Markerless motion capture and measurement of hand kinematics: validation and application to home-based upper limb rehabilitation

Cheryl D. Metcalf, *Member, IEEE**, Rebecca Robinson, Adam J. Malpass, Tristan P. Bogle, Thomas A. Dell, Chris Harris & Sara H. Demain

***Abstract*— Dynamic movements of the hand, fingers and thumb are difficult to measure due to the versatility and complexity of movement inherent in function. An innovative approach to measuring hand kinematics is proposed and validated. The proposed system utilises the Microsoft Kinect™ and goes beyond gesture recognition, to develop a validated measurement technique of finger kinematics. The proposed system adopted landmark definition (validated through ground truth estimation against assessors) and grip classification algorithms, including kinematic definitions (validated against a laboratory-based motion capture system). The results of the validation show 78% accuracy when identifying specific markerless landmarks. In addition, comparative data with a previously validated kinematic measurement technique show accuracy of MCP±10° (average absolute error (AAE) = 2.4°), PIP±12° (AAE = 4.8°) and DIP±11° (AAE = 4.8°). These results are notably better than clinically based alternative manual measurement techniques. The ability to measure hand movements, and therefore functional dexterity, without interfering with underlying composite movements, is the paramount objective to any bespoke measurement system. The proposed system is the first validated markerless measurement system using the Microsoft Kinect™ that is capable of measuring finger joint kinematics. It is suitable for home-based motion capture for the hand and therefore achieves this objective.**

***Index Terms*— Microsoft Kinect, Hand Kinematics, Markerless, Telerehabilitation.**

I. Introduction

HAND movements and hand function are intrinsic to quality of life. Dexterous ability is fundamental to gesture, communication, independence, and manipulation of, and interaction with, objects and the environment. When functional ability is impaired, the prescription of rehabilitation exercises at home is a solution that can increase the intensity of practice and optimize recovery potential.

The repetitive nature of home-based rehabilitation programs, such as those for stroke patient care, can be monotonous and often difficult for individuals to complete. Assessment of practice is essential for monitoring recovery, adapting prescriptions based on improvement or decline, and therefore providing more patient-centered care. Home-based rehabilitation and telecare are advocated by many government health organizations [1]-[3], to increase the throughput of patients for over-subscribed international healthcare systems. This however, is yet to be realized in mainstream healthcare.

One important limiting factor is the current inability to monitor and measure hand movements. Systems capable of measuring the fine dexterity of the hand are therefore required. Rehabilitation professionals require detailed data to monitor clinical progress and modify treatments, whilst patients need real-time feedback to correct their movements and to stimulate motivation. Any telecare system must therefore provide enough feedback to replace verbal feedback usually provided by therapists.

Motion capture systems have been used in a wide range of industries, including medical, occupational, sports and entertainment [4]-[6]. Four motion capture systems are currently used for hand capture; instrumented gloves [7], wrist-worn laser systems [8], inertial systems [7] and traditional optical systems [9], [10]. Glove-based systems provide a simple, quick measurement of hand position and do not suffer from occlusions. However, wearing and removing gloves can be difficult or impossible for

Manuscript received November 19, 2012. This work was supported by the University of Southampton. *Asterisk indicates corresponding author.*
R Robinson, A.J. Malpass, T.P. Bogle and T.A. Dell were with Electronics & Computer Sciences, University of Southampton, SO17 1BJ, UK.

C. Harris & R. Robinson are with Roke Manor Research Ltd. SO51 0ZN, UK.

*C.D. Metcalf & S. H. Demain are with the Faculty of Health Sciences, University of Southampton, UK. (Phone: +44 2380 598927; e-mail: c.d.metcalf@soton.ac.uk).

patients with hand deformities, spasticity and contractures. In contrast, inertial systems, such as the Nintendo Wii are widely used in rehabilitation centers [11]. However, these allow 'trick' movements, i.e. flicking the wrist to substitute whole-arm movements and therefore have limited application for unsupervised home-based rehabilitation [12].

Traditional optical or camera-based systems, considered the gold-standard, are predominantly laboratory based. They require markers attached to the participant's skin and are placed according to the underlying human anatomy, acting as reference points (landmarks). Segmental movement can then be defined and joint angles calculated; an example of marker placement for a human hand is shown in [9]. These systems generate very accurate results [13], [14] and the associated kinematic measurement techniques show repeatable errors from 1-8° [9], [15], [16]. However the feasibility of adopting these systems within a home environment is limited, not least by physical and economic constraints.

II. Markerless Motion Capture

The development of a markerless motion capture system is presented using a commercially available gaming system (Microsoft Kinect™). Comparable systems do exist for whole body tracking [17]-[20]; but require multiple cameras and hence lengthy post-processing. The proposed system will focus specifically on capturing hand and finger movements. Previous research has been undertaken in this area [21]-[23], but these systems only provide evidence of fingertip detection. Research on gesture recognition has also been performed [24], [25], but these systems do not provide kinematic measurement. For rehabilitation purposes, clinicians require detailed and accurate measurement of the hands in order to assess progress and functional recovery. The patient also requires real-time feedback of their hand posture if they are going to participate in a 'virtual' or online rehabilitation gaming platform.

III. System Overview

The initial phase of system development is presented to investigate whether the Microsoft Kinect™ was capable of tracking hand movements. An overview of the system is presented that identifies anatomical landmarks, classification between grip types and calculates joint angles from a kinematic model. The initial phase of the system was able to detect two different modes: spread hand mode and pincer grip mode. The system output was tested using a two-point validation procedure: 1) using ground truth estimation between reviewers and the associated algorithms to assess the accuracy of landmark definition (or the identification of an anatomical point of interest), and 2) joint angles generated were compared against a laboratory-based gold standard motion capture system and validated kinematic measurement technique.

The system requires a Microsoft Kinect™ suspended on a rig above a table (optimal height for reliable capture was 80cm on an adjustable rig 50 – 125cm), allowing the reviewer to place their hands above the table to use the system (see Fig. 1). The Microsoft Kinect™ was selected, as it is a commercially available device that is accessible to the public. It contains both infrared depth and color cameras and has a well-documented Application Programming Interface (API). The Kinect was used in default mode, (0.8 – 4m) rather than near mode for the field of view. The depth camera was used to identify the hand and fingers. In commercial versions of the device, only gross body segments, such as upper and lower limbs are detected. Depth data was segmented to isolate the user's hand from background surfaces, such as a table or surrounding furniture. This required an initial calibration to set the height at which segmentation was optimally performed. The segmentation of the depth data generated a binary image (see Fig. 2). A contour generation algorithm was then executed on the binary image; producing a hand outline without any background errors (see Fig. 3).

