Depth Sensor-Based In-Home Daily Activity Recognition and Assessment System for Stroke Rehabilitation

Zoë Moore
Computer Science, Mathematics
Truman State University
Kirksville, MO
zkm4326@truman.edu

Carter Sifferman
Mathematics & Computer Science
Drury University
Springfield, MO
csifferman@drury.edu

Shaniah Tullis
Computer Science, Mathematical Sciences
Lincoln University
Lincoln University, PA
shaniah.tullis@lincoln.edu

Mengxuan Ma
Electrical Engineering & Computer Science
University of Missouri
Columbia, MO
mmrnc@mail.missouri.edu

Rachel Proffitt
Occupational Therapy
University of Missouri
Columbia, MO
proffittrm@health.missouri.edu

Marjorie Skubic
Electrical Engineering & Computer Science
University of Missouri
Columbia, MO
skubicm@missouri.edu

***Abstract*—Stroke is a leading cause of long-term adult disability. Many stroke patients participate in rehabilitation programs prescribed by an occupational therapist to aid in recovery; however, occupational therapists rely on in-clinic assessments and often-unreliable self-assessments at home to track a patient's progress, limiting their ability to monitor how patients perform outside of a clinical setting. Our Daily Activity Recognition and Assessment System collects depth and skeletal data passively from within the patient's home to assess long-term recovery and provide metrics to an occupational therapist to allow for more individualized rehabilitation plans. Using data from a wall-mounted depth sensor, we adapt a hierarchical co-occurrence network to identify actions from pre-segmented skeletal data. We then perform assessments on the classified actions to track key recovery metrics: normalized jerk, speed of motions, and extent of reach. We also introduce novel filters to identify high quality data for analysis. Our sensor was installed in a stroke patient's kitchen for seven days, generating the first action recognition data set from a stroke patient in a naturalistic environment. We use this data in conjunction with the NTU-RGB-D data set to validate our recognition and assessment algorithms. We achieved 90.1% accuracy by replicating the results of the NTU-RGB-D data set and a maximum of 59.6% accuracy on our kitchen data set.**

***Index Terms*—Action Recognition, Hierarchical Co-Occurrence Network, Stroke Rehabilitation, Skeletal Data, Depth Data, Foresite DS5 Sensor**

I. Introduction

Every year, around 795,000 people are affected by a stroke in the United States, with 50% experiencing hemiparesis, or a weakening of one side of the body [1, 2]. With the help of an occupational therapist, stroke patients can avoid or mitigate these symptoms by adhering to a rehabilitation plan [3]. Every patient experiences varying degrees of symptoms, so prescribing a personalized rehabilitation plan is vital for a quick and effective recovery; however, a personalized rehabilitation plan necessitates individualized metrics, which are time- and resource-intensive to collect. Rehabilitation progress is typically measured through periodic in-clinic evaluations, but participation in rehabilitation is low, with only 30.7% of stroke patients receiving outpatient rehabilitation, and in-clinic assessments are often subjective, and do not capture patient functionality outside of a clinic [1].

Recently, work has been done to provide independent rehabilitation for patients at home [3]. Mystic Isle, an interactive video game utilizing the Microsoft Kinect depth sensor, encourages patients to simulate in-clinic rehabilitation exercises at home and tracks quantitative motion data [3]. While Mystic Isle offers a substantial improvement over previous in-home rehabilitation techniques, it still requires a time commitment from patients. Additionally, while the game does encourage the patient to participate in exercises, it does not capture and assess their movement when doing daily tasks, and it offers limited insight into long-term recovery.

We propose a Daily Activity Recognition and Assessment System (DARAS) using a DS5 depth sensor from Foresite Healthcare to recognize and assess actions in a naturalistic, in-home setting. The goal of the system is to provide metrics of the stroke patients' motions to occupational therapists to help them create more personalized rehabilitation plans and have a record of patient progress over the course of their recovery. DARAS has three main components as seen in Fig. 1: data logging, action recognition, and action assessment. The DS5 uses the Orbbec depth sensor, and our data logger utilizes the Orbbec Astra SDK to capture frames of depth data and extract

This material is based upon work supported by the National Science Foundation under Award Number: CNS-1359125. 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 National Science Foundation.

