

Stroke Patient Daily Activity Observation System

Jaired Collins*, Joseph Warren†, Mengxuan Ma‡, Rachel Proffitt§ and Marjorie Skubic¶

*†Computer Information Systems, Missouri Southern State University, Joplin, MO.

‡¶Electrical Engineering and Computer Science, §Occupational Therapy, University of Missouri, Columbia, MO.

*jairedcollins@mail.missouri.edu, †jwpsy@mail.missouri.edu, ‡mmrnc@mail.missouri.edu,

§proffittm@health.missouri.edu, ¶skubicm@missouri.edu

Abstract—Stroke is a leading cause of long-term adult disability. Stroke patients can recover through rehabilitation programs prescribed by occupational therapists (OT); however, an individualized rehabilitation program can reduce recovery times compared to traditional ones. In this paper, we propose a daily activity observation system (DAOS) that uses a Kinect v2 sensor to collect and retrieve motion data. The DAOS has a robust interface to extract depth and skeleton data, and supports data collection in an unstructured kitchen environment. Depth data are used to perform action recognition and track problematic movements, while skeleton data are used to calculate mean velocities of hand joints, max extensions, symmetry of hand movements, and other assessment metrics for therapists. Histogram of oriented 4D normals is used for action recognition. The action recognition accuracy is 97% on a multi-class kitchen action dataset. Through action recognition and accurate assessment, we present a novel system that can assist therapists and their ability to provide quality care to stroke patients.

Index Terms—KinectV2, Stroke Rehabilitation, Machine Vision, HON4D, Activity Recognition, Assessment, Daily Observation System

I. INTRODUCTION

Nearly 80,000 people each year suffer from a stroke in the U.S. [1]. Moreover, about 50% report hemiparesis which affects their ability to live independently [1]. Those who experience a stroke can recover through rehabilitation managed by a therapist whose goal is to build up the strength and capabilities of muscles that are lacking control or not functioning correctly. Traditional rehabilitation involves patients performing exercises in a hospital or clinic monitored by a therapist [2]. However, the adherence to rehabilitation programs is low [1]. Recently, with the help of new technologies, game-based physical activities have been developed for stroke rehabilitation [2]. Proffitt et al. proposed Mystic Isle, a virtual reality video game that motivates a patient through the game to target balance training and upper limb reaching [3] using an inexpensive depth sensor, the Microsoft Kinect [4]. The game can be set up in each patient's home.

To make rehabilitation treatment effective, it is essential for therapists to personalize and refine rehabilitation plans. This requires monitoring patient health status and recovery progress. To collect such information, the therapist can either observe the patient in the clinic or track the patient's motions through the game history. However, constant visits to the clinic are not convenient for most, and the in-game history usually is too short to provide enough information. Neither provides assessment on how the rehabilitation translates to everyday

activities. To accomplish this, a system that can provide daily activity assessment in the home is needed. To our knowledge, there is currently no such system for tracking rehabilitation progress as it relates to everyday movements at home, for example, tracking range of motion of actions in the kitchen.

The assessment feedback from the daily activity motions in the home poses several benefits for stroke rehabilitation. First, the daily activity assessment results provide more details of the strengths and weaknesses of a patient, which helps therapists to design a more personalized rehabilitation plan. Second, therapists can adjust the rehabilitation plan efficiently based on the change of daily activity assessment tracked over time, to enhance the quality and speed of recovery.

To achieve this goal, we propose a Kinect-based daily activity observation system for stroke patients. It classifies the action of a patient using a histogram of oriented 4D normal (HON4D) descriptor [5] of a sequence of depth images. The coordinates of joints in 3D space collected from the sensor are used for assessment metrics that are specifically tailored to stroke patients. Hand mean and maximum velocities, symmetry of hand movement, maximum hand extension, and chest sway were extracted. With this ability, a more personalized and qualitative rehabilitation plan can be made.

