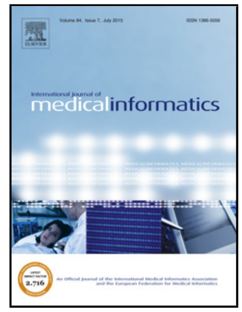
Accepted Manuscript



Title: Clinical Assessment of Depth Sensor Based Pose Estimation Algorithms for Technology Supervised Rehabilitation Applications

Authors: Joe Sarsfield, David Brown, Nasser Sherkat, Caroline Langensiepen, James Lewis, Mohammad Taheri, Christopher McCollin, Cleveland Barnett, Louise Selwood, Penny Standen, Pip Logan, Christopher Simcox, Catherine Killick, Emma Hughes

PII: S1386-5056(18)31275-9
DOI: https://doi.org/10.1016/j.ijmedinf.2018.11.001
Reference: IJB 3770

To appear in: *International Journal of Medical Informatics*

Received date: 13 November 2017
Accepted date: 5 November 2018

Please cite this article as: Sarsfield J, Brown D, Sherkat N, Langensiepen C, Lewis J, Taheri M, McCollin C, Barnett C, Selwood L, Standen P, Logan P, Simcox C, Killick C, Hughes E, Clinical Assessment of Depth Sensor Based Pose Estimation Algorithms for Technology Supervised Rehabilitation Applications, *International Journal of Medical Informatics* (2018), https://doi.org/10.1016/j.ijmedinf.2018.11.001

Clinical Assessment of Depth Sensor Based Pose Estimation Algorithms for Technology Supervised Rehabilitation Applications

Joe Sarsfield[1*], David Brown[1], Nasser Sherkat[2], Caroline Langensiepen[1], James Lewis[1], Mohammad Taheri[1], Christopher McCollin[1], Cleveland Barnett[3], Louise Selwood[4], Penny Standen[5], Pip Logan[5], Christopher Simcox[6], Catherine Killick[6], Emma Hughes[6]

[1]Interactive Systems Research Group, School of Science and Technology, Nottingham Trent University, Clifton Lane, Nottingham, UK, NG11 8NS.

[2]Sheffield Hallam University, Howard Street, Sheffield, S1 1WB

[3]SHAPE Research Centre, School of Science and Technology, Nottingham Trent University, Clifton Lane, Nottingham, UK, NG11 8NS.

[4]Nottingham CityCare, 1 Standard Court, Park Row Nottingham, UK, NG1 6GN.

[5]University of Nottingham, Queen's Medical Centre, Nottingham, UK, NG7 2UH.

[6]Nottinghamshire Healthcare NHS Foundation Trust, Stapleford Care Centre, Nottingham, UK, NG9 8DB.

**Corresponding author: joe.sarsfield@gmail.com*

Research Highlights

- Clinical analysis required to determine suitability of pose estimation algorithm.
- Joint rotation required for more in-depth and correct assessment.
- Exploit temporal information to ameliorate issues with occlusion.
- Constrain joint estimations to within the anatomical limits of the human body.
- Consult clinicians to select exercises that provide fewer challenges for algorithm.

ABSTRACT

Encouraging rehabilitation by the use of technology in the home can be a cost-effective strategy, particularly if consumer-level equipment can be used. We present a clinical qualitative and quantitative analysis of the pose estimation algorithms of a typical consumer unit (Xbox One Kinect), to assess its suitability for technology supervised rehabilitation and guide development of future pose estimation algorithms for rehabilitation applciations. We focused the analysis on upper-body stroke rehabilitation as a challenging use case. We found that the algorithms require improved joint tracking, especially for the shoulder, elbow and wrist joints, and exploiting temporal information for tracking when there is full or partial occlusion in the depth data.

Keywords: Stroke rehabilitation; Home rehabilitation; Clinical evaluation; Depth sensors; Pose estimation accuracy

1. INTRODUCTION

Stroke was the third most common cause of disability worldwide in 2010 (Murray et al., 2012), ranked by Disability-Adjusted Life Year (DALY). Even though motor control can be significantly improved by post-discharge exercises (Selzer, Michael,, unfortunately only 31% of patients adhere to their home exercise regime (Shaughnessy, Resnick, & Macko, 2006). Game-based rehabilitation systems can mitigate this (Burke et al., 2009), (Rego, Moreira, & Reis, 2010), (Flores et al., 2008), but they need to provide real-time feedback that reduces compensatory movements to enable true motor recovery.

In order to assess whether consumer-level equipment can provide this necessary quality of feedback, this paper presents a clinical analysis of a pose estimation algorithm running on a consumer depth sensor (Xbox One Kinect), in the context of stroke rehabilitation, using exercises taken from the Graded Repetitive Arm Supplementary Program (GRASP) manual (Harris J. E., 2009). A quantitative analysis of the joint position accuracy was also performed to support the clinical analysis. Section 2 indicates previous work in evaluating pose estimation algorithms in other contexts. Section 3 explains the methodology of both the clinical analysis and the quantitative analysis as applied to the stroke rehabilitation exercises. Section 4 examines the results and discusses the limitations associated with the pose estimations (both clinically and quantitatively). Section 5 draws conclusions and posits future work to be undertaken to improve the outcomes.

2. RELATED WORK

A number of studies have been undertaken on Kinect 1 and Kinect 2, though within different scenarios. Generally, these studies used marker based motion capture systems to establish ground truth. Fernández-Baena et al. (2012) examined Primesense's NITE pose estimation algorithm using depth data from Kinect 1. They claimed that joint accuracy could be improved by imposing a fixed length on the bones, and indicated that "Kinect can be a very useful technology in present rehabilitation treatments", though they performed no clinical analysis. Obdrzalek et al. (2012) examined Kinect 1 for elderly coaching exercises and concluded that measurements "could be used to assess general trends in the movement", though they made no clinical claims. Kurillo et al. (2013) found that Kinect 1's pose estimations were sufficiently accurate for reachable workspace analysis. Xu & McGorry (2015) compared the quality of Kinect 1 and 2 for poses within activities of daily living, and interestingly found that Kinect 1 produces lower errors. In contrast, Wang et al. (2015) considered that Kinect 2 was superior over a range of 12 exercises particularly when occlusion and body rotation occurred. In terms of clinical assessment, Yeung et al. (2014) found that Kinect 1 could achieve acceptable accuracy for total body centre of mass movements, but performed better for medial and lateral movements than anterior and posterior movements.

