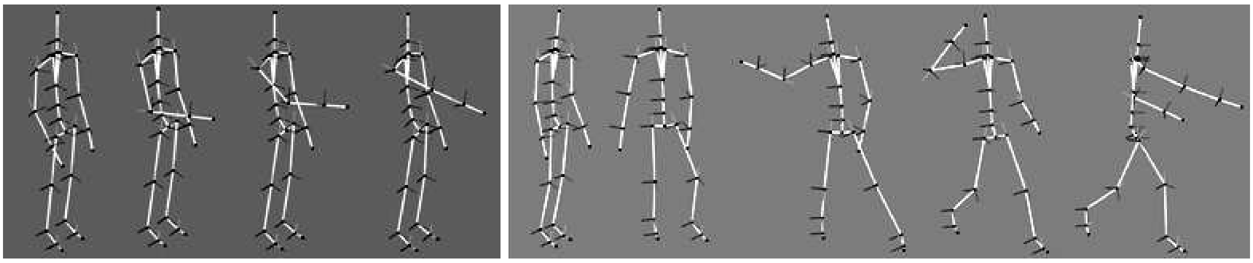

Figure 1: A representative participant performing the reach (left) and throw (right) movements.

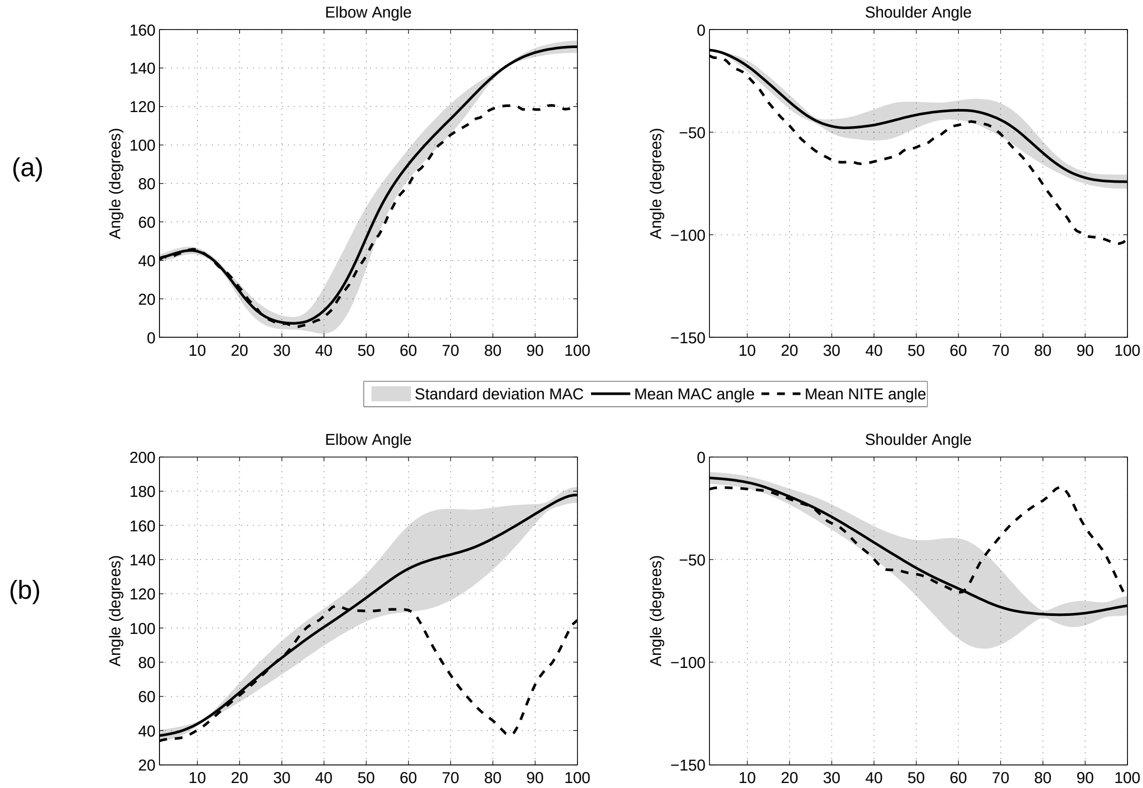

Figure 2: The mean motion of participant three as recorded by the MAC system and NITE tracking algorithm for the (a) reach and (b) throw action

3.2.4 Discussion

Of the two motions recorded, the reach is a slower action and was captured more accurately by both tracking algorithms. Generally, there was very little difference between NITE and IPIsoft in the estimation of elbow flexion-extension. The shoulder flexion-extension angle was estimated with greater accuracy using IPIsoft.

Table 1: The movements performed by each participant

Movement	**Description**
Elbow flexion-extension (1)	With the upper trunk vertical and the shoulder in a neutral position and the forearm in a supinated position, participants were asked to move their right arm from maximum elbow extension to approximately 120° of flexion before returning to full extension.
Elbow flexion-extension (2)	With the upper trunk vertical and the shoulder in a 90° abducted position and the palm of the hand facing upwards, participants were asked to move their right arm from maximum elbow extension to approximately 120° of flexion before returning to full extension.
Shoulder flexion-extension	With the upper trunk vertical, the shoulder in a neutral position and the elbow in full extension (participant in the 'military' position), participants were asked to flex their right shoulder to approximately 90°. Subsequently, they were asked to move their shoulder to a position of approximately 30° extension before returning the shoulder to a neutral position.
Shoulder abduction	With the upper trunk vertical, the shoulder in a neutral position and the elbow in full extension (participant in the 'military' position), participants were asked to abduct their right shoulder to approximately 60° before returning the shoulder to a neutral position.
Shoulder internal external rotation	With the upper trunk vertical, the shoulder in a neutral position, the elbow at approximately 90° of flexion and the forearm in a pronated position, participants were asked to internally rotate their right shoulder to approximately 30°, then externally rotate the shoulder to 30° before returning the shoulder to a neutral position.
Hip abduction	From standing in the 'military' position, participants were asked to abduct their right hip to approximately 30° before returning the hip to a neutral position.
Hip rotation	From standing in the 'military' position, participants were asked to external rotate their right hip to approximately 20° before returning the hip to a neutral position.
Star jump	A typical star jump motion was performed. This movement predominantly involves abduction-adduction motion at the shoulder and hip joints.
Throw	A maximal effort simulated overarm throw was performed.
Leg raise	Starting in the 'military' position, participants moved their right leg into a position of 90° hip and knee flexion before returning to a neutral position.
Walk on the spot	Participants were required to walk on the spot, primarily involving hip flexion-extension and knee flexion-extension.
Jump	A simple counter movement jump was performed. This movement predominantly involves flexion-extension motion at the ankle, knee and hip joints.
Cup to mouth	As an activity of daily living requiring the simultaneous execution of several anatomical movements to perform a 'cup to mouth' movement. Starting in the military position, participants reached to pickup a cup at approximately arm's length. Subsequently, they took a simulated sip from the cup before returning it to the stand.
Reach	The first phase of the cup to mouth movement - during which the participant moved from the anatomical position to pick up the cup at approximately arm's length - was analysed separately as a reaching movement.

This increase in accuracy might be due to the increased complexity of the IPIsoft skeleton, which includes a multi-joint shoulder complex - containing clavicle segments which are not included in the NITE skeleton. Furthermore, IPIsoft utilised two Kinects producing a more comprehensive point cloud, potentially improving

Table 2: Elbow and shoulder flexion-extension angle error. RMSE(Maximum error)

	Elbow Angles (°)				**Shoulder Angles (°)**			
	NITE		**IPIsoft**		**NITE**		**IPIsoft**	
ID	**Reach**	**Throw**	**Reach**	**Throw**	**Reach**	**Throw**	**Reach**	**Throw**
1	20.2(43.6)	53.8(155.8)	20.1(42.0)	31.5(86.5)	15.4(45.4)	9.7(22.4)	4.6(10.8)	11.0(20.1)
2	20.7(33.9)	84.5(169.8)	8.2(21.5)	23.7(74.6)	21.2(37.9)	7.0(11.9)	10.7(17.6)	14.1(18.8)
3	14.3(41.3)	39.4(101.4)	16.8(43.0)	22.7(60.4)	12.5(30.2)	32.8(115.0)	6.0(15.8)	8.6(29.1)
4	18.2(41.9)	65.7(189.6)	18.3(45.7)	33.2(71.5)	22.5(45.3)	24.2(71.3)	10.2(21.4)	5.8(13.5)
5	5.1(16.6)	19.7(74.9)	8.8(21.1)	12.6(42.6)	12.4(26.0)	14.5(37.0)	5.1(9.7)	8.9(22.9)
6	8.4(18.5)	35.8(95.7)	9.9(18.6)	13.1(27.4)	7.2(14.9)	15.7(47.5)	6.9(11.9)	5.2(12.7)
7	15.4(46.9)	——	16.8(34.0)	——	7.6(17.0)	——	4.9(10.5)	——
8	19.8(35.4)	22.6(75.5)	18.4(30.3)	21.2(39.5)	15.4(40.1)	17.4(74.2)	3.3(8.5)	9.4(25.3)
9	13.7(26.7)	39.3(117.1)	9.7(26.1)	17.6(40.0)	10.6(29.3)	34.4(117.2)	7.9(15.0)	21.4(35.9)
10	6.6(18.3)	98.2(167.0)	11.6(23.0)	59.1(108.3)	25.1(39.1)	18.2(71.5)	13.7(18.8)	7.4(28.2)
Mean	14.2(32.3)	51.0(127.4)	13.9(30.5)	26.1(61.2)	15.0(32.5)	19.3(163.1)	7.33(14.0)	10.2(22.9)