This paper is organized as follows: Section II presents related work. Section III describes our methodology including the daily observation system, an in-house dataset, action recognition, and assessment. Section IV shows the results of action recognition, while Section V presents the results of assessment. Finally, Section VI concludes the work.

II. RELATED WORK

A. Action Recognition

Action recognition with depth images generally follows feature extraction, quantization/dimension reduction, and classification [6]. Spacetime algorithms extract local or global features of a depth image sequence, without explicit regard for time between images [6][7][8]. Sequential algorithms extract local features for every depth image and uses a generative statistical model [6]. Since we selected a global spacetime algorithm, these methods are covered in this section.

One of the first attempts at accurate classification with depth data was bag of 3D points. It uses a sparse sampling technique to gather 80 points at key locations to be projected onto three Cartesian planes. Collected points were encoded by building an action graph [9], and then using a Gaussian Mixture Model

to obtain point distribution. Their bag of 3D points method, when applied to their own dataset, achieved around 90% initial accuracy and 76% cross subject accuracy [6].

Later work involved mostly histograms of oriented gradients or normals, with an SVM as the classification algorithm. [7] uses a histogram of oriented gradients (HOG) generated from three separate Depth Motion Maps (DMM), called DMM-HOG. DMMs were generated by summing the difference between every image in a sequence with a threshold. After three DMMs were generated—front, side, and top—a HOG is computed. In order to perform action recognition, the authors use the DMM-HOG descriptors to train an SVM. The authors also found that 35 frames were sufficient to perform their algorithm; however, that count depends on action complexity.

More recently, Oreifej and Liu [5] offer a new descriptor for a depth image sequence. Normal vectors are calculated using x , y , z , and time over an entire depth image sequence. A 120 vertices, 600 cell polychoron is used as the bins of a histogram. Performing a dot product against the normals and every bin of the histogram, a histogram of oriented 4D normals is built. The method improves upon the state of the art with an accuracy of 92% [5]; it is re-implemented here in C#.

B. Quantitative Assessment

Quantitative assessment of human movements can objectively present the outcomes and the important changes in subjects' functional states [10]. The Kinect-based game in [11] assessed the random arm movements in 3D-space. In [12], an interactive remote assessment system was provided to evaluate the trunk sway. An upper-limb functional evaluation tool in [13] makes it possible for patients to assess reaching against gravity movements at home. However, all these systems can only assess some pre-designed motions. None of them are able to assess daily activity motions.

To the best of our knowledge, no study has been conducted to combine action recognition and assessment for daily activities, even though the assessment feedback from daily in-home motions is important for rehabilitation.

III. METHODOLOGY

The proposed system, Daily Activity Observation System (DAOS), uses depth and skeleton data from a Kinect v2. Depth data are made of pixels that contain the distance from the camera plane to the nearest object [14]. Skeleton data are a list of 3D Cartesian points, specifying where a joint is at a recorded time. To test our novel solution in a home setting where an optimal viewing angle is usually not possible, we collected a dataset in a kitchen environment. Through the DAOS, we employ the HON4D algorithm for action recognition [5]. For assessment, we generate metrics such as maximum arm extension, hand velocity, hand trajectories, and chest sway.

A. Daily Activity Observation System

Using the Kinect, DAOS records unprocessed depth data (Figure 1) as well as depth data with a patient segmented out

(Figure 2). Skeleton data are also collected. Each recording has an associated .csv file with x , y , and z coordinates for all the joints tracked by the Kinect. The units are in meters, with the z coordinate encoding depth.

The skeleton data are utilized for assessment of range of motion. With x , y , and z coordinates of all joints, we calculate metrics such as mean velocities, max extension, symmetry of hand movements, and chest sway.



Fig. 1. Unprocessed depth data of a subject cutting with a knife.

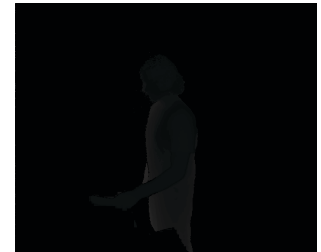


Fig. 2. Processed Figure 1 depth data.