In terms of stroke rehabilitation, Webster et al. (2014) evaluated the joint accuracy of Kinect 1 on 13 gross movements. They found that Kinect 1 accuracy is sufficient for gross movement-based rehabilitation systems for clinical and in-home use. However, there was no assessment within standard rehabilitation exercises or clinical evaluation for the possibility of detecting compensatory movement. Mobini et al. (2013) evaluated the accuracy of the flexible action and articulated skeleton toolkit (Suma, Lange, Rizzo, Krum, & Bolas, n.d.) using Kinect 1 for upper body stroke rehabilitation applications. They found that lateral variations in position did not significantly impact joint accuracy, though horizontal distance had some effect. We note that these may be relevant issues when the stroke survivors are setting up their equipment without assistance.

These previous studies concentrated on absolute joint accuracy as compared with a ground truth provided by motion capture systems. The comparison with a clinical study, where expert clinicians provide analysis on significant aspects of the poses calculated by the equipment, was not performed. During rehabilitation, clinicians stress the importance of ensuring that the patients avoid compensatory movements, and so the evaluation and assessment of the pose algorithms needs to emphasise this aspect.

3. EXPERIMENTAL METHODOLOGY

This section will describe the methods employed for capturing, processing and analysing the data used in the clinical and quantitative studies.

3.1 Experimental Setup

Our evaluation is based upon version 2.0.1410.19000 of the pose estimation algorithm of the Kinect for Windows SDK for the Xbox One Kinect. This provides pose estimations for 25 joints at 30 Hz and allows a user's skeleton to be tracked on a subset of joints. Joint locations were recorded while seated, with default tracking mode in order to capture spine joints also.

Kinect produces a depth image with a resolution of 512*424 pixels (Microsoft, n.d.), as shown in Figure 1. This depth image is then used as input to the Kinect Software Development Kit's (SDK) pose estimation algorithm, which is based on the approach presented by Shotton et al. (2011) to infer the joint positions. This reduces the original feature space of the depth image (217088 values) to a much more tractable 75 values per frame. Given that Kinect's pose estimation algorithm (Shotton et al., 2011) runs in under 5ms on an Xbox 360 graphical processing unit (GPU), further in-depth analysis of human motion can more efficiently take place on this reduced feature space.

Figure 1: A 3D representation of the depth data (Left). A 2D greyscale representation of the depth data (Centre). The RGB colour image (Right). These images are all captured from a single time step from the Xbox One Kinect.

It should be noted that when Kinect is tracking a body, joints are classified as either tracked or inferred. A joint is classed as tracked when confidence in the data is high i.e. there is little or no occlusion of the point cloud data surrounding the joint. If there is full or significant occlusion of the point cloud data surrounding the joint, its coordinates are classed as inferred.

3.2 Clinical/Qualitative Analysis Methodology

The gross upper-body exercises selected for analysis from the GRASP manual were, as labelled in the manual: Arm to Side, Arm to Front, Shoulder Shrug, Twist and Drying Off (Figure 2). They range from relatively simple exercises, e.g. arm to side, to more difficult exercises, e.g. drying off, which requires multiple limbs and a towel to be used. All exercises were recorded from a frontal view as this is the expected view for observing patients. As we are investigating the pose estimation accuracy in the context of upper-body stroke rehabilitation applications, pose positions below the hips and on the hands were not considered.

Figure 2: The five exercises selected from the GRASP manual for evaluation. Arm to Side (Top Left): Shoulder joint is abducted to 90 degrees and then adducted back to 0 degrees. Arm to Front (Top Middle): Shoulder joint is flexed to 90 degrees and then extended back to 0 degrees. Shoulder Shrug (Top Right): Shoulder joints are elevated and then depressed. Twist (Bottom Left): Shoulders are flexed to 90 degrees and hands are clasped, thorax is rotated towards 90 degrees in one direction then returned to starting position and rotated towards 90 degrees in the opposite direction. Drying off (Bottom Right): Towel is grasped and placed behind the neck, arm extension and flexion is performed along the frontal plane.

To perform a clinical analysis of the pose estimations, recordings of a user performing the GRASP exercises were captured in Kinect Studio. The physiotherapists watched a recording of each exercise with the joint skeletal data overlaid. The exercise was first performed correctly, and then repeated with each of the common compensatory movements (as listed in Table 1) deliberately included. So, for example, the 'Drying Off' exercise was performed 5 times; correctly and then 4 separate faulty versions. The physiotherapists made observations on the accuracy of the pose estimations using the key in Table 2, noting accuracy of the joint positions relevant to assessing performance of a stroke patient.

Table 1: List of exercises the physiotherapists observed including the associated common compensatory movements.

Exercise	Associated Common Compensatory Movements
Arm to side	Trunk lateral flexion, Shoulder elevation, Thorax rotation, Arm flexion
Arm to front	Trunk backward flexion, Shoulder elevation, Thorax rotation, Arm flexion
Shoulder shrug	Head side flexion, Shoulder abduction
Twist	Trunk lateral flexion, Arm flexion
Drying off	Trunk lateral flexion, Shoulder elevation, Dipped arm, Head side flexion

3.3 Quantitative Analysis Methodology

For the quantitative assessment, the ground truth (GT) was provided by a passive retro-reflective marker based motion capture system (Qualisys, 2015). This system captured the exercises at 240 Hz simultaneously with Kinect. The participants sat in an armless chair ~2 metres from the Kinect. Previous work has indicated that at this distance, Kinect has an average depth accuracy error of less than 2mm (Yang, Zhang, Dong, Alelaiwi, & Saddik, 2015). 14mm passive reflective markers were placed on the centre of the anatomical joints to be tracked. Where Kinect's counterpart anatomical joint is unclear, e.g. SpineMid, the markers were placed over the top of the Kinect joint while the user was in a seated t-pose posture, as shown in Figure 3. Multiple markers were used on certain joints to determine the location of the centre of the joint. For example, two markers were placed on the front and back of each shoulder and the position between the two markers were calculated to get the joint centre.