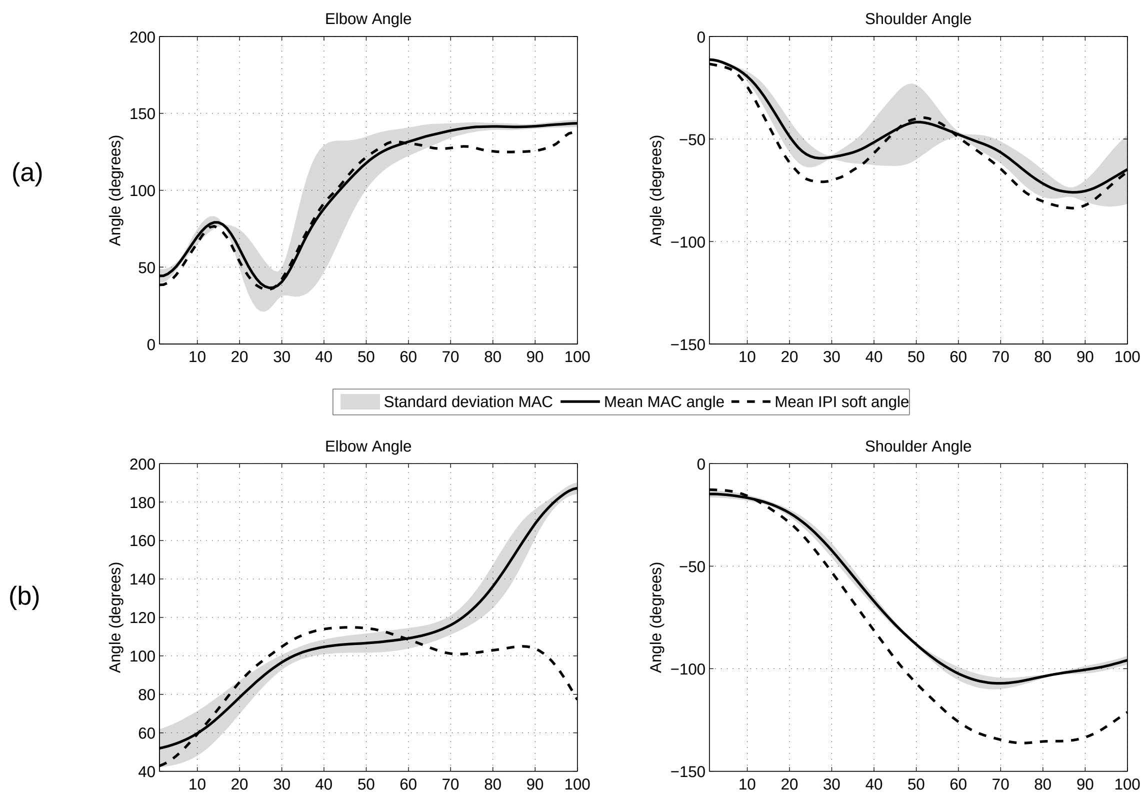

Figure 3: The mean motion of participant nine as recorded by the MAC system and IPI soft tracking algorithm for the (a) reach and (b) throw action

tracking.

During the throw, the accuracy of the elbow flexion-extension angle tracking was greatly reduced for both algorithms. It is apparent from Figures 2 and 3 that tracking accuracy reduces towards the end of the action during which period velocities are greatest, especially in the lower arm segment. Figure 2 illustrates that tracking fails completely towards the end of the action when using the NITE algorithm (mean maximum error of 127.4°). It is also clear from Table 2 that tracking failed in some IPIsoft cases, with maximum errors in excess of 100° for participants 3 and 9. The shoulder flexion-extension angle was tracked more accurately by IPIsoft, with evidence of tracking failing in the shoulder segment with the NITE algorithms (participant 10 has a maximum error of 108.3°). This breakdown in tracking is most likely a limitation of the Kinect hardware and the long exposure time of the infrared camera used in the calculation of depth. Blurring was evident when analysing recorded motion using IPIsoft and NITE, making it difficult to track a segment accurately.

The Kinect is a potentially valuable motion analysis tool. RMSE can be as low as 3° during a reach motion and errors are generally comparable to other marker-less tracking techniques [3]; although, given limitations related to the speed of movement and large maximum errors, it is unlikely the system is suitable for work where substantial accuracy is required. However, there are several significant advantages which mean this system has potential for many motion tracking applications. The Kinect is low cost (< £100), requires no calibration, no body markers and has freely available tracking functionality (NITE). However, to fully use the freely available tracking functionality, programming is required. IPIsoft offers the advantage of generally greater tracking accuracy, a more complex skeleton and a fully functioning software package, although real-time processing is currently not available. The Kinect has potential as a motion analysis tool for observing ranges of motion, or comparative studies during slower movements - this could be in sports coaching, clinical or education domains - or in other disciplines in which absolute accuracy is less of a concern - e.g. computer animation.

3.3 Investigation 2: Three-dimensional scanning

3.3.1 Introduction

Three-dimensional surface imaging techniques are important in several health and medical domains, being used for the diagnosis, monitoring and treatment of pathologies involving structural changes to bone, soft tissue and skin ([10, 2]). For example, surface scanning techniques have been used in adult facial assessment [7]. Furthermore, the use of three dimensional scanning techniques to assess breast morphology - in breast cancer and cosmetic surgery, for example - has been explored [10, 2]. Surface scanning is also important in many other disciplines. For example, Wicke and Dumas [14] recently highlighted the importance of obtaining accurate three-dimensional geometry information of the trunk segments when estimating body segment inertia parameters (BSIP) such as mass, centre of mass location and moments of inertia. Non-contact laser scanners have been used for this purpose [9, 12] and Wicke and Dumas [14] recently suggested that structured light scanning techniques would be appropriate for obtaining BSIPs.

Previously used imaging techniques - such as those based on non-contact laser scanners or stereophotogrammetry - are, generally, expensive and complex. Systems can cost in the order of £10,000, can be difficult to set-up (requiring complex calibration) and can require considerable technical expertise to operate effectively. The Kinect offers a potential, cost effective, alternative to previously used systems, offering the possibility of more widespread use of three-dimensional scanning methods across many different disciplines. The purpose of this initial, small-scale study was to investigate the use of the Kinect in three-dimensional scanning applications, focussing mainly on breast imaging. Although multiple Kinects could be registered to form more comprehensive scans, in this study we investigated the accuracy of the Kinect as a three-dimensional scanner in its most basic form, using only one Kinect. We did this to obtain 'baseline' errors and to help establish the feasibility of using the Kinect for three-dimensional scanning in its most basic form - offering the greatest potential for widespread use. We focussed on the context of breast imaging because of potential follow-up projects/grant applications and links with a surgeon and breast clinic. Furthermore, it is an application that has recently received attention in the literature [2, 10].

3.3.2 Methods

Two small scale experiments were performed to investigate the accuracy of the Kinect for three-dimensional surface imaging, in a breast scanning context. Plastic spheres of known diameter (146.5-148.0 mm) were scanned, in addition to a breast model. The same experimental setup was used for both experiments. The objects to be scanned were placed on a table, directly in front of the Kinect, approximately 900 mm away. Three-dimensional point clouds of the objects were obtained using custom written software. For the spheres testing, one scan of three spheres was collected. For the breast model, eight repeat scans were collected.

The sphere point cloud data were analysed by estimating the position of each point in the point cloud relative to the centre of the sphere - which we assumed should be constant. The centre of the sphere was found using the point cloud data by employing a gradient descent optimisation, minimising variability in radius estimates.

Analysis of the breast model was conducted to establish the accuracy with which simple, straight line three dimensional distances could be estimated using the Kinect. Markers (blue stickers, approximately 1 cm in diameter) were attached to a point approximating the suprasternal notch and an intermediate position on the dorsal aspect of the left breast. The position of these markers and the two nipples of the model were extracted from the point cloud scan of the breast model through manual digitising using Meshlab (3D-COFORM project - http://www.3d-coform.eu/). Three straight line distances were analysed (1. inter-nipple, 2. sternum-to-right nipple and 3. sternum-to-intermediate marker). Ground truth distances were measured using digital callipers - the mean of three measurements was used.

3.3.3 Results

A histogram of the estimates of the radius of the three spheres is provided in Figure 4. The standard deviation of estimates was 1.5 mm. Results of the breast model testing are provided in Table 3.

