MARKER-LESS TRACKING OF HUMAN MOVEMENT USING MICROSOFT KINECT

Simon Choppin, Jonathan Wheat

Centre for Sports Engineering Research, Sheffield Hallam

This study quantifies the accuracy of the Microsoft Kinect in two motions. Ten participants were asked to perform reaching and throwing actions which were recorded simultaneously by a Kinect and a motion analysis corporation (MAC) capture system. Elbow and shoulder angles were calculated for both motions. NITE (PrimeSense, USA) and IPI soft tracking algorithms were used. NITE tracking had an average maximum error of 32.4° for the elbow and shoulder in the reach motion and 95.3° in the throwing motion. IPI soft had equivalent maximum error values of 22.3° and 43.0° respectively. While accuracy isn't high, and suffers in high speed motions, the advantages offered by markerless tracking, low cost and zero calibration make the Kinect potentially valuable for motion analysis in coaching, clinical and educational domains.

KEY WORDS: motion tracking, marker-less, accuracy.

INTRODUCTION: Human motion analysis is an important area of sports biomechanics. The measurement of joint positions and segment orientations is useful (for example) in determining injury risk (Krosshaug, Slauterbeck, Engebretsen, & Bahr, 2007) and analysing sports technique (Lees, 2002). Commonly, motion capture data are obtained through the use of multiple high speed cameras and a retro-reflective marker set fixed to body segments and joints (Baker, 2006), although alternative sensor systems (such as electromagnetic tracking systems e.g. Polhemus, USA) or marker-less video techniques (Corazza et al., 2006; Krosshaug et al., 2007) are available. While reported accuracy of marker based systems is high (Windolf, Götzen, & Morlock, 2008) (Liu, Holt, & Evans, 2007) set-up is time consuming and the systems can be invasive.
The Microsoft Kinect is a motion sensing device for use in home entertainment which captures movement as separate colour and depth data at 30 Hz and resolution of 640 × 480 pixels. A point cloud of depth data is obtained through the interpretation of a projected infra-red speckle pattern which is deformed by objects within the scene. The method of calculating depth information is thought to be related to structured light techniques (Scharstein & Szeliski, 2003) although specific details of the methods have not been published. When used with a PC, the device (through the use of proprietary algorithms) is able to pick out a user from within the scene and use point cloud data to calculate joint positions and orientations in real time (solving a similar problem to other marker-less tracking techniques (Corazza et al., 2006)). While the original intention of this functionality was the control of computer systems through gesture commands, the device is finding favour with computer animators as a low-cost and effective solution for motion capture. As a result, third party commercial software (IPI Soft, Moscow) is now available which claims to use algorithms which track more accurately than freely available software, although this is at the expense of increased processing time. The aim of this study was to explore the accuracy of proprietary and third party markerless tracking software in combination with the Microsoft Kinect system.

METHODS: Ten participants were used in the study, appropriate ethics approval was obtained from the faculty's ethics committee. Each participant was asked to perform a simple reaching motion and a throw three times with their right arm; these actions represent similar motions performed at low and high velocity. While the participants were allowed to move naturally, for analysis purposes both movements were defined to start from a neutral standing position (arms by side) and finish when the elbow is at maximum extension in front of the participant (illustrated in figure 1). A Microsoft Kinect system and a Motion Analysis Corporation (MAC) motion capture system captured each participant's movement simultaneously. The MAC system was used as ground truth. Two different tracking algorithms were used with the Kinect; NITE (Primesense, USA) which is freely available and works in real-time, and IPI Soft, a commercially available package which processes data post-capture. Both algorithms generate a full skeleton of the body, with the IPI Soft skeleton being more complex (19 segments as opposed to 15) the IPI Soft skeleton is shown in figure 1. The skeleton structure and its orientations were exported in Biovision format (bvh) for the IPI soft system (a feature of the software), and recorded as comma separated values (csv) with NITE using custom written software. The MAC marker set was limited to the torso, upper and lower arm. As two tracking algorithms were used the data capture process was performed twice, once using IPI Soft with two Kinects (functionality which allows a more complete capture) and once using NITE software with a single Kinect. A Kinect was positioned in front of the participant and a second Kinect (when used) was placed at approximately 70° to the first viewing the participant's right side. In order to align capture between the Kinect and MAC systems a light box was placed within the field of view, a trigger illuminated a set of LEDs and simultaneously sent a +5 V signal to the MAC system, these events were used to align both recording systems.