The Kabsch algorithm (Kabsch & IUCr, 1976) was used for rotational alignment of the datasets in the X and Y coordinates. It minimises the RMS deviation between the HipLeft, HipRight and SpineShoulder joints for both datasets at the t-pose posture frames. As the markers are visually placed on top or near their counterpart Kinect joints, rotational alignment is accurate for the X and Y positions. To find a good rotational alignment for the datasets in the Z position, a reference frame was defined when the user's arms were by their side, and the Qualisys dataset rotated around the X axis by 0.5 degrees to find the minimum difference in WristLeft Z position between the Kinect dataset and Qualisys dataset on the reference frame and t-pose frame.

After the alignment of the two datasets to accurately calculate the standard deviation (SD) and mean error of Kinect's joint positions, a seated t-pose posture was selected as the GT frame of Kinect's joints (see Figure 3). This posture presents Kinect's pose estimation with little difficulty. The joint SDs were modelled as ellipsoids to enable visualisation of the variance/jitter of each joint in all axes (Wang et al. 2015). The exercises were performed by 5 volunteers, and the calculated results averaged.

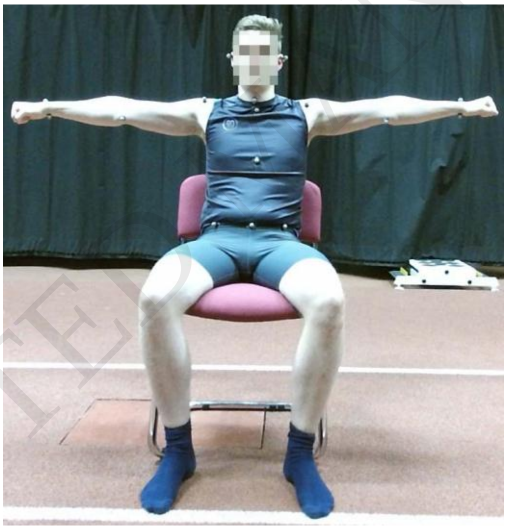


Figure 3: Seated T-Pose posture used as Kinect's ground truth for determining the SD and mean error of joint positions over an exercise.

4. RESULTS AND DISCUSSION

4.1 Qualitative Analysis Results

Four practising physiotherapists analysed the accuracy of the joint position estimations for each GRASP exercise. Their assessments are presented in Table 3. For samples from the clinical evaluation sessions see ("Clinical Evaluation of Depth Sensor - YouTube," n.d.). Physiotherapists were free to comment on any joint, using their clinical judgement to decide what was worthy of comment. Because of time constraints, only P1 watched all of the exercises (5 correct and a total of 16 with compensatory movement). Where P1 made no comment on an exercise, they confirmed that it was because they considered the pose tracking to be acceptable for assessing the exercise with regards to the compensatory movements. However they did explicitly comment on some of the exercises that were wholly acceptably tracked, and these are shown in the table. The exercises where P1 had made some comment about the tracking quality were presented individually to the other physiotherapists (without any indication of each other's views) to determine whether they also considered the tracking to have some problems. To calculate the mean (M) and standard deviation (SD), the categories AT, MT and UT from Table 2 were given a value of 1, 2 and 3 respectively.

Table 2 Key used by the physiotherapists for the evaluation of the joint position estimations in Table 3.

Key	
Acceptable tracking (AT)	A joint's estimated positions' result in an acceptable difference from the true position. The error in the position does not lead to misclassification in the assessment, e.g. a limb is showing no flexion when no flexion is occurring.
Moderately acceptable tracking (MT)	A joint's estimated positions' result in a moderately acceptable difference from the true position. The error in the position leads to a minor misclassification in the assessment, e.g. a limb is showing minor flexion when no flexion is occurring.
Unacceptable tracking (UT)	A joint's estimated positions' result in an unacceptable difference from the true position. This change in position leads to a significant misclassification in the assessment, e.g. a limb is showing severe flexion when no flexion is occurring.

Table 3 Physiotherapists' evaluation of Kinect's Pose Estimation for each GRASP exercise.

Description	P1	P2	P3	P4	M	SD	Comments
ElbowRight joint "Arm to Side" All versions	MT	AT	AT	AT	AT	0.43	**P1:** Jitter occurs along the axis of the bone, resulting in variable limb lengths.
WristRight joint "Arm to Side" All versions	MT	AT	AT	AT	AT	0.43	**P1:** Jitter occurs along the axis of the bone, resulting in variable limb lengths.
SpineMid joint "Arm to Side" Trunk lateral flexion	AT	MT	AT	AT	AT	0.43	**P2:** Angle around SpineMid joint is represented as a straight line when trunk flexion is occurring.
Hip joints "Arm to Side" Trunk lateral flexion	UT	AT	AT	MT	MT	0.83	**P1:** Hip joints give the impression that one hip is being lifted from the seat. **P4:** Hip joints showing exaggerated movements than is true.
Shoulder joints "Arm to Side" Shoulder elevation	UT	MT	AT	MT	MT	0.71	**P1:** Roughly 25% of the vertical movement is reported in the joint. **P2:** Shoulder position not accurately portraying severity of shoulder elevation. **P4:** Shoulder elevation is visible but not to the extent that is true.
ShoulderRight joint "Arm to Front" All versions except trunk backward flexion and shoulder elevation	MT	AT	AT	AT	AT	0.43	**P1:** As the arm reaches 90 degrees the shoulder joint drops to the axilla this gives the impression the arm is at a higher angle than is true.
ElbowRight joint "Arm to Front"	MT	AT	UT	MT	MT	0.71	**P1:** Jitter occurs when joint is occluded resulting in elbow flexion when the arm is straight. Joint also reports different limb lengths. **P3:** Incorrectly displaying elbow flexion when the arm is raised. **P4:** Jitter can cause confusion with knowing whether the patient kept their arm straight during the exercise.