3.3.4 Discussion

The aim of this small-scale study was to investigate the accuracy with which the Kinect could be used to obtain three-dimensional scans. Because of links to potential follow-up projects and the popularity of

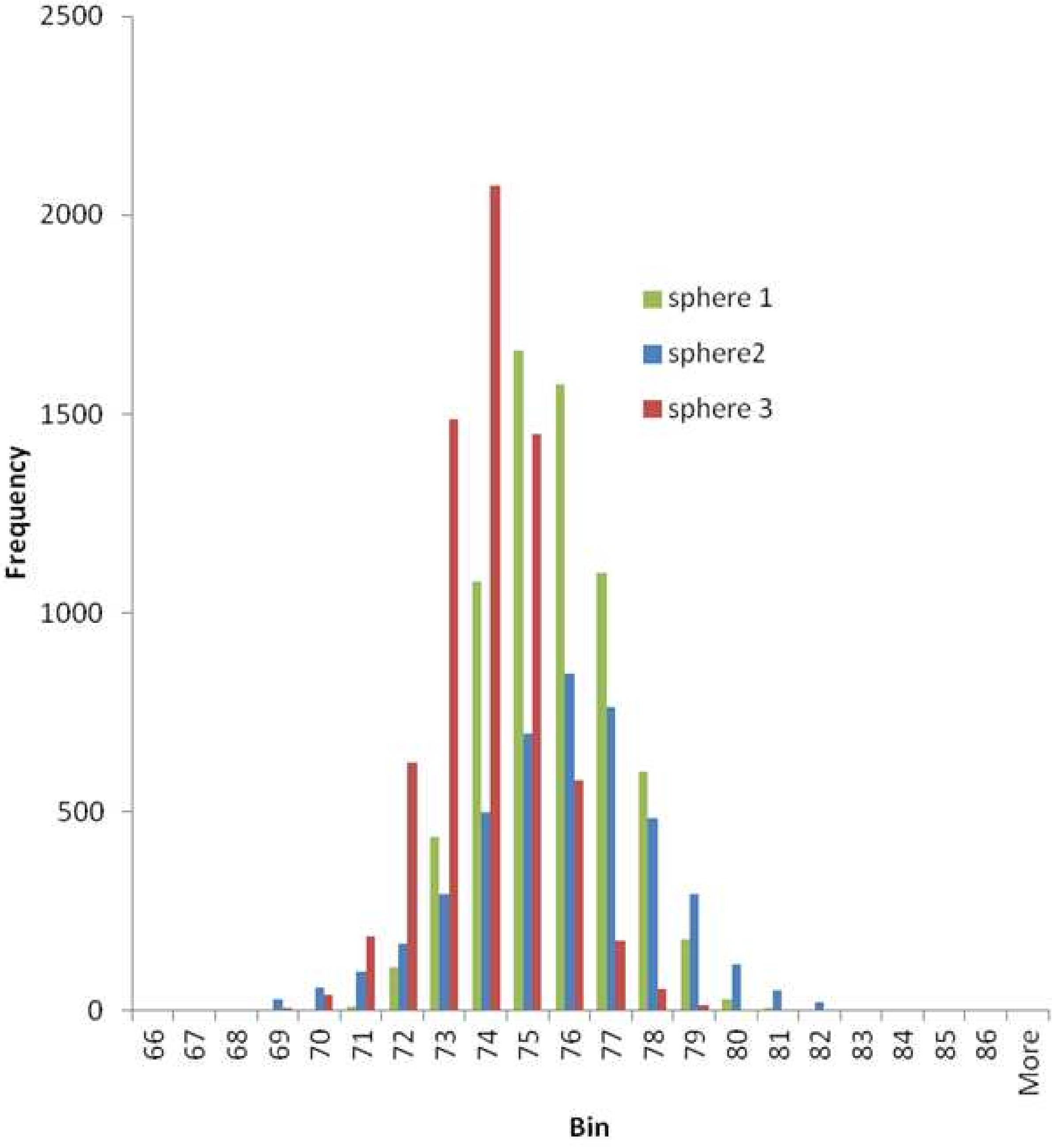

Figure 4: Estimates of the radius of the three spheres, scanned using the Kinect system (mm)

the area in the literature, the current study focussed on surface scanning of the breast. Two experiments were conducted. The first involved taking surface scans of spheres of known diameter and estimating the repeatability of estimates of the radius. Spheres were chosen as a very basic but acceptable approximation of the breast - many studies estimating breast volume use geometric equations that approximate the breast

Table 3: Estimates of the inter point, three-dimensional distances on the breast model (mm)

	Sternum-Intermediate	**Inter-nipple**	**Sternum-nipple**
Calipers	**123.3**	**172.0**	**159.0**
Trial 1	123.4	171.4	159.2
Trial 2	119.9	171.1	160.3
Trial 3	122.6	171.7	161.9
Trial 4	120.8	173.6	158.7
Trial 5	120.9	172.6	160.6
Trial 6	122.8	172.2	161.5
Trial 7	121.9	174.1	159.9
Trial 8	121.4	171.1	161.5
Mean	121.9	172.2	160.3
SD	1.2	1.1	1.2
Mean Error	-1.4	0.2	1.3
Max Error	-3.4	2.1	2.9

to be an ellipsoid. The standard deviation of the radius estimate was 1.5 mm. This is extremely promising given the cost of the Kinect and the simplicity of the scanning set-up. In the context of breast scanning and potentially informing clinical decisions, the breast surgeon with which we are collaborating indicated an acceptable error tolerance of ± 5 mm. Our data compare favourably with this analytical goal. When considering the accuracy of the estimates, the modal values for the radii were 74 mm, 76 mm and 75 mm for spheres 1, 2 and 3, respectively, indicating good accuracy for the breast scanning application and associated analytical goals.

In the second experiment, we investigated the accuracy with which the Kinect could estimate three-dimensional, straight line distances on the surface of a female mannequin. Again, in the context of the breast scanning application and the associated analytical goals, the errors were low (see Table 3). Furthermore, these errors compare favourably with similar studies in the literature, investigating the accuracy of commercially available scanning systems. For example, Catherwood et al. [2] reported mean errors in three-dimensional, straight line distances between 17 anatomical landmarks on a female mannequin of between 0.25 and 2.27 mm, with a mean error across all distances of 0.88 mm (the mean error across all distances in the present study was 0.97 mm). Importantly, the errors in the straight line distances between landmarks is considerably within the acceptable limits outlined by the collaborating breast surgeon (± 5 mm).

In these initial, small-scale experiments, we have demonstrated good accuracy and repeatability of a Kinect-based scanning system in a breast scanning context. Although, the results of the experiments have been interpreted in the context of breast scanning, the accuracy and repeatability indicate that the Kinect might be appropriate for many surface scanning applications. For example, the Kinect-based scanning system could be useful in obtaining scans of body segments to enable the individual-specific estimates of body segment inertia parameters (mass, centre-of-mass location and moments of inertia) to be made. Further, as another example, the accuracy and reliability data suggest that three-dimensional scans from the Kinect could be used to monitor changes in body morphology throughout a period of exercise training. In summary, the initial, small-scale accuracy experiments reported here suggest that three-dimensional surface scanning with the Kinect is possible and that the accuracy and repeatability might be appropriate for several potential applications.

3.4 Investigation 3: Whole body tracking

3.4.1 Introduction

Determining the location of people in a three-dimensional space is important in sport and in monitoring activities of daily living. For example, it might be required to track the location of an older person in a residential home, monitoring their movements - assessing the risk, and logging the occurrence, of falls in the home. Being able to quickly and easily obtain accurate information regarding a person's three-dimensional location is also important in many sports. Tracking the three dimensional location of players in racket sports such as squash, tennis and badminton has received much of attention. Information about three-dimensional location is used to monitor distances covered by players, their speed of movement and their position on the court at specific points in rallies - providing important tactical information. As such quantitative analysis of player activity is now considered an important aspect of the coaching process in sport [1].

In the analysis of racket sports, existing systems are primarily based on single or multiple video cameras, although GPS (or similar) based methods have also been used. An example of a video-based system is the SAGIT/squash computer tracking software. A single video camera, with a wide angle lens, is mounted on a gantry above the squash court, with the entire court in the field of view. After background subtraction, players are extracted from the image of the court and an approximation of their position in the floor plane is tracked in each video frame (see Figure 5).

It is rare for papers presenting details of person tracking techniques to report details of the accuracy of the tracking with respect to some ground truth measure. Often, a technique is deemed successful if the algorithms do not fail and are robust in the tracking of individuals. However, recently, Vuckovic et al [13] presented results of a study investigating the accuracy of the SAGIT/Squash system. By having participants stand at known locations on the court and move through predetermined movement paths, the authors demonstrated what they deemed to be acceptable accuracy. However, it would appear some results were questionable. For example, error in distance covered by the player in one minute ranged between 1.33 m and 21 m, depending on the position and nature of the movements being performed.

Many video-based methods have been presented for tracking people in sport and every day environments. Similar to the SAGIT/Squash system [13], for example, these approaches generally employ computer vision algorithms. Although many examples of successful systems have been presented, the analysis of video images can be complex, dependent on ambient lighting and can require camera positions that can be, practically, difficult to achieve. Furthermore, estimates position are often limited to two dimensions, with points normally projected onto the floor plane. The depth data provided by the Kinect - in addition to the traditional video image - offers the possibility of the, relatively, simple capture of the **three-dimensional** location of a person in many different environments. Therefore, the aim of this study was to investigate the accuracy with which the Kinect could be used to track a player during simulated badminton singles play. Two experiments were conducted. The first examined the accuracy with which the Kinect could be used to obtain the three-dimensional location of the participant's centre of mass, using an accurate gold-standard measurement system (three-dimensional motion capture system), in a laboratory setting. The second explored the accuracy with which measurements could be taken in an ecologically valid, on-court, setting - replicating many of the procedures implemented by Vuckovic et al [13].