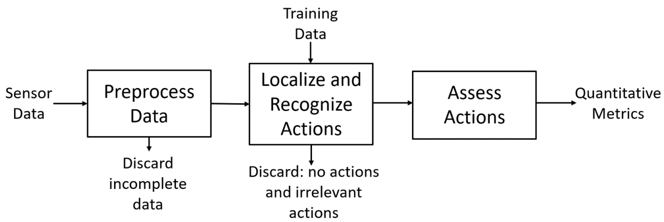

Fig. 1. Proposed components of DARAS.

skeletal data for analysis while preserving the privacy of the patient. The logger is configured to only capture data when a figure is detected in the frame, saving the data to a local storage device for later retrieval. To perform action recognition, we adapt a hierarchical co-occurrence network (HCN) from [4]; the HCN takes pre-segmented skeletal data as input and returns a recognized action. The assessment component is adapted from Mystic Isle, which evaluates metrics such as speed of motion, extent of reach, and normalized jerk [5]. Coupled with action recognition, we are able to compare metrics of similar actions over time to track and assess a patient's long-term recovery.

DARAS supports many features for stroke rehabilitation. The sensor is robust enough for extended in-home deployment, being able to accommodate multiple persons in the view and automatically segment the skeletal data from the individual bodies. Additionally, the system reboots and notifies the user should any crash or power outage occur. Depth and skeletal data are saved locally for assessment, enabling occupational therapists to longitudinally compare metrics.

The paper is organized as follows. We survey related work in Section II. We outline our implementation of the data logging, action recognition, and action assessment in Sections III, IV, and V, respectively. Our experiment and results are presented in Section VI and we discuss our results and conclusion in Section VII.

II. Related Work

A. Action Recognition

Since the release of the Microsoft Kinect in 2010, depth sensors have become an affordable and readily available method of capturing depth and skeletal data for action recognition [6]. Action recognition has typically been performed on video inputs which contain only one action. Early action recognition relied on extracting features from the visual input to guide analysis, using techniques like histogram of oriented gradients and histogram of oriented 4D normals, where vectors are extracted from RGB or depth data to interpret the motions [7], [8]. Similarly, depth motion maps are used to segment depth data into three cartesian planes that guide the recognition of actions [9]. With deep learning becoming increasingly prevalent, however, convolutional neural networks have become a more successful method of action recognition, typically outperforming other techniques [10]. Action recognition methods have been developed primarily for classifying actions, typically training and operating on pre-segmented videos; however, using pre-segmented videos limits action recognition in practical situations. Outside of a laboratory environment, individuals perform actions sequentially rather than in disjoint, isolated segments, so an optimal action recognition algorithm would be able to detect and segment actions from a continuous video feed.

B. Action Assessment

Virtual reality rehabilitation systems which utilize the Microsoft Kinect have been around for some time. Proffitt and Lange have documented the developmental research and history surrounding these systems, finding that a virtual reality environment is effective for stroke rehabilitation. The virtual reality system encourages patients to perform their rehabilitation exercises and tracks their progress over time [11]. The same researchers have also had success implementing their own virtual reality game with the Microsoft Kinect called Mystic Isle [3]. Recently, an action assessment library has been developed for Mystic Isle, which we utilize [3].

In order for these rehabilitation assessment tools to be effective, it is important that skeletal data are tracked accurately. By comparing data from the Kinect v2 to a state-of-the-art Vicon motion capture system, researchers have found that the Kinect is comparable in accuracy when tracking upper-body joints, but less accurate when tracking the lower body, especially the hips [12]. It has also been found that the extent-of-reach metrics determined by trained clinicians are comparable in accuracy to those found by the Kinect sensor [13]. However this accuracy has only been applied in structured environments such as in clinics or virtual reality games; no similar studies have been done to track the daily activities of stroke patients and quantitatively assess their movement in a natural environment, such as the home.