All versions except elbow flexion							
WristRight joint "Arm to Front" All versions	MT	AT	UT	MT	MT	0.71	**P1:** Jitter occurs when joint is occluded resulting in elbow flexion when the arm is straight. Joint also reports different limb lengths. **P3:** Incorrectly displaying elbow flexion when the arm is raised. Incorrectly showing flexion extension in the wrist. **P4:** Jitter can cause confusion with knowing whether the patient kept their arm straight during the exercise.
Shoulder joints "Arm to Front" Trunk backward flexion	UT	UT	UT	UT	UT	0	**P1:** The shoulders track inwards severely when this is not the case, thus falsely reporting elbow flexion. **P3:** Visually looks like the shoulder joints move towards the torso as trunk backward flexion occurs.
SpineMid joint "Arm to Front" Trunk backward flexion	AT	MT	AT	AT	AT	0.43	**P2:** Angle around SpineMid joint is represented as a straight line when trunk flexion is occurring.
Shoulder joints "Arm to Front" Shoulder elevation	UT	UT	UT	UT	UT	0	**P1:** UT occurs when the arms occlude the shoulder. **P2:** Shoulder dips down as the arms occlude the shoulder. **P3:** Initially elevates but when the shoulder is occluded by the arm, the shoulder joint depresses. **P4:** Not clear shoulder elevation is occurring.
Elbow joints "Arm to Front" Elbow flexion	UT	AT	AT	MT	MT	0.83	**P1:** When the elbow is flexed, jitter occurs even when the joint is not occluded. **P4:** Jitter can cause confusion with knowing whether the patient kept their arm straight during the exercise.
Shoulder joints "Shoulder Shrug" All versions	UT	UT	MT	MT	MT /UT	0.5	**P1:** Only a minor vertical movement when elevating the shoulders. **P2:** Minor shoulder elevation tracked when significant shoulder elevation occurring. **P3:** Not showing elevation to the degree the shoulders are. **P4:** Can see some elevation but not showing the range.
Wrist joints "Shoulder Shrug" All versions	MT	AT	AT	AT	AT	0.43	**P1:** Jitter occurs when joint becomes occluded by the legs.
Hips and SpineBase joints "Shoulder Shrug" All versions	AT	AT	AT	MT	AT	0.43	**P1:** Acceptable jitter can be seen, they also slightly elevate as the shoulders are lifted even though the true joint positions remain still. **P4:** Hips elevate with the shoulders.
Neck joint "Shoulder Shrug" Head Flexion	MT	MT	UT	AT	MT	0.71	**P1:** Reporting only minor head lateral flexion when severe. **P2:** Not correctly showing the severity of head flexion. **P3:** Not able to interpret the head and neck markers as flexion.
Head joint "Shoulder Shrug" Head Flexion	MT	MT	UT	MT	MT	0.43	**P1:** When severe head lateral flexion occurs, the joint has MT, resulting in reporting a minor head lateral flexion. **P2:** Not correctly showing the severity of head flexion. **P3:** Not able to interpret the head and neck markers as flexion.
SpineMid joint "Shoulder Shrug" Shoulder abduction	MT	MT	UT	MT	MT	0.43	**P1:** Falsely reporting minor trunk lateral flexion, when no trunk lateral flexion occurring.
Shoulder joints "Twist" All versions	UT	AT	MT	UT	MT	0.83	**P1:** Joints track around the axilla as the arms are raised to 90 degrees. **P4:** When arms raised shoulders become depressed down to the rib cage.
Elbow joints "Twist" All versions	MT	AT	MT	AT	AT/ MT	0.5	**P1:** Joint showing jitter and variable limb lengths during the exercise.

Wrist joints "Twist" All versions	MT	AT	UT	AT	MT	0.83	**P1:** Joint showing jitter and variable limb lengths during the exercise.
SpineShoulder, Head and Neck joints "Twist" All versions	UT	AT	MT	UT	MT	0.83	**P1:** Joints incorrectly track vertically as the joints are occluded by the arms. **P3:** Rotation is inferred by arm joint positions. **P4:** Joints elevated when occlusion occurs.
SpineMid joint "Twist" Trunk lateral flexion	AT	UT	AT	AT	AT/ MT	0.87	**P2:** Angle around SpineMid joint is represented as a straight line when trunk flexion is occurring.
Elbow joint "Twist" Trunk lateral flexion	UT	MT	MT	MT	MT	0.43	**P1:**Unacceptable jitter. **P2:** Jitter occurring. **P4:** Shows more flexion than is occurring during some of the exercise.
Wrist joint "Twist" Trunk lateral flexion	UT	MT	MT	AT	MT	0.71	**P1:** Unacceptable jitter. **P2:** Jitter occurring.
Shoulder joints "Drying Off" All versions	UT	UT	MT	UT	UT	0.43	**P1:** Unacceptable jitter and tracking on the towel. **P2:** Joint incorrectly tracks on towel. **P3:** Joint positions briefly glitch onto the towel. **P4:** Unacceptable because the joint occasionally tracks on the towel.
Elbow joints "Drying Off" All versions	UT	UT	MT	MT	MT /UT	0.5	**P1:** Unacceptable jitter and tracking on the towel. **P2:** Joint incorrectly tracks on towel. **P3:** Joint positions briefly glitch onto the towel. **P4:** Unacceptable because the joint occasionally tracks on the towel.
Wrist joints "Drying Off" All versions	UT	AT	AT	MT	MT	0.83	**P1:** Unacceptable jitter and tracking on the towel.
Hip and SpineBase joints "Drying Off" All versions	UT	AT	MT	UT	MT	0.83	**P1:** The hips and SpineBase joints show UT in the vertical axis. **P4:** Joints move around during the exercise.
SpineMid joint "Drying Off" Trunk lateral flexion	UT	UT	UT	UT	UT	0	**P2:** Angle around SpineMid joint is represented as a straight line when trunk flexion is occurring. **P4:** SpineMid does not move. Not showing any side flexion.