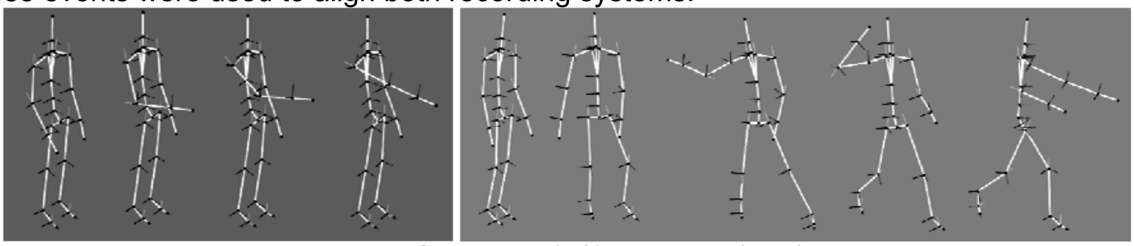

Figure 1: The start and end points of the reach (left) and throw (right) motion, showing appropriate intermediary stages.

Elbow flexion-extension and shoulder angles were calculated. In accordance with International Society of Biomechanics guidelines (Wu et al., 2005), elbow flexion-extension was defined as the first Euler angle (ZXY convention, reference segment upper arm, target segment lower arm), the shoulder angle was calculated as the second Euler angle (YXY convention, reference segment torso, target segment upper arm). All movements were normalised to 100 data points. The analysis was performed in custom matlab software written to analyse the bvh and csv data files.

RESULTS: Table 1 shows 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 IPI Soft tracking algorithms. Player 7 has been omitted from the throw data due to collection problems. For the purpose of illustration, figure 2 shows the mean motion of participant 3 in the reach and throw actions as calculated by NITE, figure 3 shows the mean motion of participant 9 in the reach and throw actions as calculated by IPI soft.

Table 1: Elbow and shoulder angle error. Maximum error values shown in bold, RMSE in plain text.

	Elbow Angles (°)				**Shoulder Angles(°)**			
	NITE		IPI Soft		NITE		IPI Soft	
Participant	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, **63.1**	7.33, **14**	10.2, **22.9**

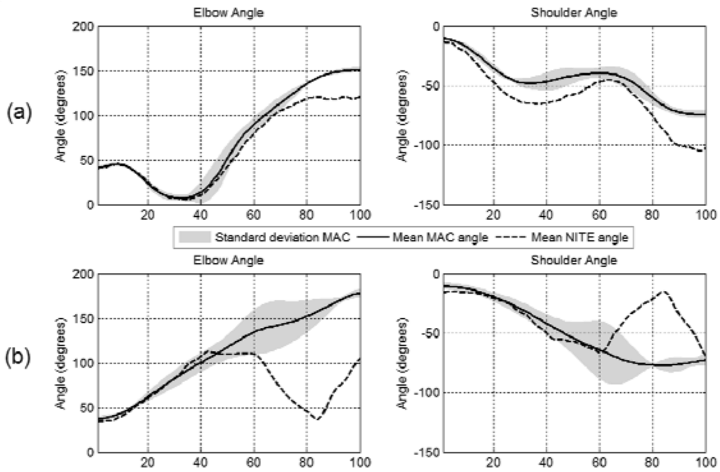

Figure 2: The mean motion of participant three as recorded by the MAC system and NITE tracking algorithm for the (a) reach and (b) throw action

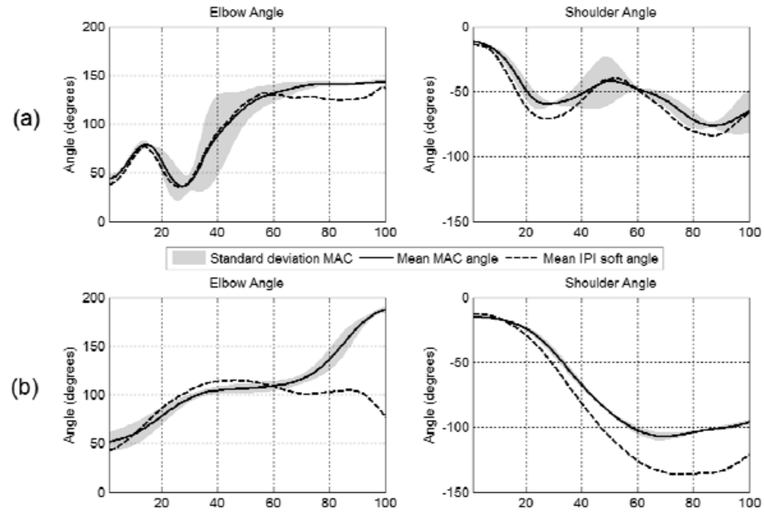

Figure 3: The mean motion of participant nine as recorded by the MAC system and IPI soft tracking algorithm for the (a) reach and (b) throw action

DISCUSSION: Of the two different motions recorded, the reach was a much slower action and was captured more accurately by both tracking algorithms. Generally, there was very little different between NITE and IPI soft in the estimation of elbow flexion-extension. The shoulder angle was estimated more accurately by IPI soft. This increase in accuracy may be due to the increased complexity of the IPI soft skeleton, which includes shoulder tracking.