III. Data Logging

A. Sensor Installation

The action logging component of our system is implemented on-board the Foresite DS5 sensor. Our logging program runs on a low-power linux computer inside the integrated DS5 system. The logger utilizes the Astra SDK to detect when a person is in view, and it only saves recordings if a person is detected. The sensor can record up to 30 frames-per-second of depth images, but the on-board Linux computer limits the speed of saving the data, reducing the recording rate to seven frames-per-second. The depth images are saved in compressed binary files which are extracted after data collection, and skeletal data is periodically written to .csv files. When recording exclusively skeletal data, the logger can reach 30 frames-per-second consistently, and because the .csv files are significantly smaller than depth images, the improved frame rate does not cause storage concerns. An example of depth and skeletal data can be seen in Fig. 2.

Body segmentation of the skeletal data is done in real time by our data collection system. When multiple bodies are detected in the frame, they are both written to the same .csv

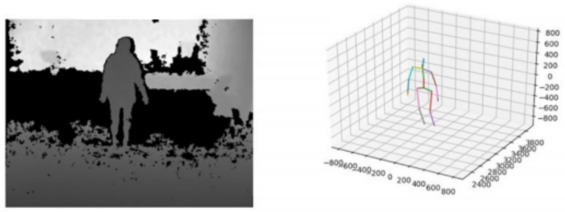

Fig. 2. Data collected by the DS5 sensor.

file and labeled with a unique body identification number to differentiate individuals between frames. Skeletal data can then be easily parsed into body-specific files after data collection. Segmentation of each frame into multiple bodies is crucial for differentiating the patient from other individuals in the home. Pre-segmented skeletal data allows us to run separate assessments on each body. Because we have the ability to capture depth and skeletal data concurrently, a depth-based body re-identification algorithm such as in [14] could be implemented on depth frames and be matched back to skeletal frames in order to identify the individual being tracked. In our case, we can manually select the body of the stroke patient for assessment.

B. Data Set

To test the capabilities of our system in a naturalistic environment, we gathered data in a stroke patient's restaurant kitchen. The kitchen was selected as a location where the patient often performs manipulation and reaching tasks. The kitchen proved to be a challenging environment for our system, as it was visually busy, with many metal surfaces which are challenging for depth sensors to detect, and it often contained multiple bodies at a time. Bodies were also prone to self-obfuscation, as they turned to the side to face workstations. Our DS5 was ceiling mounted and pointed at a busy part of the kitchen for seven days. In that time period we gathered 300,819 frames of data, equating to 12 hours and 51 minutes of recording. Of these frames, 159,343 contain exactly one person as detected by the sensor. We focus on these frames for our analysis.

For each frame of skeletal data we record a timestamp and for each body we record a body identification number as well as the x, y, and z coordinates of 19 key body joints. We also record the status of each joint as either 'highly confident' or 'not tracked' as provided by the SDK. While we capture both depth and skeletal data, only skeletal data are used for assessment and recognition. The depth data are still useful for manual labeling, and it will be necessary for future applications, such as person recognition.

C. Data Preprocessing

To compensate for the challenging kitchen environment that our data was gathered, we introduce two filters, standard deviation of bone lengths and upper body, to eliminate data which is too poor for model training or assessment.

Our novel bone length approach takes advantage of the fact that in reality some joints are connected by bones and remain an equal distance from each other, e.g. the elbow and wrist joint. We use the fluctuation in distance between these joints as a proxy for the accuracy of our joint data, with high fluctuations indicating poor data. Because we are only interested in upper body movements, we limit this analysis to the following upper body joint connections: spine-right shoulder, right shoulder-right elbow, right elbow-right wrist, right wrist-right hand, spine-left shoulder,left shoulder-left elbow, left elbow-left wrist, and left wrist-left hand. We test recognition on the 50% of actions with the lowest standard deviation, which corresponds to a standard deviation below 22.84. Additionally, we are able to localize stable sections of data within one file by finding the mean standard deviation in a neighborhood around each frame and giving priority to lower-deviation data. This could prove helpful for action localization, which is important for the longitudinal deployment of the system in a home. An example of the process is shown in Fig. 3.