As can be seen from Table 3, each of the exercises resulted in some undesirable aspects in the tracking. Even the more straightforward exercises such as 'Arm to Side', which would have little or no occlusion, caused some issues. Problems occurred with jitter at the elbow and wrist joints, which could give rise to variable bone lengths. Fernández-Baena et al. (2012) commented that fixed bone lengths might improve the joint accuracy. Trunk flexion caused problems throughout, partly due to occlusion. It was noted by several physiotherapists that to perform a correct analysis of the exercise, joint rotational information is required. This was noted when assessing trunk flexion during the twist exercise.

Of more clinical interest is the variation between the opinions of the physiotherapists. This may partly be due to familiarity, as P1 spent much longer analysing the results. The highest variation came from the SpineMid joint for the twist exercise while trunk lateral flexion occurred, where P2 rated the joint unacceptable, noting that the spine was not showing flexion, while the others rated it acceptable. The mean categorisation shows that exercises with objects or substantial occlusion leads to unacceptable or moderately acceptable tracking and therefore can be difficult to correctly assess. On occasions joint position estimations would result in an anatomically impossible pose, for example shoulder joints tracking inwards towards the spine as trunk backwards flexion occurs.

4.2 Quantitative Analysis Results

The joint names described in this section are taken from Kinect SDK. The "Twist" exercise had a relatively high SD, as shown in Table 4 and Figure 4, and shows how the pose estimation algorithm struggles with poses with limited depth data of the joint and surrounding areas, such as when the arms were extended towards the depth sensor. When comparing the exercises "Arm to Side" and "Shoulder Shrug" against the exercises "Arm to Front" and "Twist", limb joint positions, for exercises performed along the Y and Z axes, are more inaccurate than along the X and Y axes. This appears to be due to the unavoidable occlusion.

Table 4: Table showing the error and SD for each joint position estimation averaged over all repetitions. The right arm was used for the arm to side and arm to front exercises.

Joint	Arm to Side		Arm to Front		Shoulder Shrug		Twist		Drying Off	
	Error	SD	Error	SD	Error	SD	Error	SD	Error	SD
SpineBase	1.14	0.34	1.18	0.34	2.42	1.19	6.80	3.51	3.09	0.98
SpineMid	1.13	0.33	1.78	0.37	2.73	1.00	8.39	5.15	4.62	1.60
Neck	0.77	0.24	1.12	0.27	1.27	0.45	6.53	3.75	2.78	1.17
Head	0.61	0.15	0.57	0.21	1.36	0.66	4.80	2.55	3.71	1.72
ShoulderLeft	1.47	0.23	1.56	0.38	2.71	0.97	8.21	5.48	4.90	1.91
ElbowLeft	4.58	0.45	3.69	0.24	3.89	1.11	15.44	5.41	5.78	1.99
WristLeft	6.77	0.98	5.61	0.28	6.21	1.54	18.11	5.63	10.80	3.39
ShoulderRight	1.18	0.47	2.37	1.02	3.06	1.47	10.91	6.13	4.99	2.48
ElbowRight	2.35	1.15	11.48	6.19	3.75	1.12	16.84	6.58	5.99	2.67
WristRight	3.27	1.57	15.45	6.69	5.60	1.53	20.47	7.09	14.84	5.20
HipLeft	1.36	0.39	1.51	0.28	2.81	1.04	7.90	4.10	3.58	0.98
HipRight	1.31	0.52	1.99	0.38	2.83	0.83	7.90	4.00	3.46	1.01
SpineShoulder	1.08	0.28	1.79	0.33	2.30	1.01	6.48	3.51	3.35	1.45