3.4.2 Experiment 1: Laboratory based testing

Methods

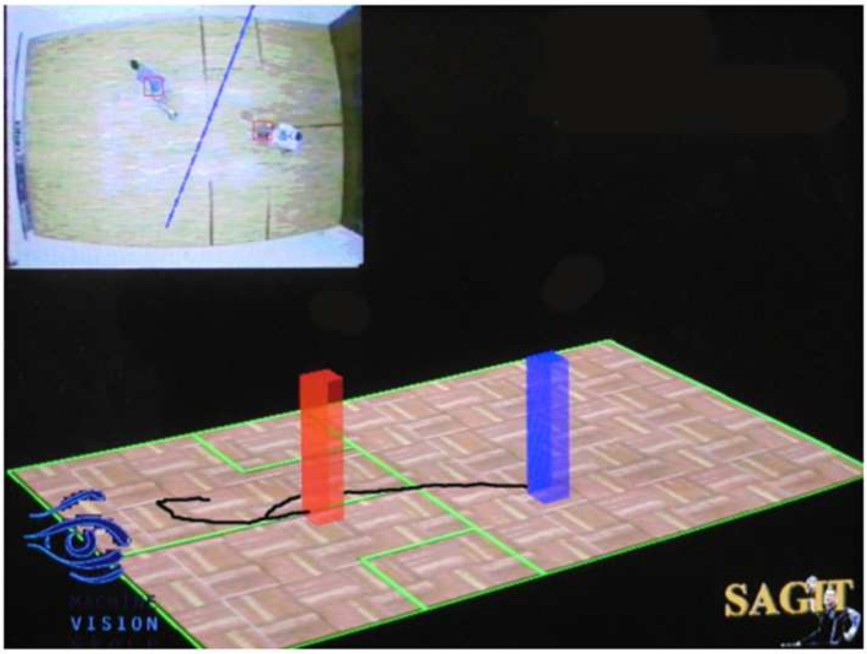

Figure 5: Example output from the SAGIT/Squash tracking system - taken from Pers et al. [11]

Participant One male, recreational badminton player volunteered to participate in the study and provided written informed consent before data collection began. The participant was free from any neuromuscular-skeletal injury that might affect completion of the protocol. The Faculty of Heath and Wellbeing Research Ethics committee approved all procedures.

Experimental Setup During all trials, data were collected concurrently from a Kinect and a 12 digital-camera motion capture system (Motion Analysis Corporation, Santa Rosa, CA, USA). The cameras of the motion capture system were set up in optimal positions around a 4m x 3m x 2.5m measurement volume and were calibrated in accordance with the manufacturer's specifications. During the anatomical calibration and movement trials, the sampling frequency of the motion capture system was 300 Hz. The motion capture system required 43 retro-reflective markers to be attached to the body at relevant anatomical landmarks before a static calibration trial. After the static trial, 10 markers were removed leaving 33 markers for the movement trials.

One eighth of a badminton court was marked out on the laboratory floor within the measurement volume of the motion capture system. This area served as one of four quarters of one half of the badminton court:

front forehand, back forehand, front backhand and back backhand (hereinafter referred to as one quarter of the court). This was deemed to be the largest area within which to obtain accurate data from the motion capture system. A Kinect was placed in one of six positions relative to the quarter of the badminton court, dependent on in which part of the court was being analysed (see Figure 6). When the calibrated measurement volume served as the back of the court, the Kinect was positioned 2.3 m away and when the measurement volume served as the front of the court, the Kinect was positioned 5.65 m away (the equivalent of 2.3 m behind the back line of the badminton court). A distance of 2.3 m behind the back line of the court was chosen through pilot testing as an appropriate compromise that maximised the field of view whilst maintaining usable depth resolution of the Kinect. Furthermore, the position complies with the Badminton World Federation's recommendation that there is at least 2 m clear space surrounding the outer lines of the court. Two positions of the Kinect were used at each distance from the measurement volume (see Figure 6). The Kinect was positioned on what would be the centre line of the court when the measurement volume acted as the forehand and backhand side.

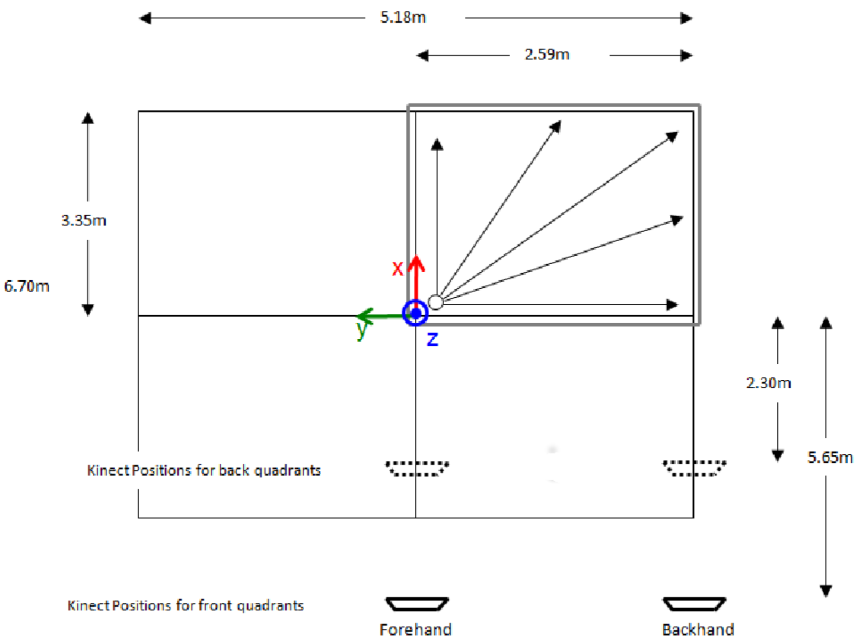

Figure 6: The positions of the Kinect relative to the MAC calibrated volume. The court-fixed coordinate system of the motion capture system is also shown.

During all trials, the sampling frequency of the Kinect was approximately 30 Hz. The Kinect and motion capture system were event synchronised using a custom-made device. A box incorporating an LED was position in the field of view of the Kinect's RGB camera. On a button press, the LED was illuminated whilst, simultaneously, a +5V signal was sent to an analogue input of the motion capture system. The first kinect video frame in which the LED was illuminated and the rising edge of the +5V signal defined an event with which to time-align the data from the motion capture system and the Kinect.

Procedure On entering the lab, time was provided for the participant to read the participant information sheet, ask any questions, and sign the informed consent form. Subsequently participants were asked to change into tight fitting lycra shorts and a t-shirt and the retro-reflective markers were attached to the body using double sided tape. Subsequently, a static calibration trial was carried out during which approximately two seconds of motion capture system data were collected, with the participant standing in the anatomical position. Subsequently, 12 movement trials were collected.

Movement trials required that the participant completed a badminton specific movement drill in each of the four quarters of the half of the court. The drill involved the participant starting in the centre of the court, moving to each of the positions represented in Figure 6 (or equivalent for the other quarters of the court). The participant was required to return to the centre of the court in between movements to each position. Movement drills for each quarter were repeated with the Kinect positioned on the centre line of the court and the centre line of measurement volume and the order in which the kinect positions were presented was:

- Front court / Forehand
- Front court / Backhand
- Back court / Backhand
- Back court / Forehand

Three repeats of the movement drill were collected for each kinect position. A rest period of approximately one minute was provided between each repeat.

Data analysis A court-fixed coordinate system, common to both the motion capture system and the Kinect was defined (see Figure 7). For the motion capture system, three retro-reflective markers were placed on the floor in an arrangement that defined a coordinate as shown in Figure 6. As the motion capture system measurement volume represented different regions of the court (e.g. front forehand), the dimensions of the badminton court were used to translate the origin to the position shown in Figure 7.

For the Kinect a coordinate system fixed to the floor was defined first by performing principal components analysis (PCA, see [4] for a description) on the points from a manually selected region of the floor. As the floor points were three-dimensional, the PCA returned three orthogonal principle components - the axes of which defined the floor-fixed coordinate system. Principal components one and two lay within the floor plane and defined the $x-axis$ and $y-axis$, respectively. Principal component three was perpendicular to the floor plane and defined the $z-axis$. The floor-fixed coordinate system was then aligned with the court markings by translation in the established floor plane and rotation around the $z-axis$. This was achieved by projecting representations of the unit vectors of the floor-fixed coordinate system onto the video image of the Kinect, facilitating alignment through visual inspection.

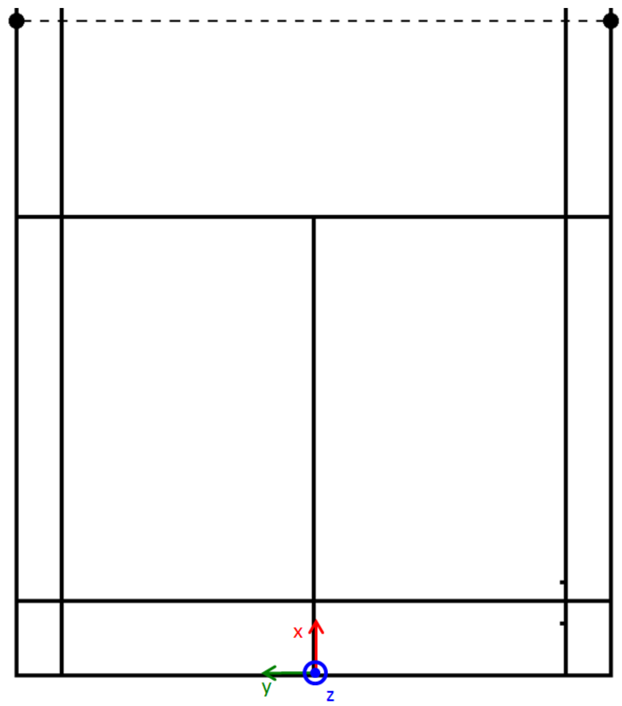

Figure 7: The court-fixed coordinate system

Raw marker coordinate data from the motion capture system were filtered using a fourth order, zero-lag, low-pass Butterworth filter. Subsequently, whole-body centre of mass location was estimated using the default geometric model in Visual3D (C-Motion, Germantown,MD, USA) - similar to that presented by Hanavan [6]. After assuming uniform density within each body segment (Dempster [5]), geometric solids were used to model the segments - truncated cones, with the exception of the head which was modelled as an ellipsoid - from which estimates of their centre of mass location were made. Using this information and an estimate of each segment's mass, the location of the whole-body centre of mass in the court-fixed coordinate system was calculated. This was assumed to be the ground truth location of the participant's centre of mass.