The upper body filter limits our data to only the upper body joints. This is done because the upper body is tracked more consistently, and is more relevant to our actions than the lower body. We restrict the input to the following 12 joints: head, shoulder-spine, left shoulder, left elbow, left hand, right shoulder, right elbow, right hand, mid-spine, left wrist, right wrist, and neck.

IV. Action Recognition

In order to provide in-home quantitative metrics to occupational therapists, data must be parsed into individual actions so assessment can be on an action-by-action basis. Each frame in our novel data set has been manually labeled for training and validating our recognition model. The actions were decided with help from an occupational therapist; they allow assessment on both halves of the body and are heavily impacted by stroke.The labels are: walking, reaching above the shoulder, reaching forward between the shoulder and the waist,

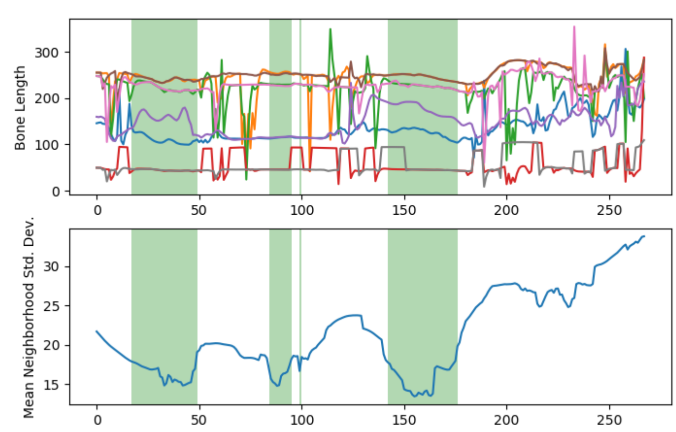

Fig. 3. Graph representing standard deviations in bone length. The highlighted regions represent standard deviations below a set threshold.

reaching below the waist, object manipulation, sweeping, and background. Background is being used as a catch-all category for any actions that do not fit into any other category.

We have adopted the Hierarchical Co-Occurrence Network (HCN) as seen in [4] which performs action recognition on skeletal data. The HCN was developed to perform recognition on the NTU-RGB-D data set, a pre-segmented data set containing 3D coordinates of 25 joints performing 60 unique actions. The novelty of the HCN network comes from the way it extracts co-occurrence features from sequences of skeletal data. Features are aggregated from joint-level to the global level and fed to a convolutional neural network. This enables HCN to outperform other methods of joint-based action recognition [4].

We adapt the HCN to train on our novel data set, which contains 19 joints and 7 unique actions. This required reformatting the kitchen dataset into the NTU-RGB-D data format. Adapting the network required reducing the number of input nodes from 25 to 19, and reducing the number of classifications from 60 to 7. All other details of the model are left the same as in [4], including a learning rate of 0.001 and the use of the Adam optimizer from [15].

V. Action Assessment

The end goal of our assessment component is to provide quantitative data to track patient progress over time, enabling occupational therapists to create customized rehabilitation plans. We accomplish this by providing metrics which are routinely gathered in the clinic, such as extent of reach, and metrics which are not typically gathered by clinicians, such as normalized jerk. Additionally, our system gathers metrics from within the home, which gives a more holistic view of recovery. For example, our system could track how often a patient reaches with their right hand compared to their left, which could give insight into the symptoms of their hemiparesis.