All data collected, processed and unprocessed, are stored locally by the computer hosting the program. In predefined increments, a server will request the data, which are then transferred where they are stored for future use by a therapist.

An in-house dataset was collected to mimic the actions of a kitchen environment, where the counter and appliances were on one wall. To address a lack of forward facing data, our Kinect was placed at a side angle. As a result, our kitchen environment was highly susceptible to self-occlusion.

The dataset included 9 different subjects, 2 recordings of each action, and 28 different actions that correspond to general kitchen tasks. This amounts to 504 different image sequences. The 28 actions can be considered as 5 categories which are washing, meal preparation, kitchen gadget manipulation, general picking tasks, and walking. In the washing tasks, the participants were asked to put an item into the sink (WashSink), wash and rinse a dish (WashRinse) and place an item into the dishwasher (WashDishwasher). In the meal preparation tasks, the participants cut (PrepCut) and stirred (PrepStir) food. They were also asked to open (PrepOpen) and close (PrepClose) a container. In the manipulation tasks, participants performed actions including using the stove (ManipulateStove), microwave (ManipulateMicrowave) and refrigerator (ManipulateFridge), and turning on and off the sink faucet (ManipulateSinkOn/Off). In general picking tasks, the participants were asked to pick up and place an object onto a counter or a cabinet (PickUpCounter, PutDownCounter, PickUpTop, PutDownTop, PickUpBottom, PutDownBottom). They also opened and closed a cabinet (OpenBottomCabinet, CloseBottomCabinet, OpenTopCabinet, CloseTopCabinet). In the walking tasks, the participants walked in the kitchen either holding an object or without holding an object.

B. Descriptor for Action Recognition

It is important to have a rigorous global descriptor of a sequence of depth images so that actions that resemble

each other are distinguishable. Many kitchen actions have an opposite, and only differ in the temporal sense by reversing the movement, such as opening and closing various items. By including temporal information, a HON4D can be constructed [5]. We implemented a novel C# version that uses the first half of creating a HON4D descriptor, without adding projectors to the already-calculated histogram [5].

The first step in the methodology is to calculate normals for every pixel in a given set of depth images $\mathcal{I} = \{i_1, i_2, \dots, i_k\}$, as in Figure 3. The components of the normals are the changes in depth, which is summarized by

$$\mathbf{n} = \left(\frac{\partial z}{\partial x}, \frac{\partial z}{\partial y}, \frac{\partial z}{\partial t}, -1 \right). \quad (1)$$

```

1: procedure CALCULATENORMALS(images)
2:   for k = 0; k < images.Count - 1; k++ do
3:     img1 ← images[k]
4:     img2 ← images[k+1]
5:     for x = 0; x < img1.Width; x++ do
6:       for y = 0; y < img1.Height; y++ do
7:         currentPixel = img1.GetPixel(x, y)
8:         nextPixel = img2.GetPixel(x, y)
9:         rightPixel = img1.GetPixel(x + 1, y)
10:        leftPixel = img1.GetPixel(x - 1, y)
11:        upPixel = img1.GetPixel(x, y - 1)
12:        downPixel = img1.GetPixel(x, y + 1)
13:        x = rightPixel - leftPixel
14:        y = downPixel - upPixel
15:        z = currentPixel - nextPixel
16:        normalList.Add(x, y, z, -1)
17:       end for
18:     end for
19:   end for
20:   return normalList
21: end procedure

```

Fig. 3. Pseudocode to generate a list of oriented 4D normals for a given sequence of images.

Once all normals are computed, they are normalized since only the direction matters for bin contribution. A polychoron is then initialized uniformly in 4D space, where the vertices are treated as vectors and called projectors to be used as the bins of the histogram [5]. The contributions are calculated as the dot product of every normal and projector, as in Figure 4. After every dot product is calculated and added to the proper bin of the histogram, it is normalized.