An estimate of the location of the participant's centre of mass in the court-fixed coordinate system, was also made using the kinect. Three-dimensional points returned by the kinect that lay within a bounding box coincident with the measurement volume of the motion capture system were assumed to represent the participant - the minimum height of the bounding box was set to 400 mm above the floor to avoid the influence of noisy points from the floor at large distances from the Kinect. Each three-dimensional point was assigned an equal, nominal, mass and the centre of mass of the points was assumed to represent the centre of mass of the participant.

The centre of mass (COM) time series calculated using the motion capture system were re-sampled to the same time points as the Kinect data using cubic spline interpolation (300 Hz to approximately 30 Hz). This was especially important as the sampling frequency of the Kinect was sometimes erratic. The root mean square error (RMSE) across all time points was calculated between the Kinect estimates of centre of mass and the ground truth estimate from the motion capture system at each court region and Kinect location. RMSE was calculated for x,y and z position of the centre of mass in the court coordinate system as well as the speed of the centre of mass. In addition a comparison was made of the estimates of the total distance moved by the participant during each movement drill - the position data from the Kinect were filtered using a simple moving average filter (with a 10 point radius) before estimates of total distance were made.

Results For brevity, results of one repeat from each of the four court positions are reported here. RMSEs for participant COM position and speed are outlined in Table 4. The total distance moved by the participant during the trial, estimated with both the MAC and Kinect data, is reported in Table 5. Time series for MAC and Kinect derived position and speed are presented in Figures 8, 9, 10 and 11. Finally, a comparison of participant COM location in the floor plane, derived from the Kinect and MAC systems is presented in Figure 12.

Table 4: RMSE for COM position and speed in the four court locations

Court Region	**Position RMSE (cm)**			**Speed RMSE (m/s)**
	X	**Y**	**Z**	
Front Forehand	25	9	6	0.21
Front Backhand	16	11	5	0.21
Back Forehand	13	9	8	0.18
Back Backhand	11	16	6	0.20

Table 5: Distance travelled by the participant, estimated using the Kinect-based system and the MAC motion capture system

Court Region	**Kinect (m)**	**MAC (m)**	**Difference (m)**
Front Forehand	21.4	21.2	0.2
Front Backhand	21.2	20.9	0.3
Back Forehand	17.6	19.0	-1.4
Back Backhand	18.5	19.5	-1.0

3.4.3 Experiment 2: On-court testing

Methods

Participant One male, recreational badminton player volunteered to participate in the study and provided written informed consent before data collection began. The participant was free from any neuromuscular injury that might affect completion of the protocol. The Faculty of Heath and Wellbeing Research

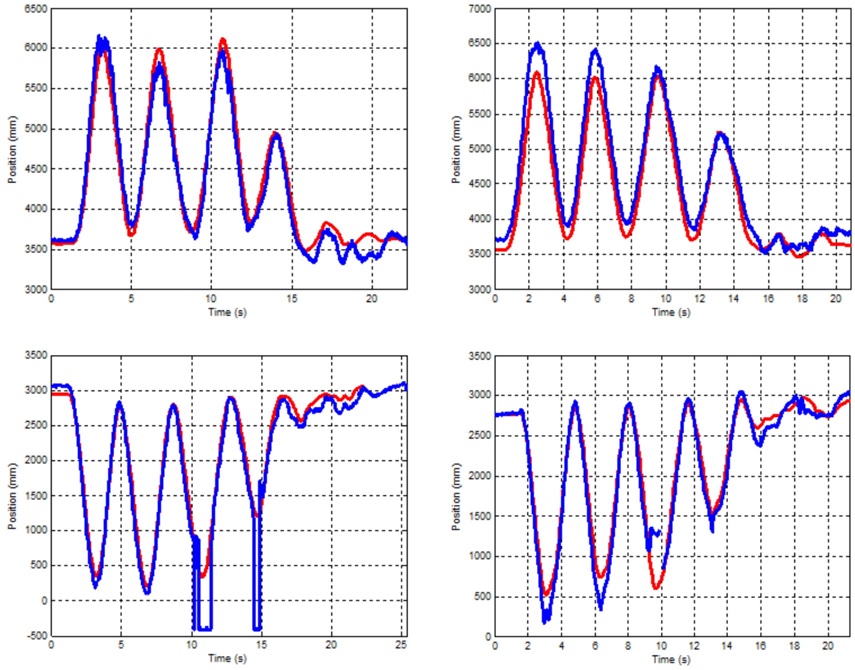

Figure 8: Centre of mass position along the $x - axis$ (fore-aft) of the court coordinate system. Top left: backhand front court, top right: forehand front court, bottom left: backhand back court, bottom right: forehand front court. Red line - MAC, Blue line - Kinect

Ethics committee approved all procedures.

Experimental Setup A standard badminton court in the University sports hall was used for data collection. A Kinect was placed on a tripod at a height of approximately 1.5m, at the back of the court, 2.3 m behind the back court line (see Figure 13). A distance of 2.3 m behind the back line of the court was chosen through pilot testing as an appropriate compromise that maximised the field of view whilst maintaining usable depth resolution of the Kinect. Furthermore, the position complies with the Badminton World Federation's recommendation that there is at least 2 m clear space surrounding the outer lines of the court. Additional markings were added to the court using tape. A grid of discrete points at known locations was added (Figure 13) and a 2 m diameter circle were marked on the court (Figure 13).

Procedure The participant completed two tests. During the first, they stood on each of the discrete points marked on the floor of the court for one minute (Figure 13). Participants were instructed to stand as

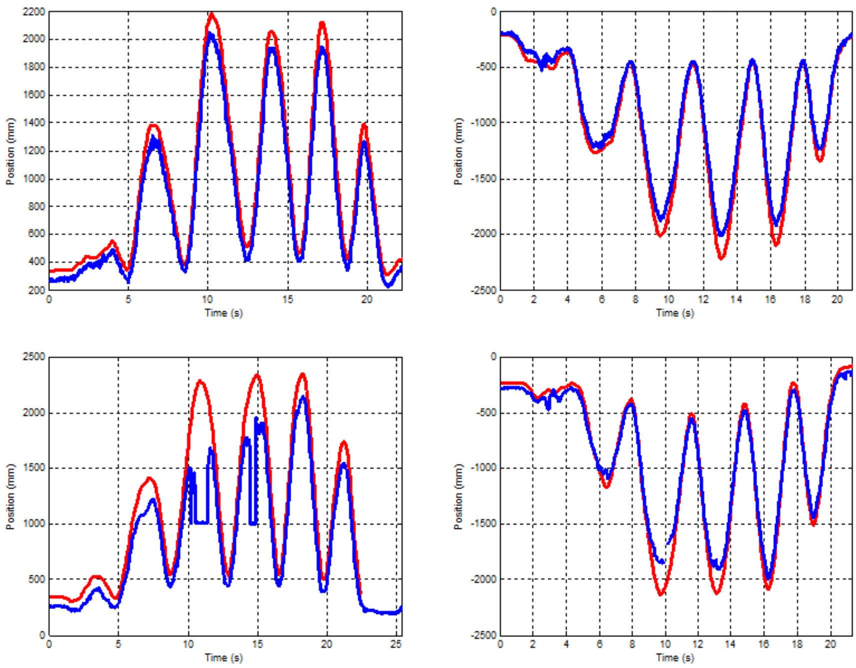

Figure 9: Centre of mass position along the $y - axis$ (side-to-side) of the court coordinate system. Top left: backhand front court, top right: forehand front court, bottom left: backhand back court, bottom right: forehand front court. Red line - MAC, Blue line - Kinect

still as possible, facing away from the Kinect, with feet approximately shoulder width apart and the marking on the floor at the centre of stance. The second test was a dynamic trial in which participants moved around a circular trajectory marked on the floor of the court. Starting from a position closest to the Kinect, participants completed laps of the circle at three self-selected speeds: walking, running and sprinting. Three laps of the circle were recorded at each speed.

Data Analysis A court-fixed coordinate system similar to that used in experiment 1 was defined (see Figure 7). For each frame of all trials, the Kinect returned a three-dimensional point cloud representing the scene in the field-of-view. Each point in the point cloud was transformed into the court-fixed coordinate system and the participant was segmented from the scene by assuming that any point with a z coordinate greater than 200 mm represented the participant - a vertical threshold of 200 mm was used to avoid the influence of noisy points from the floor surface sometimes present at large distances ($\sim$8 m) from the kinect.