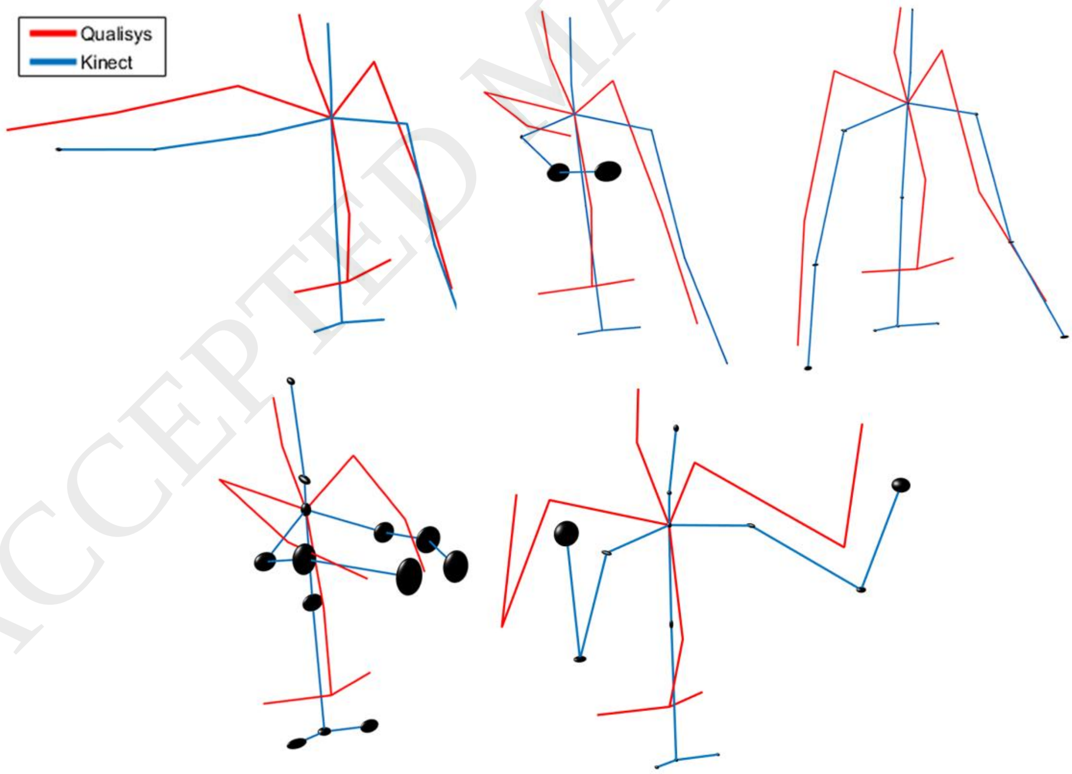


Figure 4 depicts the SD of the error for all repetitions of that exercise modelled as ellipsoids. Exercise order from top left; Arm to Side, Arm to Front, Shoulder Shrug, Twist, Drying Off.

In Figure 5, the error of the joint positions are larger for the arm to front exercise than the arm to side, this is understandable, as there would be limited depth data for the arm as it is raised to 90 degrees.

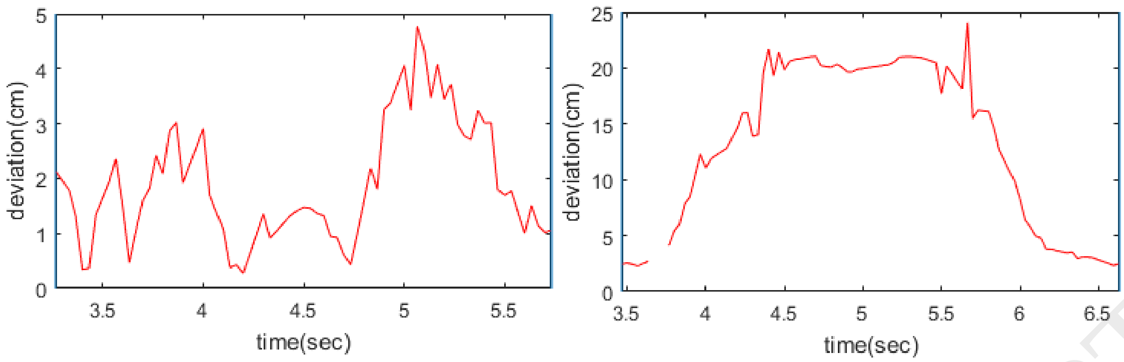


Figure 5: Plots showing a participant's WristRight joint deviation from the ground truth joint position. Arm to side (left) and arm to front (right).

Figure 6 shows the algorithm struggling to track the true movement of the shoulder joint even though there is no occlusion in the depth data around the shoulder joint. This could be due to the pose estimation algorithm being trained on a dataset containing no or limited data of correctly labelled elevated shoulder joints.

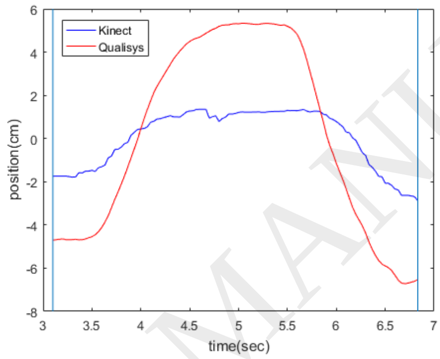


Figure 6: Plot showing a participant's ShoulderRight joint position in the Y axis when performing the shoulder shrug exercise.

Figures 7-9 show the errors for each exercise repetition for the shoulder, elbow and wrist joints - these are the main important joints for assessing the exercises. Figure 7 shows the shoulder joint has a relatively large error when being tracked during the twist exercise. Figure 8 shows the elbow joint with relatively large error on the arm to front and twist exercise. Figure 9 shows the wrist joint has the largest mean error when compared to the elbow and shoulder joint. It also has a relatively large error during the arm to front, twist and drying off exercises.

Interestingly in Figure 7 the ShoulderRight joint does not appear to be relatively erroneous for the shoulder shrug exercise, but the physiotherapists reported UT and MT for this joint on this exercise. This suggests absolute joint error is not a definitive measure of acceptability.

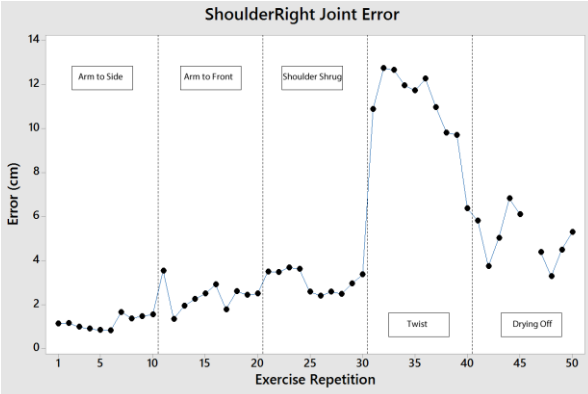


Figure 7: Exercise repetition errors of the ShoulderRight joint.

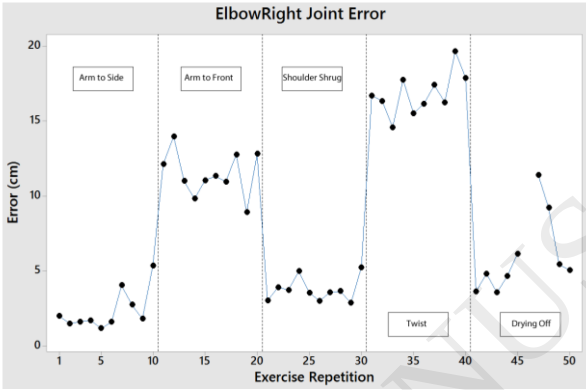


Figure 8: Exercise repetition errors of the ElbowRight joint.