To enhance the uniqueness of a HON4D descriptor, it is essential to subdivide the sequence of images into cells [5]. The Kinect sensor records depth information at a resolution of 512×424 . Our cells are $4 \times 4 \times 3$ ($w \times h \times d$). As normals are computed as in Figure 3, they are placed into their proper cells. Once all normals are placed, a separate HON4D is calculated for each cell. Afterwards, they are concatenated and normalized. This produces a 120 bin histogram.

```

1: procedure CREATEHON4D(proj, normList, hon4d)
2:   for k = 0; k < proj.Count; k++ do
3:     for n = 0; n < normList.Count; n++ do
4:       hon4d[k] += max(0, dotP(proj[k], norm-
5:         List[n]))
6:     end for
7:   end for
8:   return hon4d
9: end procedure

```

Fig. 4. Pseudocode to generate a histogram of oriented 4D normals, where proj is the list of projectors, normalList is the list of normals calculated from Figure 3, and hon4d is the histogram.

C. Assessment

The selected metrics in our novel solution were chosen to specifically track stroke patients. Each piece of information can be used to track improvement over time, or indicate a decline where intervention is needed. Stroke patients can suffer from weakness or paralysis, whether it is one side or the whole body [15]. Additionally, problems with balance or coordination can make walking difficult [15]. If a therapist can track improvements and declines, critical intervention points can be identified and care can be adjusted accordingly. For this reason, max extension of arms, speed and trajectories of hands, and chest sway are calculated for our DAOS.

Skeleton data are recorded and stored in a .csv file, with a 3D Cartesian point corresponding to each depth image. These points are utilized in several metrics. Occluded joints are set to be zeros, and are not used in assessment.

Max extension is calculated by first creating a list of distances using the distance formula. To find the max extension of the arm, we picked the wrist and shoulder joint. Searching for the maximum value yields the desired measurement.

Speed is calculated by finding the difference between each two successive distances in the list generated for max extension, and then dividing by the time elapsed between the two frames. After adding all the speed values to a list, the maximum or average speed can be found.

For trajectories, to center and decrease superfluous data, the middle spine joint is used as the origin. This is accomplished by subtracting the middle spine joint coordinates from any chosen joint's coordinates.

Chest sway is our final chosen metric. Points are plotted for the middle spine joint. The salient information is how far left and right the chest moves during an action, so points are plotted only for the X and Z plane, the plane parallel to the floor according to the Kinect's coordinates.

IV. ACTION RECOGNITION RESULTS

The HON4D descriptor is used to extract 3D position and time from a sequence of depth images for each action. An SVM with a quadratic kernel was then chosen [5] to classify actions using histograms. Since the DAOS was made to work in a challenging home environment, some data can be confused with one another. PrepCut, PrepOpen, and PrepClose look

similar in the feature space since they involve standing in the same spot with comparable hand movements. To this end, we tested our action recognition algorithm by subdividing the dataset described in Section III.A into smaller datasets.

TABLE I
DATASETS GENERATED

Dataset Name	Actions	SVM Accuracy
Prep1	All Prep Actions, Separated	56.3%
Prep2	PrepCut, PrepStir PrepOpen + Close	66.7%
Manipulate1	All Manipulate Actions, Separated	73.3%
Manipulate2	ManipulateStove ManipulateMicrowave ManipulateSinkOn + Off ManipulateFridge	77.1%
Wash	WashRinse WashSink WashDishwasher	72.2%
Walk1	WalkIntoKitchen(WIK)Hold WalkAroundKitchen(WAK)Hold WalkOutKitchen(WOK)Hold	29.2%
Walk2	WIKHold + NotHold WAKHold + NotHold WOKHold + NotHold	41.7%
WalkHold	WIKHold WAKHold WOKHold	69.3%
WalkNotHold	WIKNotHold WAKNotHold WOKNotHold	52.8%
PickPut1	All PickUp and PutDown Actions, Separated	37.5%
PickPut2	PickUpCounter + PutDown PickUpTop + PutDown PickUpBottom + PutDown	69.4%
OpenClose1	OpenTopCabinet CloseTopCabinet OpenBottomCabinet CloseBottomCabinet	54.2%
OpenClose2	OpenTopCabinet + Close OpenBottomCabinet + Close	75.0%
Mixed1	ManipulateFridge CloseBottomCabinet WIKHold	97.2%
Mixed2	WashSink WAKNotHold PrepStir	97.2%
Mixed3	OpenTopCabinet + Close PrepOpen + Close + Cut WIK + WOK, Hold + NotHold ManipulateMicrowave ManipulateStove ManipulateFridge	86.9%