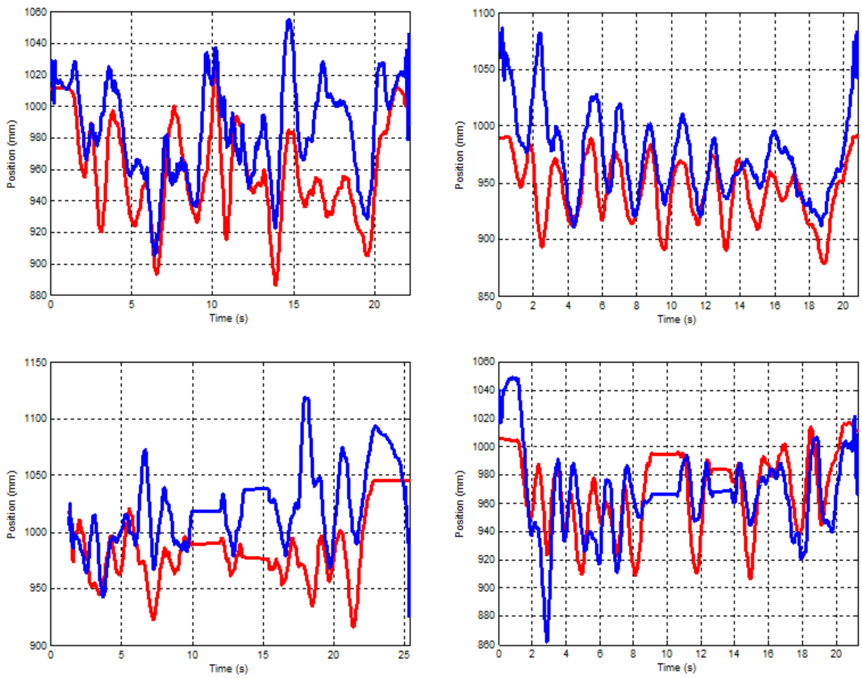

Figure 10: Centre of mass position along the $z-axis$ (vertical) of the court coordinate system. Top left: backhand front court, top right: forehand front court, bottom left: backhand back court, bottom right: forehand front court. Red line - MAC, Blue line - Kinect

The position of the participant on the court was determined by calculating the 'centre of mass' (COM) of the resulting point cloud. Each point representing the player was assigned an equal, nominal mass and the participant's COM was calculated by determining the location about which the mass of the points was evenly distributed.

Despite the fact that the three-dimensional position of the participant was determined using the Kinect, in the on-court experiments, the true position of the participant could only be constrained in the plane of the floor. As such, the position of the participant's COM was projected onto the floor plane (x-y plane of the court coordinate system). When analysing the position of the participant at each grid location (see Figure 15), the root mean square error (RMSE) between the participant's location and the reference location was calculated for each trial using equation 1:

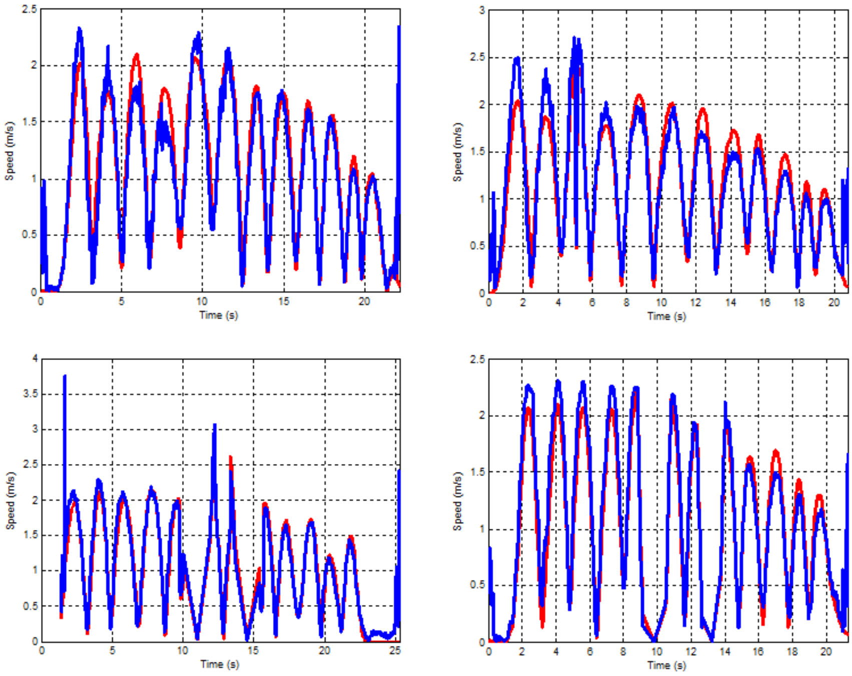

Figure 11: Centre of mass speed in the court floor plane. Top left: backhand front court, top right: forehand front court, bottom left: backhand back court, bottom right: forehand front court. Red line - MAC, Blue line - Kinect

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i - x_r)^2 + (y_i - y_r)^2} \tag{1}$$

where N is the number of data points, x_i and y_i are the estimates of the x and y positions, respectively, at the ith data point and x_r and y_r are the x and y reference positions, respectively.

For the trials in which the participant moved along the circular trajectory, agreement between the participant's movement and the reference trajectory was assessed using equation 2.

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(r_i - r_r)^2} \tag{2}$$

where N is the number of data points, r_i is the radial distance from the centre the circle to the ith data point and r_r is the radius of the circle marked on the floor of the court.

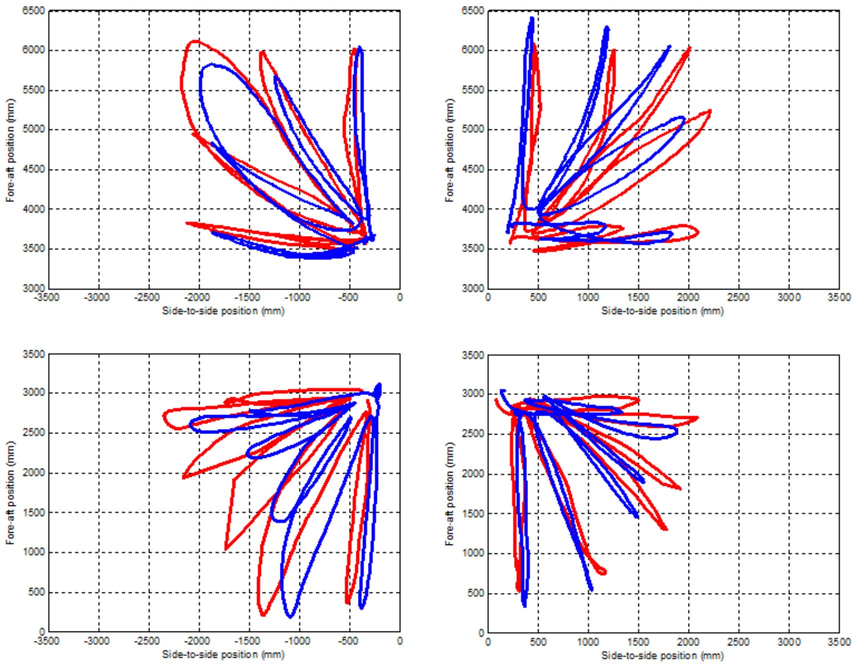

Figure 12: Centre of mass position the court floor plane. Top left: backhand front court, top right: forehand front court, bottom left: backhand back court, bottom right: forehand front court. Red line - MAC, Blue line - Kinect

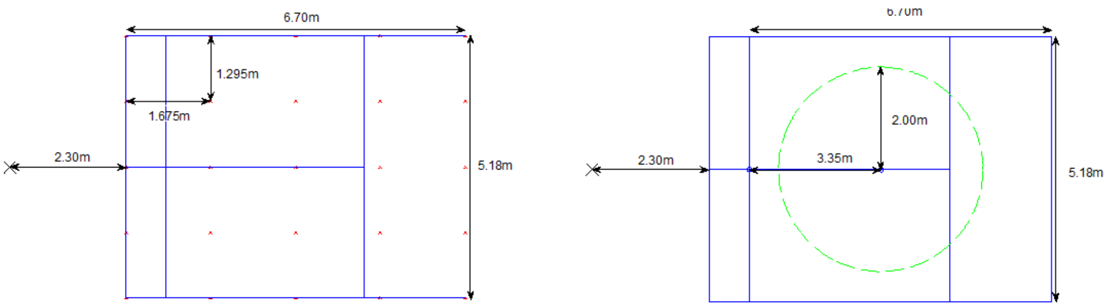

Figure 13: Additional markings on the badminton court. Left, a grid of positions over which the participant stood, stationary. Right, the circular trajectory.

3.4.4 Overall Discussion

The accuracy with which the position of a participant in three dimensional space can be determined using Microsoft's Kinect was investigated. Tracking the on-court movements of badminton players was used as an example situation in which knowledge of the three-dimensional position of a person is important. However, it

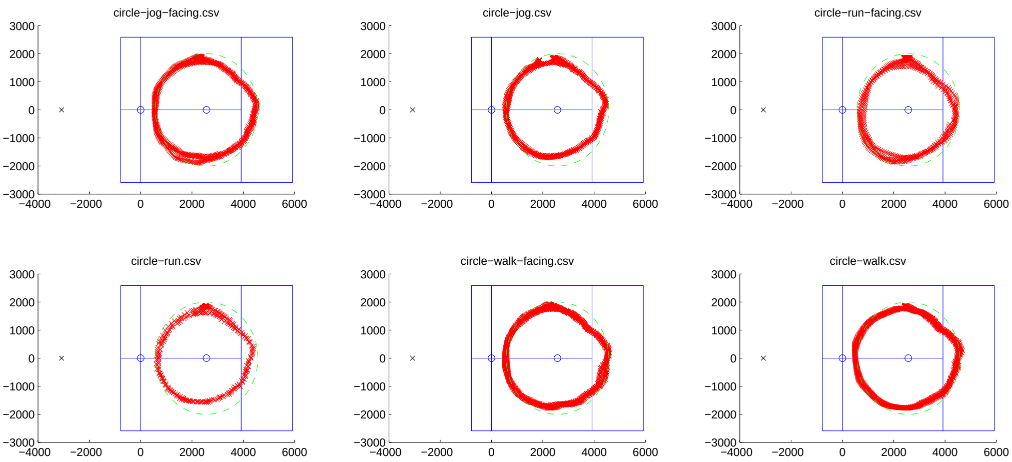

Figure 14: Circular trajectories measured using the Kinect for the participant moving along the reference trajectory at different speeds. The red cross hairs indicate the trajectory of the player completed three laps, measured using the Kinect based system. The dotted green line indicates the reference trajectory. The black cross indicates the location of the Kinect

is equally important in many, more general, contexts such as monitoring the location of an older person in a residential home. Importantly, badminton is a very dynamic activity and presents challenging conditions in which to track the location of a player. The experiments reported in this document investigated the accuracy of the Kinect-based system in two experimental settings: in the lab and in the 'field' (on a badminton court).