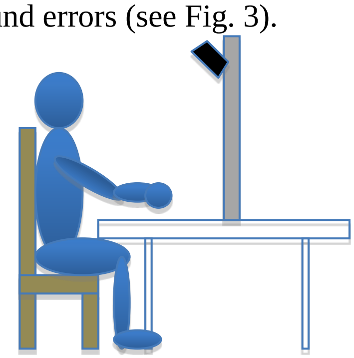

Fig. 1. System set-up showing participant, frame and Microsoft Kinect™.

1) *Ground Truth Validation Testing*

Ground truth estimation [26] was used to validate the system's landmark definitions; landmarks are directly compared to equivalent points generated from reviewer input. To account for human error, three trained reviewers in landmark identification performed the ground truth estimation in spread hand mode. If the correct pattern of points could not be found, then that frame was not tested or included in the results.

Spread hand mode (see Fig. 2) consisted of the hand (right or left) placed flat with the fingers separated. The fingertips were defined as points of highest convexity and extracted, as were the spaces between the digits (highest concavity).

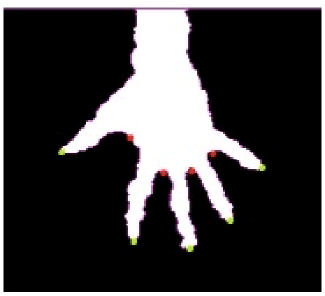

Fig. 2. Binary image extracted from the infrared depth camera data and minima and maxima of convexity and concavity depicted as points of interest.

Each reviewer tested 14 different videos (42 videos in total). These videos contained subtly different spread hand positions (moving finger positions or waving). Frames were presented to each reviewer in random order to reduce anticipation of feature point locations. The videos portrayed three different users' hands to increase the potential variability between frames and provide versatility that will be required of the final system implementation. For each video, the distance between the two co-ordinates (one generated by the algorithm, the other by the reviewer performing the ground truth estimation) was found. The average of these magnitudes was then taken from all the points found in the video. This was repeated for all spread hand mode videos and reviewers.

Pincer grip (see Fig. 3) consisted of an active movement where the thumb opposes the fingertips.

Four trained reviewers performed the ground truth estimation on 8 different pincer grip videos (33 videos in total).

A single factor analysis of variance (ANOVA) was undertaken to assess reviewer variability during ground truth estimation.

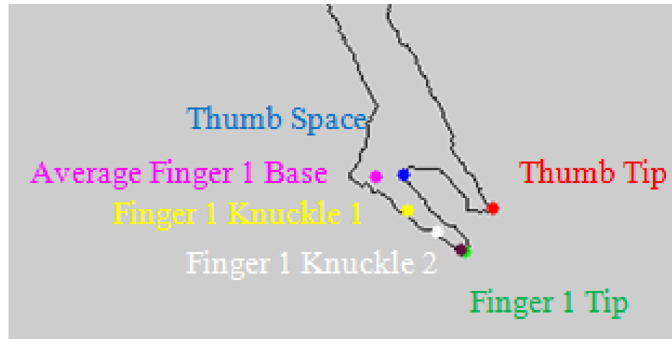

Fig. 3. A contour image extracted directly from the program of a hand in pincer mode position with labeled points of interest.

B. *Grip Classification*

Following the landmark definition, the system was extended to classify transitions between spread hand to pincer grip modes. The latter grip posture was chosen as it is a common functional movement strategy [27] often present in many activities of daily living, such as when reaching for a cup during drinking.

Points of high curvature were extracted from the contour around the hand and classified into "maxima" or "minima" depending on the sign of the curvature (convexity or concavity respectively). Transitions were then defined depending on the posture of the hand. To define the mode (posture), the program would look for a specific pattern of "maxima = fingertip" (FT) and "minima = finger spaces" (FS). The pattern for Spread Hand mode was FT, FS, FT, FS, FT, FS, FT, FS, FT. As this pattern is symmetrical, the same pattern was identifiable for both left and right hands.

For pincer grip mode, different patterns were searched depending on whether a left or right hand was detected. If a left hand was observed, the systems defined a pattern of FT, FS, FT & FT, with the points then being further classified as thumb tip, thumb space, base knuckle on first finger and first finger tip. For a right hand, the pattern was reversed, FT, FT, FS & FT.

The grip classification algorithm was designed to classify both spread hand and grip pincer modes following transitions between these two grips. This was an important consideration for development of the system in future phases; the seamless transition between grip postures is an important aspect of classifying the functional grips used.

1) *Joint Angle Algorithm*

The human finger is comprised of three joints; the metacarpophalangeal (MCP), proximal interphalangeal (PIP) and distal interphalangeal (DIP) joints. Many kinematic methods measure MCP finger joint movement as a hinge [9], [10], [15], although there is anatomically some rotational movement around the long axis of the finger. Given the negligible movement in the third dimension, and limited functional impact of the movement, the proposed system assumes hinge joints for the fingers.

Having defined this constraint, it was hypothesized that for a fixed base position of the MCP, *B*, there will be only one combination of $\theta_{MCP}, \theta_{PIP}$ and θ_{DIP} to place the fingertip, D, into a particular position. Thus, if the locations of B and D are known, in addition to the lengths of the phalangeal segments (l_{MCP}, l_{PIP} and l_{DIP}), then the joint angles j^1 and j^2 can be inferred.

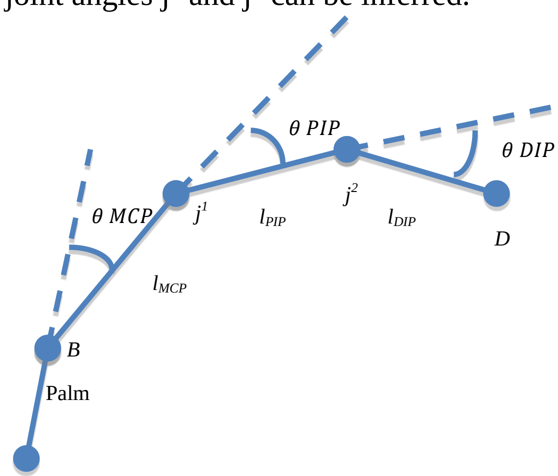

Fig. 4. Two-dimensional assumed kinematic model of the finger.