Initial datasets were put together under their theme. After running a raw category, actions that are similar are combined. In each instance the accuracy improved. The last three datasets have members that are from mixed categories to better imitate the diverse actions possible in a kitchen. Accuracy is important so therapists can have a better understanding of what actions a stroke patient is performing.

The cross-validation method chosen was the holdout method with 4 items in the test set. The rest of the data were assigned to a training set. This was performed 3 times and averaged. The datasets and their accuracies are summarized in Table I.

The confusion matrix in Figure 5 was produced from a dataset of 50 random actions from each group described in Section III.A, which amounts to 250 items. 20 actions from each category of this new subset were put into a test set, totaling 100. The rest, 150, were used as a training set.

Overall, the precision is 70%. Individually, Walking, Preparation, Picking, Manipulation, and Washing have the precisions 83.3%, 81.3%, 73.7%, 58.3%, and 53.0%, respectively. Manipulation and Washing have the lowest precisions because of similar hand movements and standing positions. Manipulation is incorrectly labeled as Washing 25% of the time, and Washing is falsely labeled as Manipulation 35% of the time.

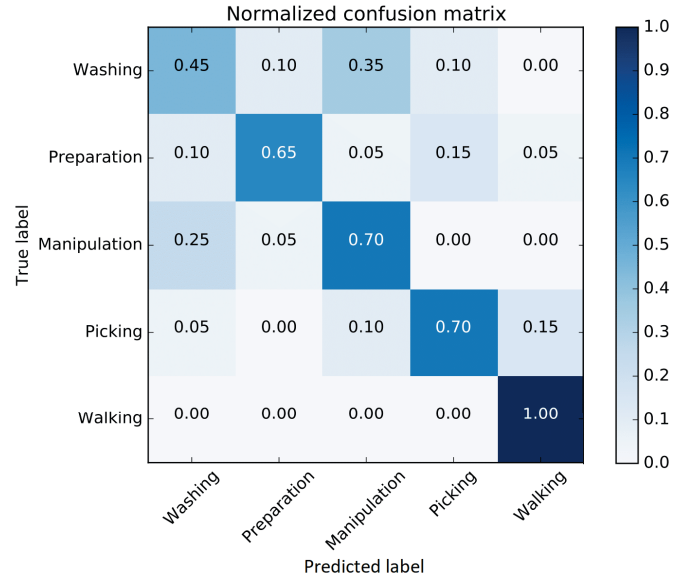


Fig. 5. A confusion matrix generated from a 100 item test set and 150 training set.

V. ASSESSMENT RESULTS

DAOS also quantitatively evaluated kitchen actions using assessment metrics including maximum/mean speeds, maximum hand extension, symmetry analysis of hands, and chest sway. Assessment results for speed metrics and max extensions are summarized in Table II for selected actions.

To calculate metrics for assessment, it is crucial for the Kinect device to record accurate skeleton data. For assessment of extension, the length between two joints can be reasonably accurate with the Kinect, with the average inaccuracy around 50mm [16]. The assessment method from DAOS is similar to the assessment provided by a stroke game, Mystic Isle [3], whose results were validated in [17].