We adapt the assessment component from Mystic Isle [3]. From any segment of joint data, we gather the following metrics for each hand: maximum speed, mean speed, normalized jerk$_{x,y,z}$, and extent of reach$_{x,y,z}$. From these metrics we are able to calculate the following secondary metrics: mean / maximum speed, normalized jerk, 3D extent of reach away from shoulder, area of reach, and volume of reach. These metrics can be split into two categories: those measuring shakiness of movement (speed and jerk) and those measuring extent of reach. These metrics have been developed with an occupational therapist to capture recovery [3].

VI. Experiments and Results

A. Action Recognition

To ensure the integrity of the HCN, we first replicate the results on a cross-view evaluation protocol of the NTU-RGB-D data set. The default training method uses a learning rate of 0.001. After 400 epochs, our results were on par with those reported in [4], at 90.1% accuracy compared to their 91.1%.

We then run the HCN on our field-collected data, and we perform a number of tests on filtered data in order to optimize our recognition accuracy. We filter by three parameters: relevant actions, quality of the skeletal data, and part of the body. We do not perform any filtering on the NTU-RGB data set, because the data are already of very high quality, and filtering the data only reduces the amount of information the model has to train on, without increasing the data quality significantly. To assess the accuracy of our model, we train on a random 70% of the data and test on the remaining 30%; all accuracy results are the results on the testing data set. A summary of our results are shown in Table I and corresponding confusion matrices in Fig. 4. The results achieved on our data set are considerably lower than those achieved on the NTU-RGB-D data set. We attribute this largely to the noisy data from the kitchen environment. The rest of this section will go into depth cross-comparing tests using different filters.

We ran tests limiting input to only the 12 upper body joints, as shown in Table II. This was an attempt to reduce noise coming from the lower body. In addition, most of the actions we aim to recognize feature primarily upper body motions. In general, we find that using only upper body joints does improve recognition accuracy. In the test that filters both by bone length and removes background and sweeping actions, the reverse is seen. We attribute this to a significant reduction in the quantity of training data, which is explored primarily with Table III.

We also perform action recognition on data filtered to only the 50% of actions with the lowest standard deviation of bone length, as explained in section III. In general, using the less variable half of our data improved our recognition accuracy by approximately 10% on both the entire body and the upper body data, suggesting that it is a reliable filtering method. However, when we use the bone length filter in addition to filtering out the background and sweeping actions, the bone length filter negatively impacts recognition accuracy. We attribute this to filtering out too much data, resulting in a data set too small to train on. The bone length filter halves the size of our data set, so removing any additional actions in conjunction with this filter leaves us with a data set that is a fraction of what we normally have. In the instance of filtering out both background and sweeping actions from the bone length filters, we're left with roughly a third of our original data.

Finally, we compare data sets with all actions against data sets without both background and sweeping actions, as seen in Table IV. We remove background data because it functions as a catch-all for actions which did not fit into any other category or were too obscured to be fully recognized. We find that only applying the action filter on our data noticeably improves recognition accuracy. However, when the action filter is used with the bone length filter, recognition accuracy decreases. Once again, we attribute this to the significantly reduced quantity of training data.

With each of our trials, we find that these filters improve our recognition accuracy until we have too little data to successfully train the HCN. These data-filtering techniques are useful methods for improving action recognition accuracy. We see reduction in recognition accuracy when combining filters;