An algorithm was developed to experimentally verify this hypothesis; calculating every possible location of the fingertip, given the physical constraints of the model, for all angles of $\theta_{MCP}, \theta_{PIP}$ and θ_{DIP}. Adjusting the step size of each angle as the full range was traversed set the precision. The algorithm then analyzed the results for possible matches, which would signify that there was more than one possible set of joint positions that could determine D_k, where D_k is the output of the k-curvature algorithm (detailed below) [28]. *B* was set as the mid-point between the thumb space and the point of high negative curvature on the contour when the finger is in flexion. When the finger is in adduction, i.e. when the fingers and thumb are brought together, its value can no longer be directly inferred from the contour due to a lack of distinguishing features, so temporal analysis is performed to calculate the most likely value based on its previous positions. The phalanges' lengths can be measured by a clinician and input to the program before first-time use. These data are then stored in a database allowing the patient and clinician access to a personalised rehabilitation environment.

By comparing the number of matches found with the total number of unique positions, the overall certainty of the algorithm working was calculated for various levels of precision. A 'match' is defined as multiple sets of different joint angles resulting in the fingertip being located in the same position, whereas a unique position is one for which there is only one set of joint angles that can result in the fingertip being located in any one position, thus confirming the hypothesis. The total number of end states was calibrated to reflect the number of pixels in the area of interest that can be seen by the Microsoft Kinect™ and this was approximated as being 10% of its resolution of 640x480.

Therefore the principle of the hypothesis was to calculate all possible values of j_1, j_2 and the DIP tip, D¸ to find the optimal match with the DIP tip found by the k-contour algorithm, D_k. The values for θ_{MCP}, θ_{PIP} and θ_{DIP} are then generated, as are the positions of j_1 and j_2. All possible coordinates of j_1 are calculated using (1) for all values of θ_{MCP}, where B_x, B_y *are the x and y components of B respectively:*

$$j_1 \binom{x}{y}_{\theta_{MCP}} = \begin{pmatrix} B_x + l_{MCP} * \cos\theta_{MCP} \\ B_y + l_{MCP} * \sin\theta_{MCP} \end{pmatrix}$$

$$\theta_{MCP} \in \left\{\frac{\pi}{4} \rightarrow \frac{3\pi}{4}\right\} \quad (1)$$

Then all the possible coordinates of j_2 can be determined using (2), also accounting for all values of θ_{PIP}:

$$j_2 \binom{x}{y}_{\theta_{MCP},\theta_{PIP}} = \begin{pmatrix} j_1(x)_{\theta_{MCP}} + l_{PIP} * \cos(\theta_{MCP} - \theta_{PIP}) \\ j_1(y)_{\theta_{MCP}} + l_{PIP} * \sin(\theta_{MCP} - \theta_{PIP}) \end{pmatrix}$$

$$\theta_{MCP} \in \left\{\frac{\pi}{4} \rightarrow \frac{3\pi}{4}\right\},\ \theta_{PIP} \in \left\{0 \rightarrow \frac{\pi}{2}\right\} \quad (2)$$

All the possible coordinates of *D* can be found, additionally accounting for all values of θ_{DIP}. Due to the fixed size of the tendons in the finger, the amount through which θ_{DIP} can rotate anatomically is limited by θ_{PIP}:

$$D\binom{x}{y}_{\theta_{MCP},\theta_{PIP},\,\theta_{DIP}} = \begin{pmatrix} j_2(x)_{\theta_{MCP},\theta_{PIP}} + l_{DIP} * cos(\theta_{MCP} - \theta_{PIP} - \theta_{DIP}) \\ j_2(y)_{\theta_{MCP},\theta_{PIP}} + l_{DIP} * sin(\theta_{MCP} - \theta_{PIP} - \theta_{DIP}) \end{pmatrix}$$

$$\theta_{MCP} \in \left\{\frac{\pi}{4} \to \frac{3\pi}{4}\right\}, \theta_{PIP} \in \left\{0 \to \frac{\pi}{2}\right\}, \theta_{DIP} \in \{0 \to \theta_{PIP}\} \quad (3)$$

The algorithm then calculates the minimum distance between D_k and all possible values for D. This result corresponds to the system's best estimate of all joints' positions, and hence the optimum values of θ_{MCP}, θ_{PIP} and θ_{DIP}.

$$\min\left\{\sqrt{\left[D(x)_{\theta_{MCP},\theta_{PIP},\,\theta_{DIP}} - D_k(x)\right]^2 + \left[D(y)_{\theta_{MCP},\theta_{PIP},\,\theta_{DIP}} - D_k(y)\right]^2}\right\}$$
$$\theta_{MCP} \in \left\{\frac{\pi}{4} \to \frac{3\pi}{4}\right\}, \theta_{PIP} \in \left\{0 \to \frac{\pi}{2}\right\}, \theta_{DIP} \in \{0 \to \theta_{PIP}\} \quad (4)$$

Refinements were made to optimize accuracy and performance. Due to the data acquisition rate (30FPS), and the limits of human movement, the finger was unlikely to move very far between successive frames. Thus the search for θ_{MCP} can be limited to just a few degrees either side of its previous value, θ_{MCP}^{-1} (±4° was found to be optimal; higher values negated the performance advantages without significantly improving results, and lower values lacked precision). This reduced the likelihood of anomalous data corrupting the intended values, whilst also significantly lowering the required processing time for each frame, reducing the algorithm's complexity to (where p is the degree of precision calculated by the algorithm):

$$\mathrm{O}\left(\frac{8}{p^2}\right) \quad (5)$$

It was satisfactory to constrain:

$$\theta_{DIP} = \frac{2}{3}\theta_{PIP} \quad (6)$$

which is an empirical relationship established by Byron, *et al*. [29] and to enhance model output [24], [30]. Whilst this reduced the degrees of freedom (DOF) of the finger's movement by one, it modeled the relationship between the PIP and DIP, and further reduced the program's complexity to:

$$O\left(\frac{8}{p}\right) \quad (7)$$

This allowed for improved real-time performance. These enhancements are shown in (8):

$$\theta_{MCP} \in \left\{\theta_{MCP}^{-1} - \frac{\pi}{45} \to \theta_{MCP}^{-1} + \frac{\pi}{45}\right\},\ \theta_{PIP} \in \left\{0 \to \frac{\pi}{2}\right\}$$

$$j_1\binom{x}{y}_{\theta_{MCP}} = \begin{pmatrix} \frac{B_x + B_x^{-1}}{2} + l_{MCP} * cos\,\theta_{MCP} \\ \frac{B_y + B_y^{-1}}{2} + l_{MCP} * sin\,\theta_{MCP} \end{pmatrix}$$

$$j_2\binom{x}{y}_{\theta_{MCP},\theta_{PIP}} = \begin{pmatrix} j_1(x)_{\theta_{MCP}} + l_{PIP} * cos(\theta_{MCP} - \theta_{PIP}) \\ j_1(y)_{\theta_{MCP}} + l_{PIP} * sin(\theta_{MCP} - \theta_{PIP}) \end{pmatrix}$$

$$D\binom{x}{y}_{\theta_{MCP},\theta_{PIP}} = \begin{pmatrix} j_2(x)_{\theta_{MCP},\theta_{PIP}} + l_{DIP} * cos(\theta_{MCP} - \frac{5}{3}\theta_{PIP}) \\ j_2(y)_{\theta_{MCP},\theta_{PIP}} + l_{DIP} * sin(\theta_{MCP} - \frac{5}{3}\theta_{PIP}) \end{pmatrix}$$

$$\min\left\{\sqrt{\left[D(x)_{\theta_{MCP},\theta_{PIP}} - D_k(x)\right]^2 + \left[D(y)_{\theta_{MCP},\theta_{PIP}} - D_k(y)\right]^2}\right\} \quad (8)$$