People with stroke may have decreased range of motion on one or both hands; the symmetry can be illustrated by plotting the hand trajectories as 2D graphs using the skeleton data. Figure 6 shows the left and right hand trajectories of opening a container. Since the data were collected from a healthy subject, the movements of both hands cover almost the same amount of area in the X-Z plane in Figure 6. Figure 7 shows the

chest sway for the same action. Most of the points are focused around the jar to be opened, which is to be expected.

TABLE II
SPEED METRICS IN M/S AND MAX EXTENSION METRICS IN METERS

Hand	Max Speed		Mean Speed		Max Extension	
	Left	Right	Left	Right	Left	Right
CloseTopCabinet	2.49	2.87	1.63	1.62	0.63	0.64
PrepCut	0.92	3.03	0.20	0.88	0.63	0.64
PutDownCounter	0.30	0.23	0.22	0.19	0.60	0.61
WalkOutNotHold	2.23	2.64	1.18	1.04	0.68	0.62

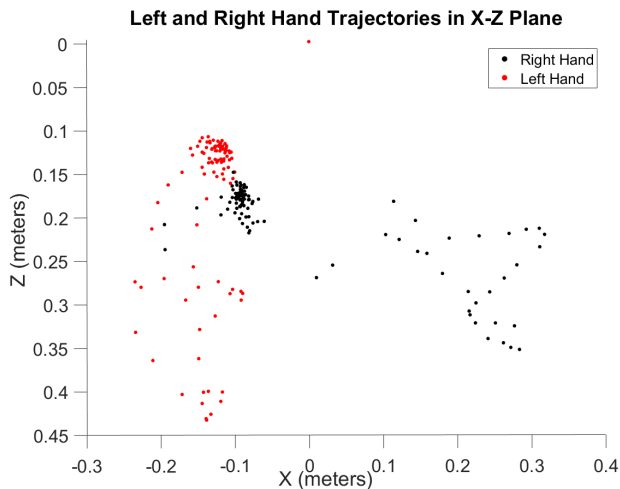


Fig. 6. A 2D plot of hand movements in PrepOpen. Most of the points are centered around the jar being opened.

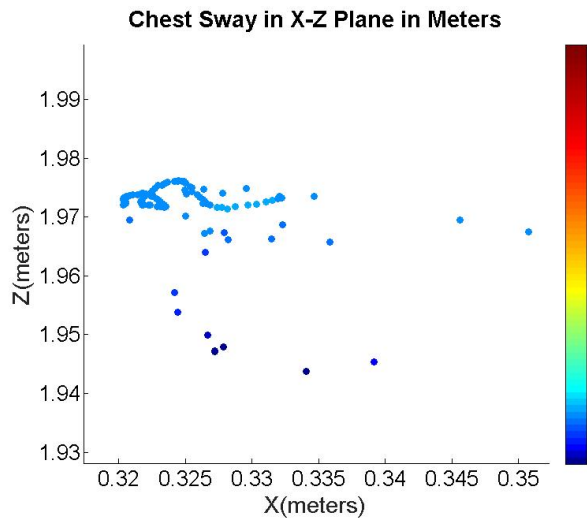


Fig. 7. A 2D plot of chest sway in PrepOpen. While leaning closer to the jar, the chest sways. When the jar is opening, the points are in a tight group.

VI. CONCLUSION

We presented a novel solution for occupational therapists to create customized care for stroke rehabilitation patients. Through a daily observation system using a Kinect v2, we

collect depth information and skeleton data. With these data, we are able to recognize several actions with a high degree of certainty by implementing HON4D as a global descriptor. In addition to activity recognition, we are able to perform assessment on critical metrics such as arm extension, mean velocity, max velocity, hand trajectories, and chest sway. With the extensive knowledge our daily observation system reports, a therapist can see problem areas or improvements over time to provide superior, personalized care for a stroke patient.

ACKNOWLEDGMENT

This material is based upon work supported by the National Science Foundation under Award Number: CNS-1659134. Any opinions, findings, and conclusions or recommendations expressed in this publication are those of the author(s) and do not necessarily reflect the views of the NSF.