During the throw, the accuracy of elbow angle tracking is greatly reduced for both tracking algorithms, NITE has the largest RMSE and maximum errors. It is apparent from figures 2 and 3 that tracking accuracy decreases towards the end of the action during which period velocities are greatest, especially in the lower arm segment. Figure 2 shows that, towards the end of the action, tracking breaks down completely in the NITE algorithm (mean maximum error of 127.4°). It is also clear from table 1 that tracking failed in some IPI soft cases, with maximum errors in excess of 100° for participants 3 and 9. The shoulder angle was tracked more accurately by IPI soft, 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 exposure time used to collect depth data. Blurring was evident when analysing recorded motion using IPI soft and Nite, making it difficult to track a segment accurately.

As a marker-less tracking system, the Kinect's RMSE can be as low as 3° during a reach motion and errors are comparable to other marker-less tracking techniques (Corazza et al., 2006) in the best cases; although given limitations related to the speed of movement and large maximum errors, the system is unsuitable for work where substantial precision is required. However, there are several significant advantages to the Kinect system; low cost (< £100), no calibration, no body markers and freely available tracking functionality (NITE). However, in order to fully use this, programming is required. IPI soft gives the advantage of generally higher tracking accuracy (as reflected by this study) a more complex skeleton and a fully functioning software package, although real-time processing is currently not available. Given the limitations of the Kinect, potential motion analysis applications should be considered carefully. Quantitative analysis where precision in segment angles is required is unsuitable, although figures 2 and 3 suggest that timing points for the start of motions could be obtained accurately during slower motions. The Kinect could be used as a qualitative analysis tool to (for example) demonstrate techniques/principles, assess relative ranges of motion or as a low cost educational tool. Future developments in hardware may yield higher resolutions, higher frame-rates and more accurate tracking algorithms.

CONCLUSION: The Kinect is a novel motion analysis tool which allows markerless tracking at very low cost. However, given the large maximum errors seen in this study, careful consideration should be given as to whether the Kinect is appropriate given the constraints in terms of accuracy and speed of movement. The Kinect could be used for qualitative analysis in situations where ecological validity is of greater importance than accuracy of measurement. This could include certain applications in coaching, teaching or clinical practice. IPI soft provides a more complex tracking algorithm for use with the Kinect with the disadvantages of increased cost and processing time. More work needs to be conducted to investigate more body segments and a wider range of movements.

ACKNOWLEDGEMENT AND ADDITIONAL INFORMATION

The authors would like to acknowledge Sheffield Hallam University's 'Engineering for life' project and the EPSRC who provided funding for this study. Ben Lane completed the data capture for this study and deserves our thanks. Copies of the software used in this study and further information on using the Kinect for motion analysis is available at www.depthbiomechanics.co.uk

REFERENCES:

Baker, R. (2006). Gait analysis methods in rehabilitation. *Journal of neuroengineering and rehabilitation*, *3*, 4.

Corazza, S., Mündermann, L., Chaudhari, a M., Demattio, T., Cobelli, C., & Andriacchi, T. P. (2006). A markerless motion capture system to study musculoskeletal biomechanics: visual hull and simulated annealing approach. *Annals of biomedical engineering*, *34*(6), 1019-1029.

Krosshaug, T., Slauterbeck, J. R., Engebretsen, L., & Bahr, R. (2007). Biomechanical analysis of anterior cruciate ligament injury mechanisms: three-dimensional motion reconstruction from video sequences. *Scandinavian journal of medicine & science in sports*, *17*(5), 508-519.

Lees, A. (2002). Technique analysis in sports: a critical review. *Journal of Sports Sciences*, *20*(10), 813–828. Taylor & Francis.

Liu, H., Holt, C., & Evans, S. (2007). Accuracy and repeatability of an optical motion analysis system for measuring small deformations of biological tissues. *Journal of biomechanics*, *40*(1), 210-214.

Scharstein, D., & Szeliski, R. (2003). High-accuracy stereo depth maps using structured light. *2003 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2003. Proceedings.*, 195-202. IEEE Comput. Soc.

Windolf, M., Götzen, N., & Morlock, M. (2008). Systematic accuracy and precision analysis of video motion capturing systems--exemplified on the Vicon-460 system. *Journal of biomechanics*, *41*(12), 2776-2780.

Wu, G., Vanderhelm, F., Dirkjanveeger, H., Makhsous, M., Vanroy, P., Anglin, C., Nagels, J., et al. (2005). 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.