Ground truth estimation was used to validate the accuracy of Landmark Definitions; however, further validation of the system was required to assess the validity of the Joint Angle Algorithm. The output of the system was therefore compared directly with a laboratory-based motion capture system as a gold-standard.

Twenty-six 3mm hemisphere retro-reflective markers were placed on the dorsal aspect of the right wrist, hand, fingers and thumb of a single participant in accordance with the Hand & Wrist Kinematics (HAWK) protocol; a previously published kinematic measurement technique that has been validated and tested for reliability [9], [10]. HAWK has been shown to generate joint angles of the wrist, hand, fingers and thumbs accurately to <1° [9], [10].

The participant was asked to sit at a table with their hand positioned above the surface as illustrated in Fig. 1 holding their right hand in a pincer grip. They were then asked to open and close their hand to mimic grasping a cup. Finger joint angles (MCP, PIP & DIP) were generated whilst the participant moved their hand from an open position (Open1), to a mid-point (Mid1), to a closed-grip (Closed), then back to the mid-point (Mid2) and finally back to the open position (Open2). This process was then repeated.

The Microsoft Kinect™ was used alongside a Vicon T-Series infrared motion capture system sampling at 100Hz. The Vicon system consists of 12 cameras (6 T40 and 6 T160 cameras). These data were then directly compared to assess the accuracy of the proposed system to that of the HAWK kinematic measurement technique.

IV. Results

A two-phase validation process of the grip classification algorithms was undertaken. For spread hand mode, the method of ground truth estimation was used, but due to potential errors inherent in human-based validation, further validation was undertaken comparing the algorithm output against a laboratory-based motion capture system.

A. Ground Truth Validation Testing

Three reviewers assessed 42 videos in total for Spread Hand mode. In total 788 frames were successfully analyzed, and 221 were rejected due to classification not being able to be performed, or a suitable contour not found by the algorithm. Four reviewers assessed 33 videos in total for Pincer Mode. In total 904 frames were successfully analyzed, and 146 were rejected. The system obtained a result 78% of the time. When the system performed a classification, the points were classified correctly compared to the reviewer classification. The average variance for this approach was 2.34 pixels for spread hand mode and 9.47 in pincer mode.

Pearson's correlation was calculated between the points of interest generated by the algorithm and the reviewer's classification in order to determine whether there was a linear relationship, which would be used as an indicator of output reliability. The correlation was investigated separately between the x and y co-ordinates. The results are shown in Table I.

Due to errors inherent in the recording process (clicking on a screen), there may be errors within this process of validation, which was assessed using an ANOVA.

TABLE I
UNITS RESULTS OF THE PEARSON'S PRODUCT MOMENT CORRELATION FOR X AND Y CO-ORDINATES BETWEEN DIFFERENT REVIEWERS AND ALGORITHMICALLY GENERATED FEATURES

Reviewer	Correlation between X values	Correlation between Y values
Reviewer 1	R = 0.996948	R = 0.990262
Reviewer 2	R = 0.997922	R = 0.996943
Reviewer 3	R = 0.996778	R = 0.989879

A single factor ANOVA was then calculated from these data. The results show that F is significant at $p < 0.0036$, indicating a significant relationship between the validation errors.

B. Joint Angle Algorithm

The results from testing the joint angle algorithm shows that precision decreases with a higher required prediction value, i.e. the algorithm can predict within 6° with 78.2% certainty, within 8° with 92% certainty and within 10° with 97% certainty (see Fig. 5).

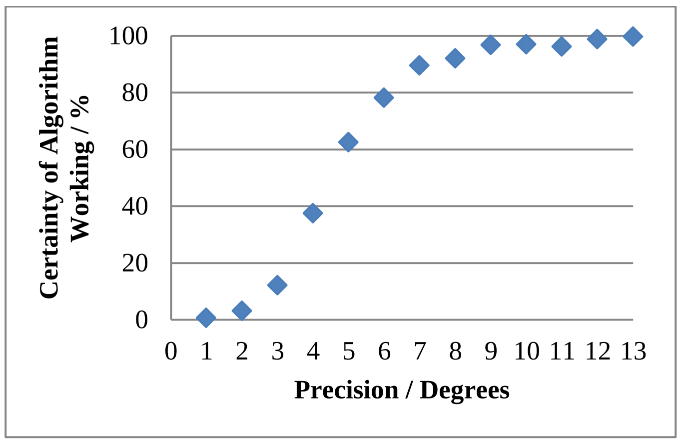

Fig. 5. Graph showing how the certainty of the algorithm working correctly varies with the degree of precision required.

The joint angle algorithm was therefore set to find the angles correct to the nearest 8°, due to the high probability of getting correct results (92% according to Fig. 5.) and to ensure real-time performance. The complexity of the algorithm is inversely proportional to the cube of the degrees of precision, p, due to the triplet of nested loops that generate the three angles:

$$O\left(\frac{1}{p^3}\right) \tag{9}$$

Thus, the complexity becomes too large to run in real-time if p is too low. In practice, $p \geq 4°$ was used to ensure the program runs smoothly.

The results of the comparative analysis between the proposed system and the laboratory-based motion capture system are presented in Table II. Fig. 6 to Fig. 8 illustrates the full dynamic movement for θ_{MCP}, θ_{PIP} and θ_{DIP}, respectively. They show direct comparative data between the proposed system and motion capture system including HAWK.