REFERENCES

- [1] R. Proffitt and B. Lange, "Considerations in the efficacy and effectiveness of virtual reality interventions for stroke rehabilitation: Moving the field forward," *Physical Therapy*, vol. 95, no. 3, pp. 441–448, 2014.
- [2] K. Nair and A. Taly, "Stroke rehabilitation: traditional and modern approaches," *Neurology India* 50, pp. S85 – 93, 2002.
- [3] R. Proffitt and B. Lange, "The feasibility of a customized, in-home, game-based stroke exercise program using the microsoft kinect sensor," *Int. J. of Telerehabilitation*, vol. 7, no. 2, pp. 23–34, 2015.
- [4] B. Galna, G. Bary *et al.*, "Accuracy of the microsoft kinect sensor for measuring movement in people with parkinson's disease," *Gait & Posture*, vol. 39, no. 4, pp. 1062 – 1068, 2014.
- [5] O. Oreifej and Z. Liu, "Hon4d: Histogram of oriented 4d normals for activity recognition from depth sequences," in *2013 IEEE Conf. on Comput. Vision and Pattern Recognition*, June 2013, pp. 716–723.
- [6] M. Ye, Q. Zhang *et al.*, *A Survey on Human Motion Analysis from Depth Data*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2013, pp. 149–187.
- [7] W. Li, Z. Zhang, and Z. Liu, "Action recognition based on a bag of 3d points," *2010 IEEE Comput. Soc. Conf. on Comput. Vision and Pattern Recognition - Workshops*, 2010.
- [8] X. Yang, C. Zhang, and Y. Tian, "Recognizing actions using depth motion maps-based histograms of oriented gradients," *Proc. of the 20th ACM international conference on Multimedia - MM 12*, 2012.
- [9] W. Li, Z. Zhang, and Z. Liu, "Expandable data-driven graphical modeling of human actions based on salient postures," *IEEE Trans. on Circuits and Sys. for Video Technology*, vol. 18, no. 11, pp. 1499–1510, Nov 2008.
- [10] A. D. L. Reyes-Guzmn, I. Dimbwadyo-Terrer *et al.*, "Quantitative assessment based on kinematic measures of functional impairments during upper extremity movements: A review," *Clinical Biomechanics*, vol. 29, no. 7, pp. 719–727, 2014.
- [11] M. E. Nixon, A. M. Howard, and Y.-P. Chen, "Quantitative evaluation of the microsoft kinect for us in an upper extremity virtual rehabilitation environment," *2013 Int. Conf. on Virtual Rehabilitation*, Nov 2013.
- [12] A. K. Mishra, M. Skubic, and C. Abbott, "Development and preliminary validation of an interactive remote physical therapy system," *2015 37th Annual Int. Conf. of the IEEE Engineering in Medicine and Biology Soc. (EMBC)*, Nov 2015.
- [13] A. Scano, M. Caimmi *et al.*, "Using kinect for upper-limb functional evaluation in home rehabilitation: A comparison with a 3d stereoscopic passive marker system," *2014 5th IEEE RAS & EMBS Int. Conf. on Biomedical Robotics and Biomechanics*, Oct 2014.
- [14] "Data streams," [Online]. Available: <http://bit.ly/MMSDN>. [Accessed: 6–Oct–2017].
- [15] *Recovering After a Stroke: A Patient and Family Guide*, 16th ed., ser. 95-0664. AHCPR, Rockville, MD, 2017.
- [16] M. Elgendi, F. Picon, and N. Magnenat-Thalmann, "Real-time speed detection of hand gesture using kinect," 05 2012.
- [17] M. Ma, R. Proffitt, and M. Skubic, "Quant. assessment and validation of a stroke rehab. game," *2017 IEEE/ACM Int. Conf. on Connected Health: Applicat., Syst. and Eng. Technol.*, pp. 255–257, Jul 2017.