The more ecologically valid, badminton court setting, presented difficulties with defining the true position of player against which to compare the Kinect based estimates. However, an approach similar to that of Vuckovic et al [13] - used when investigating the accuracy of the SAGIT system in the analysis of squash play - was employed. This first involved the participant standing at known, discrete reference locations on the court (see Figure 13). In general, estimates of participant location using the Kinect based system were good, especially when the participant was relatively close to the Kinect (approximately <6m). The results at these distances compare favourably with the results presented by Vuckovic et al. [13] who presented a lower bound on error (taken from a position in the centre of the field of view) of 110 mm with an upper bound on error of 420 mm (taken from positions at the extremities of the field of view). In the present study,

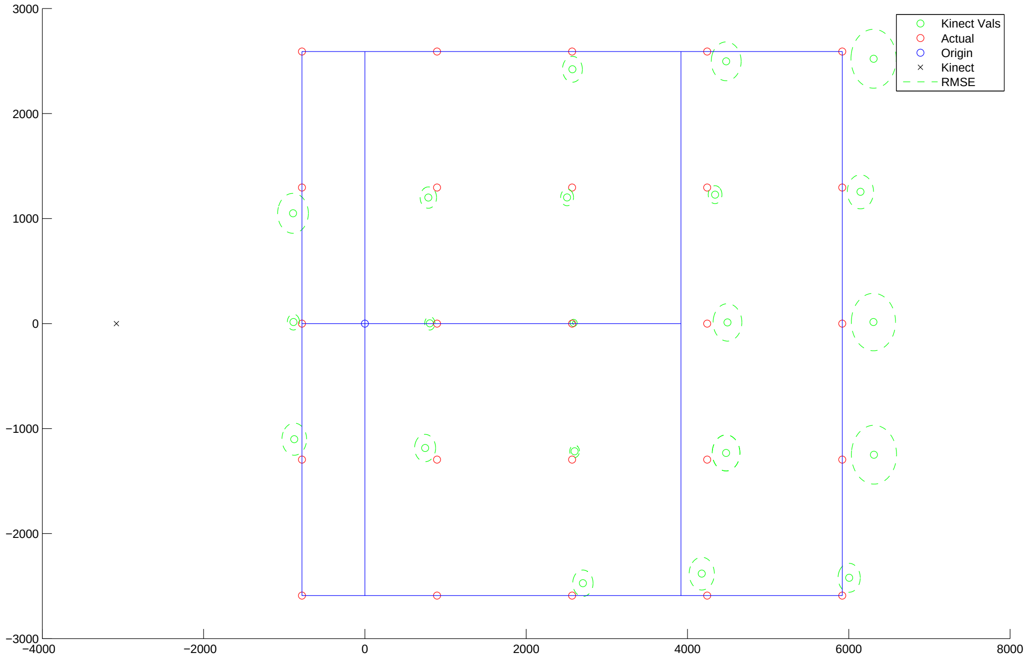

Figure 15: Positions of the participant estimated by the Kinect-based system at each of the predetermined reference grid positions on the badminton court.

at distances below 6m from the Kinect, RMSE errors ranged between 64 mm and 190 mm.

When further away from the Kinect - at distances greater than 6 m - errors in participant location determined using the Kinect-based system were larger (RMSE 85-280 mm). These increased errors are likely to have been caused by the inaccuracies and limits of the Kinect at these relatively large depths. Indeed, although the OpenNI driver (openni.org) for the Kinect (used in this study) returns depth values up to 10 m, it is interesting to note that Microsoft's Kinect for Windows SDK (Microsoft, Redmond, WA,USA) only returns depth up to a maximum distance of 4 m. Depth values for the Kinect are returned in 'bins' on an 11-bit scale - hence there are 2048 possible depth values. To improve resolution closer to the Kinect the real world distance between bins is smaller than at larger depths. Close to the Kinect ($\sim$1 m) the distance between bins is approximately 3 mm whereas towards the extremities of the depth range ($>$ 9 m) the distance between bins is $>$ 0.1 m. It is also conceivable that the accuracy with which the depth is calculated would decrease with increasing depth. However, even at the greater distances from the Kinect, the errors still compare favourably with the data presented by Vuckovic et al. [13].

During the on court testing, participants also completed trials during which they moved along a predefined movement path at three different speeds: walking, jogging and running. A 2m radius circle was marked on the floor of the badminton court for the participant to follow during the movement trials. Again, the resulting RMSE values compared favourably with those presented by Vuckovic et al. [13]. These authors reported errors of 170 mm, 320 mm and 500 mm for walking, running and sprinting, respectively, when a squash player moved around a circular reference trajectory with a radius of 2.7 m. In the current study, RMSE errors were slightly greater for walking (196 mm) but lower for the two running speeds (252 mm for running and 276 mm for sprinting). Other studies reporting the magnitude of tracking errors include an investigation of the tracking of beach vollyball players [8]. Marginally greater errors were reported by Mauthner et al [8] - approximately 225-350 mm - than those presented in the present study, again indicate the QUALITY of the Kinect-based tracking system.

Similar to the results presented by Vuckovic et al. [13], it can be seen in Figure 14 that the radius of the participant's trajectory is systematically smaller than the reference trajectory. This could have been caused the participant leaning their upper body towards the centre of the circular path [13]. Figure 14 illustrates that this issue becomes more pronounced at faster speeds, as would be expected.

The lab-based testing enabled the use of an accurate 'gold standard' position of the player to be determined (using a complex, expensive three-dimensional motion capture system), against which the Kinect-based estimates could be compared. We could find no studies in the literature which have used this approach to investigating the accuracy of person tracking algorithms. Agreement between the measurement systems in estimating participant position in the x and y directions (in the floor plane) was generally good (Table 4 and Figures 8 to 10 and 12). Similar to the on-court testing, the Kinect-based tracking system generally produced more accurate estimates of participant location (centre-of-mass in the floor plane during lab-based testing) when the participant was closer to the Kinect, at the back of the court. The results of the lab-based testing again compare favourably with the accuracy of other tracking systems presented in the literature [13, 8]. Results of the lab based testing also indicate good agreement between the motion capture and Kinect-based systems in estimating the speed of movement of, and total distance travelled by, the participant Tables 4 and 5).

Unlike previous tracking systems reported in the literature (e.g. [13, 8] which report only the two-dimensional position of a participant, the Kinect-based tracking system estimates three-dimensional position. Obviously, this is important in analysing activities such as badminton in which a great deal of movement takes place in the vertical direction, with players jumping to hit to hit high, and lunging to hit low, shuttlecocks. Furthermore, movements in the vertical direction can affect the accuracy with which estimates of the movement of a participant in the floor plane can be made with some two-dimensional tracking systems. In situations where the camera used for tracking is not perpendicular to the floor/ground plane, vertical movements of the participant can introduce error due to the projection of the participant's position onto the floor/ground plane [8]. The RMSEs presented in Table 4 suggest good agreement between the Kinect-based tracking system and the motion capture system in the vertical direction (z). However, in the context of the range of movement of the participant in the vertical direction, tracking was quite poor. Indeed, although inspection of the example time series data in Figure 10 suggests that the overall trends of vertical motion of the participant are present (certainly in the forehand front court), agreement between the Kinect and motion capture systems is considerably worse than tracking in the floor plane (x and y directions - Figures 8 and 9). Strangely, tracking in the vertical direction appeared to be more accurate in the front court quadrants - when the participant was a relatively large distance away from the Kinect. A simple tracking algorithm was used in this study that did not differentiate between the player and the racket. At times, when the participant was relatively close, the Kinect was able to track the racket. This meant that, occasionally, the racket was included in the point cloud from which an estimate of the participant's centre of mass was made. The 'gold standard' model of the centre of mass location did not include the racket. The presence of the racket in the point cloud generated by the Kinect would lead to poorer agreement between the Kinect and motion capture system based estimates of participant location. This might help explain the better agreement in vertical position estimates in the front quadrants as - given the greater distance to the Kinect - the racket was less often tracked than in the back quadrants.

Finally, it is important to note that, in this study in which a badminton player was tracked, the environment was free from objects that might obscure the filed-of-view of the Kinect. For other person tracking applications - e.g. tracking people in the home such objects might be present. More work is required to explore the difficulties associated with tracking people in more cluttered environments.