TABLE II
MCP ANGLES MEASURED BY MOTION CAPTURE + HAWK (LAB) AND KINECT SYSTEMS DURING A DYNAMIC HAND MOVEMENT

	θ_{MCP} / °			
	TEST 1		TEST 2	
	LAB	KINECT	LAB	KINECT
OPEN 1	43	40	25	15
MID1	20	22	20	16
CLOSED	7	7	5	3
MID2	22	28	23	29
OPEN2	41	42	41	39
	θ_{PIP} / °			
	TEST 1		TEST 2	
	LAB	KINECT	LAB	KINECT
OPEN 1	36	24	36	28
MID1	25	23	26	31
CLOSED	2	8	4	8
MID2	24	22	26	22
OPEN2	38	28	41	45
	θ_{DIP} / °			
	TEST 1		TEST 2	
	LAB	KINECT	LAB	KINECT
OPEN 1	27	18	29	18
MID1	17	16	16	21
CLOSED	0	6	3	4
MID2	15	14	11	15
OPEN2	27	22	31	30

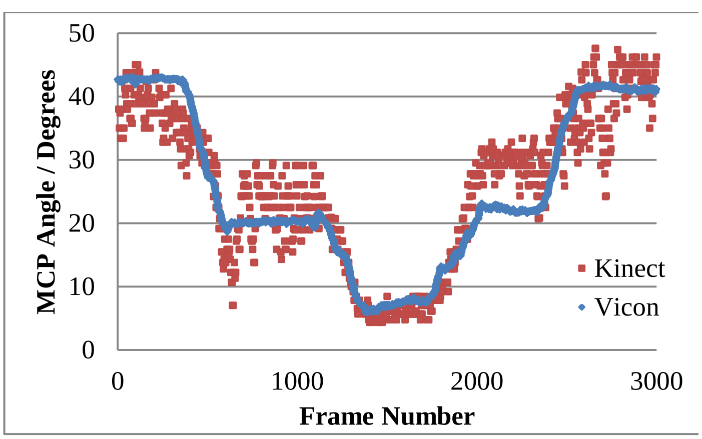

Fig. 6. Comparative data between the Microsoft Kinect™ and the Motion Capture system + HAWK for measuring θ_{MCP}.

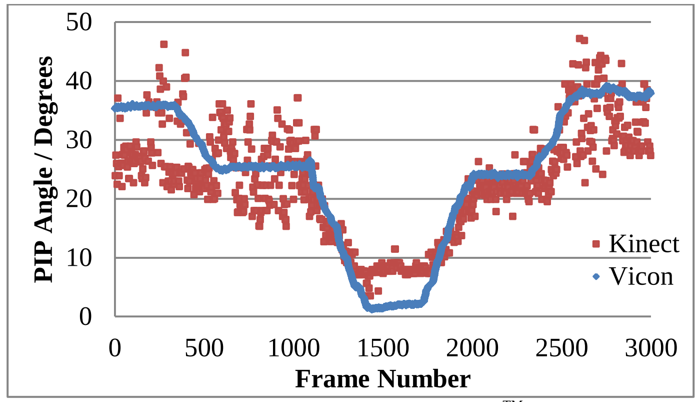

Fig. 7. Comparative data between the Microsoft Kinect™ and the Motion Capture system + HAWK for measuring θ_{PIP}.

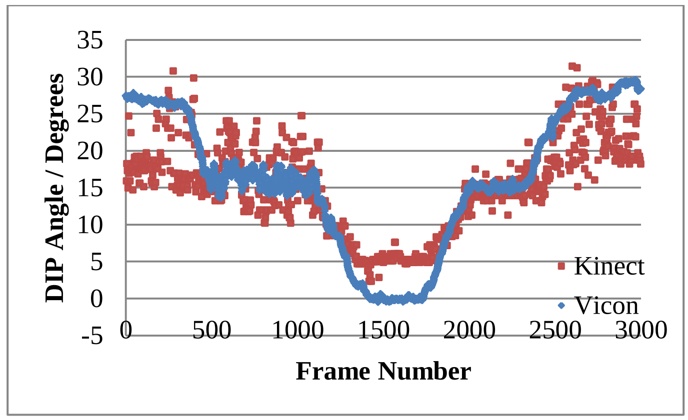

Fig. 8. Comparative data between the Microsoft Kinect™ and the Motion Capture system + HAWK for measuring θ_{DIP}.

V. DISCUSSION

A need to develop a home-based rehabilitation tool to capture hand and finger movement has been identified. To provide such an innovative step-change in home-based healthcare, this new technology must be able to robustly track and accurately measure the movements of the hand. This paper has presented a novel system using the Microsoft Kinect™ to identify anatomical (hand) landmarks, develop classification algorithms for different grip modes and the definition of a joint angle algorithm for remote measurement. This information was validated using a two-phase process; ground truth estimation against reviewers, and a comparative analysis against the motion capture industry's gold standard system.

In home-based rehabilitation, some remote technologies do include measurement of the hand and fingers, often using glove-based technology [7], [31]. Gianstanti [32], for example, adopted a glove-based approach for tracking and graphically rendering hand gestures. However, some Authors advocate moving away from traditional instrumented gloves as a home-based rehabilitation solution due to the difficulty in donning and doffing [12], [33]. Alternative solutions include miniature inertial measurement units (IMUs) attached to each phalanx of the fingers and thumbs [7]. However the practical application of such a system requires a series of wired nodes attached to a central power and processing source, which is cumbersome to wear and therefore unrealistic as a home-based solution.

The proposed system, once realized, can be deployed in either a home-based or clinical setting. In terms of accuracy, the proposed system must be comparable, if not competitive to existing systems and methods of measurement. In clinical practice, the most common measurement technique to assess joint range of movement is manual goniometry. These traditional measurements are known to have a large component of human error, averaging 7-9° for each joint movement measured [34]. Furthermore, these manual systems can only obtain measurements for a static position and do not generate continuous data as the user moves their hand. Thus the proposed system would give a greater quantity and quality of information than previously available using these traditional methods.

The system was validated using two methods; ground truth estimation and a direct comparison with a laboratory-based motion capture system. The process of ground truth estimation was limited and the Authors advocate a large sample size be used to avoid individual variability.