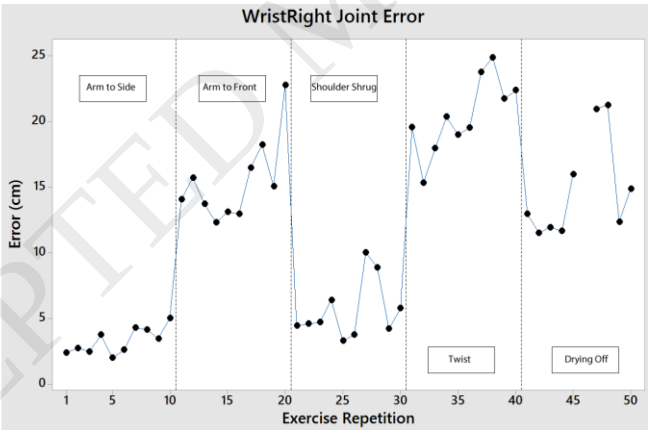


Figure 9: Exercise repetition errors of the WristRight joint.

5. CONCLUSION AND FUTURE WORK

Based on the clinical analysis supported by the quantitative measures we conclude that the pose estimations are mostly inadequate for correctly assessing stroke rehabilitation exercises.

When performing upper-body gross exercises the shoulder joints act as indicators for incorrect movement of limbs. For example, elevated shoulders are a common compensatory movement among stroke patients and needs to be detected during rehabilitation exercises. However, the algorithm failed to accurately track the true movement of the shoulder joints even when the joints were in a tracked state. This could be improved by retraining the pose estimation algorithm with correctly labelled shoulder joints that contain training data with elevated shoulders.

Partial or full occlusion in the depth data surrounding a joint causes unacceptable jitter and tracking. When objects are required for performing an exercise (eg the towel used in "Drying off"), unacceptable jitter and tracking can also occur as the algorithm can transiently misclassify it as a body part, resulting in incorrectly tracked joint positions. Similarly, for seated exercises it is recommended a user be seated on a perching stool to eliminate the chances of seat arms being misclassified as joints.

When assessing the suitability of a pose estimation algorithm intended for rehabilitation applications, solely performing a quantitative measure does not provide conclusive answers, a clinical analysis optionally supported by a quantitative measure is required to determine suitability. This is because the measured accuracy of the joint estimations does not take into account that joints require a varying degree of accuracy to correctly assess a given exercise. This is evident by the shoulder joints for the arm to front and shoulder shrug exercise, whereby the ShoulderRight joint displays similar error in Figure 7 for these exercises but has a mean classification of AT and MT/UT respectively from the clinicians presented in Table 3.

Future pose estimation algorithms should consider using temporal information and extrapolating from previous frames for inferring joint positions that have full or partial occlusion. This should also reduce the possibility of inferring a joint position incorrectly by ruling out sudden and extreme changes in position. We are currently working on techniques that use temporal information in a scalable way to improve joint tracking. Estimating joint rotation should be considered for a more in-depth and correct assessment of a patent's performance. Constraining joint estimations to within the anatomical limits of the human body should ameliorate severe tracking error and solve the issue of anatomically impossible poses.

In order to make the task of automatically assessing for compensatory movements easier, clinicians should be consulted to try to select exercises that are useful for rehabilitation but provide fewer or easier challenges for pose estimation algorithms. For example, clinicians highlighted using less obtrusive objects such as a rod or walking stick to perform the drying off exercise, resulting in fewer tracking errors.

Author Contribution

Authors: Joe Sarsfield, David Brown, Nasser Sherkat, James Lewis, Mohammad Taheri, Cleveland Barnett, Louise Selwood, Penny Standen, Pip Logan, Caroline Langensiepen, Christopher McCollin, Christopher Simcox, Catherine Killick, Emma Hughes

Conception or design of the work

Joe Sarsfield, David Brown, Nasser Sherkat, James Lewis, Mohammad Taheri, Cleveland Barnett, Louise Selwood, Penny Standen, Pip Logan

Data collection

Joe Sarsfield, Louise Selwood, David Brown, Nasser Sherkat, James Lewis, Mohammad Taheri, Cleveland Barnett

Data analysis and interpretation

Joe Sarsfield, David Brown, Nasser Sherkat, James Lewis, Mohammad Taheri, Louise Selwood, Penny Standen, Pip Logan, Christopher McCollin, Christopher Simcox, Catherine Killick, Emma Hughes

Drafting and/or Revising the article

Joe Sarsfield, David Brown, Nasser Sherkat, James Lewis, Mohammad Taheri, Caroline Langensiepen, Cleveland Barnett, Louise Selwood, Penny Standen, Pip Logan, Christopher McCollin, Christopher Simcox, Catherine Killick, Emma Hughes

Final approval of the version to be published

Joe Sarsfield, David Brown, Nasser Sherkat, James Lewis, Mohammad Taheri, Caroline Langensiepen, Cleveland Barnett, Louise Selwood, Penny Standen, Pip Logan, Christopher McCollin, Christopher Simcox, Catherine Killick, Emma Hughes

Conflict of Interest

Mr. Sarsfield has nothing to disclose.

Summary Points

What was known before this study?

- Consumer-level depth sensors combined with pose estimation algorithms have potential for practical technology supervised rehabilitation applications.
- Current pose estimation algorithms designed to run on consumer-level depth sensors are not specifically designed for rehabilitation applications and therefore the pose estimations need to be evaluated in this context to determine suitability.

What did this study add to our body of knowledge?