Conclusions The aim of this study was to investigate the accuracy with which the three-dimensional location of a participant could be determined using a Kinect-based tracking system. The results of both the on-court and lab testing suggest that the Kinect-based system exhibits accuracy that compares favourably with previously reported tracking systems that estimate participant location based on data from traditional video cameras. The added depth information available from depth cameras such as Microsoft's Kinect offers the possibility of more accurate tracking of a person's location in several contexts. In this study, the accuracy of the Kinect-based system was explored in the context of tracking a badminton player. The algorithms used for estimating the location of the badminton player were deliberately simple. However, even with the simple tracking algorithm, during the time when the participant was in the field-of-view of the Kinect, the participant was *never* misidentified (all tracking was automatic without the requirement for any manual intervention), emphasising the robustness that depth information adds to tracking algorithms. In the badminton context, improvements could be made by, for example, developing algorithms to ignore the racket. Obviously, more complex algorithms would also be required to track two players in badminton doubles play (importantly, with the added depth information provided by the Kinect, the problem of tracking multiple participants would be much more simple than with traditional two dimensional image based analysis). However, this complexity was avoided in the present study because the simple algorithm increases the degree to which the accuracy results presented in this study can be generalised to contexts other than badminton. It is likely that the accuracy of tracking presented in this study could be improved by tailoring bespoke algorithms to the constraints of a particular measurement context.

4 Stage 3

A further project steering group meeting was held on Monday 19th March 2012, during which the results of the work carried out in Stage 2 (see section 3) were discussed. The team were appraised of the findings of the accuracy/feasibility experiments and the results were interpreted in the context of potential follow-up interdisciplinary research projects. Outlined below is a brief overview of the discussions in interpreting the findings of the experiments during stage 2.

4.1 Investigation 1: Segment tracking

The accuracy of the segment tracking algorithms is quite poor when compared with more traditional lab-based equipment. Errors were generally an order of magnitude greater than would be expected by in a lab-based research context. However, there are many benefits associated with the Kinect including its cost and accessibility. This offers the possibility of taking biomechanical measurements in the field. Where the need for ecological validity or convenience outweighs the need for internal validity, the Steering Group agreed that the tracking algorithms associated with the Kinect may still be appropriate.

4.2 Investigation 2: Three-dimensional scanning

The Steering Group agreed that the results of the initial three-dimensional scanning accuracy experiments are promising. In general three-dimensional vector distances can be measured to a RMSE of approximately 1.5 mm with maximum error of approximately 3 mm. These errors are considerably within the analytical goals of many applications. For example, a potential application of the Kinect as an affordable three-dimensional scanner is taking measurements to monitor outcomes in breast reconstruction surgery. A breast surgeon at Royal Derby Hospital - with which we are developing a collaboration - has indicated that an error tolerance of approximately 5 mm would be appropriate for this context. This degree of accuracy would be appropriate across many other applications - e.g. calculation of body segment inertia parameters in biomechanical analyses.

4.3 Investigation 3: Whole body tracking

Similarly to Investigation 2, the results of the whole body tracking studies were promising. RMS errors were between approximately 5-30 cm. In the context of the size of a badminton court - or, indeed a living room, for example, with in which we might be interested in tracking the location of an older person - we deemed these errors to be quite small. Indeed, some of the errors could be explained by the nature of what is tracked by the Kinect. Our gold standard motion capture system provided a measure of the location of the participant's COM. The COM is normally located inside the body - often, approximately in the centre. However, the Kinect approximates this position of the COM by calculating the COM of a cloud of points on the surface of the skin - the aspect of the surface in the field of view of the Kinect. This offset could explain some of the error seen in Investigation 3. Notwithstanding these errors, the Kinect would appear useful for tracking the position of a person in a three-dimensional space.

5 Stage 4

The aim of the fourth stage of the project was to explore the possibilities related to improving the algorithms/techniques used in the analysis of the Kinect data, using the expertise of the multi-disciplinary team. Due to time constraints, work in the area was limited. However, initial work has explored the development of bespoke segment tracking algorithms to, potentially, improve on the accuracy presented in section 3. This has led to a multi-disciplinary collaboration related to an MSc Games Software Development dissertation project. Work is also in progress to improve the accuracy of the scanning and person tracking approaches.

6 Outcomes

The aim of this project was to establish the feasibility and accuracy of the Microsoft Kinect in various contexts related to potential funding applications - some already in development - across several research centres and departments across the university. A further aim was to explore potential ways in which we can improve the accuracy and applicability of the data produced by the Kinect. The degree to which the individual objectives of the project have been met is outlined in Table 6.

References

[1] Sian Barris and Chris Button. A review of vision-based motion analysis in sport. *Sports medicine (Auckland, N.Z.)*, 38(12):1025–43, January 2008.

[2] T Catherwood, E McCaughan, E Greer, R a J Spence, S a McIntosh, and R J Winder. Validation of a passive stereophotogrammetry system for imaging of the breast: A geometric analysis. *Medical engineering & physics*, March 2011.

[3] S Corazza, L Mündermann, a M Chaudhari, T Demattio, C Cobelli, and T P Andriacchi. A markerless motion capture system to study musculoskeletal biomechanics: visual hull and simulated annealing approach. *Annals of biomedical engineering*, 34(6):1019–29, June 2006.

[4] Andreas Daffertshofer, Claudine J C Lamoth, Onno G Meijer, and Peter J Beek. PCA in studying coordination and variability: a tutorial. *Clinical biomechanics (Bristol, Avon)*, 19(4):415–28, May 2004.

[5] W Dempster. Space requirements of the seated operator. In *WADC Tech. Report*, chapter Space requ, pages 55–159. Wright-Patterson Air Force Base, Dayton, OH., 1955.

[6] E.P. Hanavan. A mathematical model of the human body. chapter A mathemat, pages 64–102. Aerospace Medical Division, Ohio, 1964.

[7] Chung How Kau and Stephen Richmond. Three-dimensional analysis of facial morphology surface changes in untreated children from 12 to 14 years of age. *American journal of orthodontics and dentofacial orthopedics : official publication of the American Association of Orthodontists, its constituent societies, and the American Board of Orthodontics*, 134(6):751–60, December 2008.

[8] Thomas Mauthner, Christina Koch, Markus Tilp, and Horst Bischof. Visual Tracking of Athletes in Beach Volleyball Using a Single Camera. *Journal of Computer Science*, 6:21–34, 2004.

[9] Jonathan Norton, Nicholas Donaldson, and Laura Dekker. 3D whole body scanning to determine mass properties of legs. *Journal of biomechanics*, 35(1):81–6, January 2002.

[10] Scott M Paul, Andrew P Chamberlin, Charles Hatt, Amritha V Nayak, and Jerome V Danoff. Reliability, validity, and precision of an active stereophotogrammetry system for three-dimensional evaluation of the human torso. *Medical engineering & physics*, 31(10):1337–42, December 2009.

[11] Janez Pers, G. Vuckovic, and S. Kovacic. A low-cost real-time tracker of live sport events. *Image and Signal*, pages 362–365, 2001.

[12] Alison L Sheets, Stefano Corazza, and Thomas P Andriacchi. An automated image-based method of 3D subject-specific body segment parameter estimation for kinetic analyses of rapid movements. *Journal of biomechanical engineering*, 132(1):011004, January 2010.

[13] Goran Vučković, Janez Perš, Nic James, and Mike Hughes. Measurement error associated with the SAGIT/Squash computer tracking software. *European Journal of Sport Science*, 10(2):129–140, March 2010.

[14] Jason Wicke and Genevieve a Dumas. Influence of the volume and density functions within geometric models for estimating trunk inertial parameters. *Journal of applied biomechanics*, 26(1):26–31, February 2010.

[15] G Wu, F Vanderhelm, H Dirkjanveeger, M Makhsous, P Vanroy, C Anglin, J Nagels, a Karduna, K Mcquade, and X Wang. ISB recommendation on definitions of joint coordinate systems of various joints for the reporting of human joint motion?Part II: shoulder, elbow, wrist and hand. *Journal of Biomechanics*, 38(5):981–992, May 2005.

Table 6: Outcomes of the EFL Kinect project

Objective	**Achieved**	**Notes**
Confirm potential applications of the Kinect - aligned with multi-disciplinary/cross research centre funding applications - on which to focus during the project	YES	Potential application areas were identified in Stage 1 of the project.
Establish the accuracy of the data from the Kinect in the identified contexts	YES	Three separate studies were conducted highlighting the accuracy of the Kinect
If the accuracy in some contexts is not appropriate, we will explore ways in which further/post processing using discipline-specific analysis techniques can be used to improve the accuracy.	YES	Preliminary work has been conducted in this area starting to investigate the feasibility of bespoke segment tracking algorithms to meet the requirements in potential follow-up work
Explore the practicalities and feasibility of using the Kinect in the identified contexts.	YES	Feasibility and practicalities of using the Kinect in various contexts have been identified through the development of software tools for, and completion of, the experimental studies reported in section 3 of this report.
Publish the results of the accuracy analyses	IN PROGRESS	Aspects of all the studies reported in this document were presented in two invited seminar presentations at the Max Plank Institute for Biological Cybernetics in January 2012. Furthermore, a overview paper regarding the use of Kinect in Biomechanics - including data from this project - is in review with the Journal of Sports Science. A paper reporting the results of Investigation one is soon to be submitted with a manuscript for Investigation three in preparation. The results of of investigation one were be presented at the International Society for Biomechanics in Sports conference in Melbourne, Australia in July 2012. Furthermore, a website to disseminate the results of this programme of research as well as share resources and information related to the use of the Kinect and other depth cameras is currently in development (www.depthbiomechanics.co.uk).
Generate data with which to inform future multi-disciplinary, cross research centre funding applications.	YES	During the course of the project the overlaps between this and another EFL funded project, PhysioFu emerged. We have agreed to adopt a joined-up approach to the submission of funding application informed by the data from the Kinect accuracy/feasibility project. An expression of interest for the Healthcare Technology Challenges for Engineering EPSRC call was submitted on May 28th.