The inclusion of the joint angle algorithm is highly novel in markerless motion capture of the hand, and particularly within the context of utilizing the Microsoft Kinect™. In the current phase, the Microsoft Kinect™ was used in default mode. It is arguable that the proposed system is limited by the resolution of the images received. Therefore further work using the Microsoft Kinect™ in near mode may reduce errors and achieve higher levels of accuracy. The results of the comparison between the joint angle algorithm and the laboratory-based system showed a good to high correlation in θ_{MCP}, θ_{PIP} and θ_{DIP} across the entire dynamic movement. The maximum error for θ_{MCP} is ±10°, although the average absolute deviation is much smaller at just 2.4°. The maximum error for θ_{PIP} is ±12° and the average absolute error is 4.8°. These errors are slightly larger since θ_{PIP} is measured relative to the plane created by θ_{MCP}. Therefore any errors in θ_{MCP}will be compounded in the result of θ_{PIP}, in addition to any intrinsic errors in the measurement of this angle. The maximum error for θ_{DIP} is ±11° and the average absolute error is 4.8°. These are very similar to the results for θ_{PIP} due to the linear relationship between the angles used to calculate θ_{DIP}. These results are comparable to results of manual goniometry, where 2-18.9° was found when measuring hip motion [35], 6-7° when measuring cadaveric wrists [36] and 7-9° for measuring finger position [37]. The proposed system is therefore comparable to manual goniometry and an improvement due to the disparity in reliability of the traditional technique, and the proposed system being capable of working in real-time. In addition, the proposed system is markerless, unlike other motion capture solutions aimed at home-based rehabilitation [38].

Direct comparisons between the Microsoft Kinect™ and laboratory-based motion capture has also been undertaken by Dutta [39], who found root mean squared errors (SD) of 0.0065m (0.0048m), 0.0109m (0.0059m) and 0.0057m (0.0042m) in the x, y and z directions respectively of an absolute position. In this study, Dutta was measuring the position of a series of static points within a laboratory. Dutta found the largest errors farthest away >3.0m from the sensor. This implies that close data capture, like those proposed by the Authors for hand rehabilitation would be less likely to incur these identified errors. This assumption would need validation and the release of a ‘near mode’ in the Microsoft SDK would also help facilitate this aim. In other work, Alnowami, *et al*., [40] used the Microsoft Kinect™ to capture absolute position on the thorax for measuring respiratory inhalation/exhalation. During validation work, Alnowami, *et al*., found the Microsoft Kinect™ to have millimeter precision at depths 0.8-1.5m.

An obvious benefit of the proposed system is its ability to run in real-time; making it a viable solution as a biofeedback device in the home. Whilst some laboratory-based motion capture systems, such as the Vicon system can find and classify POIs in real-time, the additional process of producing model outputs, and, hence joint angles, currently requires post-processing.

In a recent study of 24 markerless hand motion capture systems only two were found to work in real-time [19]. However, of these, one system performs at 10 FPS [20] and one requires a cluster of six PCs to run at 30 FPS. By contrast, the proposed system reaches 30 FPS on a single PC and is limited by the processing speed of the Microsoft Kinect™ (30 FPS). The marker-based system by Aristidou & Lasenby [21] also achieves real-time performance of 76 FPS, but requires 10 separate cameras and presents no data to ascertain the level of accuracy achieved. El-Sawah, *et al*. [22] propose a real-time system with a reported accuracy of ±1°, but the system proposed is purely theoretical with no indication that it has been implemented. Therefore, the system proposed by the Authors is highly novel being the only real-time system suitable for use by a patient with standard desktop equipment commonly available in the home.

The tracking component of the proposed system is not novel. Many gesture-tracking systems have been proposed using fingertip tracking for Natural User Interfaces [41], [42]. The landmark definition and consequently, the grip classification defined in the proposed system, and validated through ground truth estimation, may, however, improve gesture recognition systems further by utilizing less constrained and natural hand movements. Like other clinical systems [43]-[45], the proposed system can only classify a limited number of gestures which is suitable for a specific application where there are a finite number of rehabilitation gestures. However, the proposed system has been shown to detect gestures following transitions, and is the first system to apply a kinematic technique to measure finger joint movement.

Previous research has [25] adopted gesture recognition and reported a total measurement of 22 DOF for the whole hand, with three DOF for each finger. However, the Authors assert that $\theta_{DIP} = \frac{2}{3}\theta_{PIP}$ and that $\theta_{PIP} = \theta_{MCP}$. Therefore, in reality, each finger had only one DOF. In addition, real-time performance was achieved at 10.2 FPS using a cluster of 9 dual-core CPUs (2.16GHz), which is not practical for home use. The results presented suggest a maximum error in θ_{MCP} of ±6°, however these are referenced only to a ground truth. This would include a large intrinsic error, which was not reported by the Authors, so it was not possible to ascertain the credibility of these results. This system is the closest to the proposed system in terms of features and performance.

The results presented here were calculated in real-time, with a markerless solution using one Microsoft Kinect™ camera. There is currently no other system available that combine these features, although several exist that can achieve part of this solution. As previously discussed, laboratory-based motion capture systems can achieve higher accuracy but are impractical for use in a home-based rehabilitation environment. An alternative to this would be to use a marker-based approach with the Microsoft Kinect™. Cordella, *et al*., [38] adopts this approach using a comprehensive marker set. This solution however poses problems analogous to the use of gloves and inertial measurement systems: the accuracy of the system relies on the patient being able to accurately position and wear the device, in this case a marker set reliant on palpation for accurate placement. This is not a suitable solution for rehabilitation or remote monitoring in a home environment.

VI. Conclusion

Ubiquitous and unobtrusive systems are paramount to patient adherence in the next generation of home-based, remote healthcare. The development and preliminary validation of a real-time markerless system has been presented, which can measure finger movement using the Microsoft Kinect™. This is the first system that goes beyond gesture recognition; defining the basis for a home-based rehabilitation platform.

References

[1] United States of America Government, "Patient protection and affordable care act," *Office for the Legislative Council.* 1 May. 2010.

[2] United Kingdom of Great Britain Government, "The power of information: putting all of us in control of the health and care information we need," *Dept. of Health.* Gateway Ref: 17668. 21 May. 2012.

[3] European Commission, "TeleHealth," *Straeqo Cornsortium.* 11 December. 2007.

[4] F. L. Buczek, E. W. Sinsel, D. S. Gloekler, B. M. Wimer, C. M. Warren, J. Z. Wu, "Kinematic performance of a six degree-of-freedom hand model (6DHand) for use in occupational biomechanics," *J. Biomech.*, vol. *44*, no. 9, pp. *1805-1809*, June, 2011.

[5] J. Condell, G. Moore, "HandPuppet3D: motion capture and analysis for character animation," *Artifical Intel. Rev.* vol. *31*, no. 1-4, pp. 45-59, 2009.

[6] M. I. Jackson, M. J. Hiley, M. R. Yeadon, "A comparison of Coulomb and pseudo-Coulomb friction implementations: application to the table

contact phase of gymnastics vaulting," *J. Biomech.*, vol. *44*, no. *15*, pp. 2706-2711Oct., 2011.

[7] H. G. Kortier, H. M. Schepers, V. I. Sluiter, P. H. Veltink, "Ambulatory assessment of hand kinematics: using an instrumented glove," *XII Int. Symp. 3D Analysis Human Movement*, 18-29 July, Bologna, Italy. 2012.