- Based on the clinical analysis supported by the quantitative measures we conclude that the pose estimations are mostly inadequate for correctly assessing stroke rehabilitation exercises.
- Future pose estimation algorithms intended for rehabilitation applications should consider; exploiting temporal information to ameliorate issues with occlusion, constraining joint estimations to within the anatomical limits of the human body, estimating joint rotational information for a more in-depth and correct assessment.
- A methodology for clinically analysing the performance of a pose estimation algorithm for use in rehabilitation applications.

REFERENCES

Burke, J. W., McNeill, M. D. J., Charles, D. K., Morrow, P. J., Crosbie, J. H., & McDonough, S. M. (2009). Optimising engagement for stroke rehabilitation using serious games. *The Visual Computer*, *25*(12), 1085–1099. https://doi.org/10.1007/s00371-009-0387-4

Clinical Evaluation of Depth Sensor - YouTube. (n.d.). Retrieved November 9, 2017, from https://www.youtube.com/watch?v=_oMcZC8rilM&feature=youtu.be

Fernández-Baena, A., Susín, A., & Lligadas, X. (2012). Biomechanical validation of upper-body and lower-body joint movements of kinect motion capture data for rehabilitation treatments. *Proceedings of the 2012 4th International Conference on Intelligent Networking and Collaborative Systems, INCoS 2012*, 656–661. https://doi.org/10.1109/iNCoS.2012.66

Flores, E., Tobon, G., Cavallaro, E., Cavallaro, F. I., Perry, J. C., & Keller, T. (2008). Improving patient motivation in game development for motor deficit rehabilitation. In *Proceedings of the 2008 International Conference in Advances on Computer Entertainment Technology - ACE '08* (p. 381). New York, New York, USA: ACM Press. https://doi.org/10.1145/1501750.1501839

Kabsch, W., & IUCr. (1976). A solution for the best rotation to relate two sets of vectors. *Acta Crystallographica Section A*, *32*(5), 922–923. https://doi.org/10.1107/S0567739476001873

Kurillo, G., Chen, A., Bajcsy, R., & Han, J. J. (2013). Evaluation of upper extremity reachable workspace using Kinect camera. *Technology and Health Care*, *21*(6), 641–656. https://doi.org/10.3233/THC-130764

Microsoft. (n.d.). Kinect for Windows features. Retrieved August 6, 2015, from https://www.microsoft.com/en-us/kinectforwindows/meetkinect/features.aspx

Mobini, A., Behzadipour, S., Foumani, M. S., & Saadat Foumani, M. (2013). Accuracy of Kinect's skeleton tracking for upper body rehabilitation applications. *Disability and Rehabilitation. Assistive Technology*, *9*(4), 1–9. https://doi.org/10.3109/17483107.2013.805825

Murray, C. J. L., Vos, T., Lozano, R., Naghavi, M., Flaxman, A. D., Michaud, C., … Lopez, A. D. (2012). Disability-adjusted life years (DALYs) for 291 diseases and injuries in 21 regions, 1990-2010: A systematic analysis for the Global Burden of Disease Study 2010. *The Lancet*, *380*(9859), 2197–2223. https://doi.org/10.1016/S0140-6736(12)61689-4

Obdrzalek, S., Kurillo, G., Ofli, F., Bajcsy, R., Seto, E., Jimison, H., & Pavel, M. (2012). Accuracy and robustness of Kinect pose estimation in the context of coaching of elderly population. *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS*, 1188–1193. https://doi.org/10.1109/EMBC.2012.6346149

Qualisys. (2015). A complete Running Analysis solution. Retrieved from http://www.qualisys.com/applications/sports/running/

Rego, P., Moreira, P. M., & Reis, L. P. (2010). Serious Games for Rehabilitation A Survey and a Classification Towards a Taxonomy. *5th Iberian Conference on Information Systems and Technologies*, (November 2015), 1–6. https://doi.org/978-1-4244-7227-7

Shaughnessy, M., Resnick, B. M., & Macko, R. F. (2006). Testing a model of post-stroke exercise behavior. *Rehabilitation Nursing : The Official Journal of the Association of Rehabilitation Nurses*. https://doi.org/10.1002/j.2048-7940.2006.tb00005.x

Shotton, J., & et al. (2011). Real-time human pose recognitiom in parts from single depth images. In *IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*. IEEE.

Suma, E. A., Lange, B., Rizzo, S., Krum, D., & Bolas, M. (n.d.). FLEXIBLE ACTION AND ARTICULATED SKELETON TOOLKIT. Retrieved January 1, 2016, from http://projects.ict.usc.edu/mxr/faast/

Wang, Q., Wang, Q., & Kurillo, G. (2015). Evaluation of Pose Tracking Accuracy in the First and Second Generations of Microsoft Kinect Evaluation of Pose Tracking Accuracy in the First and Second Generations of Microsoft Kinect, (November).

Webster, D., & Celik, O. (2014). Experimental evaluation of Microsoft Kinect's accuracy and capture rate for stroke rehabilitation applications. *IEEE Haptics Symposium, HAPTICS*, (February 2014), 455–460. https://doi.org/10.1109/HAPTICS.2014.6775498

Xu, X., & McGorry, R. W. (2015). The validity of the first and second generation Microsoft Kinect™ for identifying joint center locations during static postures. *Applied Ergonomics*, *49*, 47–54. https://doi.org/10.1016/j.apergo.2015.01.005

Yang, L., Zhang, L., Dong, H., Alelaiwi, A., & Saddik, A. El. (2015). Evaluating and Improving the Depth Accuracy of Kinect for Windows v2. *IEEE Sensors Journal*, *15*(8), 4275–4285. https://doi.org/10.1109/JSEN.2015.2416651

Yeung, L. F., Cheng, K. C., Fong, C. H., Lee, W. C. C., & Tong, K.-Y. (2014). Evaluation of the Microsoft Kinect as a clinical assessment tool of body sway. *Gait & Posture*, *40*(4), 532–8. https://doi.org/10.1016/j.gaitpost.2014.06.012