[8] D. Kim, O. Hilliges, S. Izadi, A. D. Butler, J. Chen, I. Oikonomidis, P. Oliver, "Digits: freehand 3D interactions anywhere using a wrist-worn gloveless sensor," *Proc. 25th Annu. ACM Symp. On User Interface Software Tech.*, Cambridge, MA, USA. 2012.

[9] C. D. Metcalf, S. V. Notley, P. H. Chappell, J. H. Burridge and V. T. Yule, "Validation and application of a computational model for wrist and hand movements using surface markers," *IEEE Trans. Biomed. Eng.*, vol. 55, no. 3, pp. 1199-1210, 2008.

[10] C. D. Metcalf, S. V. Notley, "Modified kinematic technique for measuring pathological hyperextension and hypermobility of the interphalangeal joints," *IEEE Trans. Biomed. Eng.*, vol. 58, no. 5, pp. 1224-1231, 2011.

[11] Wii-Habilitation, "For those involved in the use of the Nintendo Wii as part of therapy or rehabilitation," [Online]. Available Oct. 2012: http://wiihabilitation.co.uk/main.shtml.

[12] S. Demain, C. D. Metcalf, G. V. Merrett, D. Zheng, S. Cunningham, "A narrative review on haptic devices: relating the physiology and psychophysical properties of the hand to devices for rehabilitation in central nervous system disorders," *Dis. Rehab.: Assist. Tech.*, [In Press].

[13] M. Windolf, N. Götzen, M. Morlock, "Systematic accuracy and precision analysis of video motion capturing systems – exemplified on the Vicon 460 system", *J. Biomech.*, vol. 41, no. 12, pp. 2776-2780. 2008.

[14] Y. Ehara, "Comparison meeting of motion analysis systems," *Clin. Gait Anal. Forum Japan*, [Online]. Available Oct. 2012: http://www.ne.jp/asahi/gait/analysis/comparison2002/index-eng.html.

[15] I. Carpinella, J. Jonsdottir, M. Ferrarin, "Multi-finger coordination in healthy subjects and stroke patients: a mathematical modelling approach," *J. Neuroeng. Rehab.*, vol. 8, no. 19., 2011. [Online], Available Oct. 2012: http://www.jneuroengrehab.com/content/pdf/1743-0003-8-19.pdf.

[16] R. DeGeorges, J. Parasie, D. Mitton, N. Imbert, J. N. Goubier, F. Lavaste, "Three-dimensional rotations of human three-joint fingers: an optoelectronic measurement. Preliminary results," *Surg. Radiol. Anat.*, vol. 27, pp. 43-50. 2005.

[17] S. Corazza, L. Mundermann, A. M. Chaudhari, T. Demattio, C. Cobelli, T. P. Andriacchi, "A markerless motion capture system to study musculoskeletal biomechanics: visual hull and simulated annealing approach," *Ann. Biomed. Eng.*, vol. 34, no. 6, pp. 1019-1029, June 2006.

[18] H. Rijpkema and M. Girard, "Computer animation of knowledge-based human grasping," *18th Ann.Conf. Comp.Graphics Interact.Tech.*, Groningen, The Netherlands, 1991.

[19] A. Erol, G. Bebis, M. Nicolescu, R. D. Boyle, X. Twombly, "A review on vision-based full DOF hand motion estimation," *IEEE Comp. Soc. Conf. Comp.Vis. Pattern Recog.*, University of Nevada, USA, 2005.

[20] J. M. Rehg, T. Kanade, "DigitEyes: vision-based hand tracking for human-computer interaction," *IEEE Workshop Motion Non-Rigid Articulated Obj.*, Nov, 1994.

[21] A. Aristidou, J. Lasenby, "Motion capture with constrained inverse kinematics for real-time hand tracking," *Proc. 4th Int. Symp. Comm., Cont. Sig. Process.*, Limmasol, Cyprus, 2010.

[22] A. El-Sawah, N. D. Georganas, E. M. Petriu, "Finger inverse kinematics using error model analysis for gesture enabled navigation in virtual environments," *IEEE Int. Workshop Haptic Audio Vis. Envir. App.*, Ottawa, Canada, 2006.

[23] B. Stenger, P. Mendonca, R. Cipolla, "Model-based 3D tracking of an articulated hand," *Proc. IEEE Comp. Soc. Conf. Comp. Vis. Pattern Recog.*, Cambridge, UK, 2001.

[24] J. Lee, T. L. Kunii, "Model-based analysis of hand posture," *IEEE Comp. Graph. App.*, vol. 15, no. 5, pp. 77-86, September, 1995.

[25] T. Gumpp, P. Azad, K. Welke, E. Oztop, R. Dillmann, G. Cheng, "Unconstrained real-time markerless hand tracking for humanoid interaction," *6th IEEE-RAS Int. Conf. Humanoid Robots*, Saitama, Japan, 2006.

[26] A. M. Biancardi, A. C. Jirapatnakul, A. P. Reeves, "A comparison of ground truth estimation methods," *Int. J. Comp. Assist. Radiol. Surg.*, vol. 5, no. 3, May 2010.

[27] C. Sollerman, "The use of eight main hand-grips in activities of daily living," Ph.D. dissertation, Univ. of Göteborg, Sweden, 1980.

[28] S. Malik, "Real-time hand tracking and finger tracking for interaction," M.Sc. dissertation, Comp. Sci. Univ. Toronto, Canada, 2003.

[29] K. Byron, L. Pham, J. Situka, "Markerless motion tracking," M.Sc. dissertation. Iowa State University, USA, 2009.

[30] J. L. Crowley, F. Berard, J. Coutaz, "Finger tracking as an input device for augmented reality," *Int. Workshop Face Gesture Recog.*, Zurich, Switzerland, 1995.

[31] Z. Luo,C. K. Lim, I-M. Chen, S. H. Yeo, "A virtual reality system for arm and hand rehabilitation", *Front. Mech. Eng.* Vol. 6, no. 1, pp. 23-32, 2011.

[32] D. Gianstanti, "Validation of a tele-home-care for hand-telerehabilitation," *Proc. IEEE Eng. Med. Biol. Soc.*, pp. 3830-3832, 2007.

[33] M. Huber, B. Rabin, C. Docan, G. C. Burdea, M. Abdelbaky, M. R. Gomolb, "Feasibility of modified remotely monitored in-home gaming technology for improving hand function in adolescents with cerebral palsy", *IEEE Trans. Info. Tech. Biomed.*, vol. 14, no. 2, pp. 526-534, 2010.

[34] L. Ballan, G. M. Cortelazzo, "Marker-less motion capture of skinned models in a four camera set-up using optical flow and silhouettes," *Proc. 4th Int. Symp. 3D Data Process., Vis. Transmission*, Georgia Inst. of Tech., Atlanta, USA, 2008.

[35] S. Nussbaumer, M. Leunig, J. F. Glatthorn, S. Stauffacher, H. Gerber, N. A. Maffiuletti, "Validity and test-retest reliability of manual goniometers for measuring passive hip range of motion in femoroacetabular impingement patients," *BMC Musc. Disord.*, vol. 11, pp. 194, 2010.

[36] T. I. Carter, B. Pansy, A. L. Wolff, H. J. Hillstrom, S. I. Backus, M. Lenhoff, S. W. Wolfe, "Accuracy and reliability of three different techniques for manual goniometry for wrist motion: a cadaveric study," *J. Hand. Surg.*, vol. 34, no. 8, pp. 1422-1428, 2009.

[37] B. Ellis, A. Bruton, "A study to compare the reliability of composite finger flexion with goniometry for measurement of range of motion in the hand," *Clin. Rehab.*, vol. 16, no. 5, pp. 562-570, 2002.

[38] F. Cordella, F. Di Corato, L. Zollo, B. Siciliano, P. van der Smagt, "Patient performance evaluation using Kinect and Monte Carlo-based finger tracking," *4th IEEE RAS/EMBS Int. Conf. Biomed. Rob. Biomechatronics*, Rome, Italy, June, 2012.

[39] T. Dutta, "Evaluation of the Kinect™ sensor for 3-D kinematic measurement in the workplace," *App. Ergo.*, vol. 43, pp. 645-649.

[40] M. Alnowami, B. Alnwaimi, F. Tahavori, M. Copland, K. Wells, "A quantitative assessment of using the Kinect for Xbox360™ for respiratory surface motion tracking," *Proc. SPIE*, vol. 8316, 2012.

[41] J. A. K. S. Segen, "Gesture VR: vision based 3D hand interface for spatial interaction," *Proc. ACM Multimedia*, pp. 455-464, 1998.

[42] S. S. Rautaray, A. Agrawal, "Design of gesture recognition system for dynamic user interface," *IEEE Int. Conf. Tech. Enhanced Educ.*, Allahabad, India, Jan. 2012.

[43] B. You, P. Siy, W. A. T. S. Anderst, "In vivo measurement of 3-D skeletal kinematics from sequences of biplane radiographs: Application to knee kinematics," *IEEE Trans. Med. Imag.*, vol. 20, pp. 514-525, 2001.

[44] D. Massimini, J. A. L. G. Warner, "Non-invasive determination of coupled motion of the scapula and humerus: An in-vitro validation," *J. Biomech.*, vol. 44, pp. 408-412, 2011.

[45] Y. A. N. M. Jang-Hee, "Automated Markrless Analysis of Human Gait Motion for Recognition and Classification," *ETRI J.*, vol. 33, pp. 259-266, 2011.

Cheryl D. Metcalf (S'04–M'08) received the B.S. degree (Hons.) in computer studies from the University of Sunderland, Sunderland, U.K., in 2002, specializing in adaptive control strategies of robotic grippers, the M.S. degree in evolutionary and adaptive systems from the University of Sussex, Brighton, U.K., in 2003, specializing in evolving optimized object recognition using haptics, and the Ph.D. degree from the University of Southampton, Southampton, U.K., in 2008, specializing in kinematic analysis of wrist and hand movements and function. She completed a prestigious Research Council's (U.K.) Roberts Fellowship in Life Sciences Interface at the University of Southampton, Southampton, U.K (08-11) and is currently a Lecturer in Biomechanics in the Faculty of Health Sciences, University of

Southampton, U.K. Her research interests include the hand and wrist biomechanics in unimpaired, neurological and musculoskeletal impairment, kinematic assessment of complex skill acquisition, developing and assessing novel biomedical sensors for telemonitoring and biomechanical assessment of assistive devices.

Rebecca Robinson received the First Class M.Eng. degree (Hons.) in Electronic Engineering from the University of Southampton, U.K., in 2012. She is currently a Graduate Engineer at Roke Manor Research and has previously worked as an intern at Ericsson Televisions.

Adam Malpass received the First Class M.Eng. degree (Hons.) in Electronic Engineering from the University of Southampton, UK, in 2012, during which time he was awarded the IET Jubilee Scholar. He has previously worked internships at Drumgrange Ltd, Surrey, UK (2006-9); Ernst and Young LLP, London, UK (2010); Analog Devices Ltd, Newbury, UK (2010); and Dialog Semiconductor KK, Tokyo, Japan (2011). He is currently a Graduate Engineer at Dialog Semiconductor in Swindon, UK.

Tristan Bogle received the M.Eng. degree (Hons.) in Electronic Engineering with Artificial Intelligence from the University of Southampton, UK, in 2012. He is currently working as a Graduate Electronic Engineer at WheelRight, Oxford, UK.

Thomas Dell received the M.Eng. degree (Hons.) in Electronic Engineering from the University of Southampton, UK, in 2012. He has worked as an intern at: Analog Devices Ltd, Newbury, UK (2009); Roke Manor Research Ltd, Romsey, UK (2010); and Aptina Imaging Ltd, Bracknell, UK (2011). He is currently a Software Engineer at McLaren Electronic Systems Ltd, Woking, UK.

Chris Harris received the Ph.D. degree in theoretical physics (quantum field theory) at the University of Kent, UK in 1978. He has worked for 30 years at Roke Manor Research Ltd. on image processing and computer vision. An early success was the invention of the Harris corner detector, facilitating the analysis of geometric structure of computer vision systems based on structure-from-motion algorithms. He did initial development of the algorithms for tracking cricket and tennis balls, now used worldwide in the HawkEye system. His current work concerns the use of computer vision and optical sensors for the guidance of autonomous robots and measurement systems.

Sara Demain received the Graduate Diploma of Physiotherapy from the Chartered Society of Physiotherapy, London, U.K., in 1990; the M.S. degree, from the University of Southampton, U.K. in 2001, specializing in Rehabilitation and Research, and the Ph.D. from the University of Southampton, U.K. in 2009, in physiotherapy following Stroke. She is a Lecturer in Physiotherapy in the Faculty of Health Sciences at the University of Southampton, U.K., and is currently undertaking a prestigious National Institute of Health Research (UK) Post-Doctoral Fellowship in understanding and measuring the burden of self-care rehabilitation post-stroke.