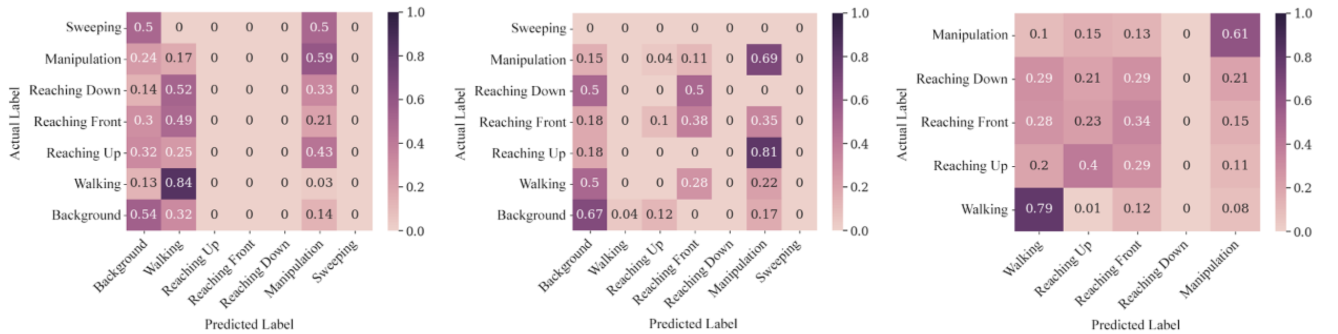

Fig. 4. Confusion matrix results for the following experiments: no filter, best 50% of bone length filter on upper body, no background and no sweeping filter.

however, on larger data sets, this could be done without too greatly reducing the quantity of training data available. The ideal DARAS, when fully clinically-ready, will accommodate background actions interspersed with clinically-relevant actions. Therefore, we place greater importance on the bone length and upper body filters.

TABLE I
SUMMARY OF ACTION RECOGNITION RESULTS

	NTU-RGB-D	**Default**	**Best Without Removing Actions**	**Best Overall**
Part of Body	N/A	Entire body	Upper body	Entire body
Actions	N/A	All actions	All actions	Without background and sweeping
Filter	N/A	None	Best 50% of bone length	None
Accuracy	90.1%	40.0%	55.6%	59.6%

TABLE II
UPPER BODY FILTER RESULTS

Test	**Whole Body**	**Upper Body Only**
No Filter All Actions	40.0%	**44.9%**
No Filter No Background / Sweeping	59.6%	**50.0%**
Best 50% Bone Length All Actions	49.4%	**55.6%**
Best 50% Bone Length No Background / No Sweeping	**46.2%**	44.1%

TABLE III
BONE LENGTH FILTER RESULTS

Test	**No Filter**	**Bone Length**
Whole Body All Actions	40.0%	**49.4%**
Upper Body All Actions	54.9%	**55.6%**
Whole Body No Background / Sweeping	**59.6%**	46.2%
Upper Body No Background / No Sweeping	**50.0%**	44.1%

B. Action Assessment

We apply our assessment component to the skeletal data from our data set. In an ideal system, recognition results would be fed directly into assessment, however, we manually select segments for assessment to ensure that the segments are accurate and demonstrate our system's capabilities. Our assessment results demonstrate that meaningful metrics can be extracted from skeletal data, even in a naturalistic setting. Comparing metrics such as these over time could prove beneficial to occupational therapists. In our case, our patient's right side was affected by stroke, which can be seen in the assessment results displayed in Table V.

VII. DISCUSSION

There are a number of limitations that impact the success of our system. The foremost limitation is data quality. It is clear from our action recognition results, specifically comparing the NTU-RGB-D data set to our field data, that there is a reduction in quality when moving from a laboratory to a real-world environment, which significantly impacts recognition and assessment performance. The lower frame rate of our data set when compared to the NTU-RGB-D data set might also have an impact on recognition accuracy. Additionally, despite the week-long installation of our system, the size of our data set limits the use of filters, impacting recognition results when multiple filters are applied due to lack of remaining data. These limitations and proposed future solutions are discussed below.

TABLE IV
ACTION FILTER RESULTS

Test	**All Actions**	**No BG / Sweeping**
Whole Body No Filter	40.0%	**59.6%**
Upper Body No Filter	44.9%	**50.0%**
Whole Body Top 50% Bone Length	**49.4%**	46.2%
Upper Body Top 50% Bone Length	**55.6%**	44.1%

We learned many things about the installation process that may help to improve data quality in the future. First, an optimal installation angle and installation environment are important for capturing accurate skeletal data. Specifically, our kitchen environment was primarily metal, which the depth camera had trouble detecting, and the counters of the kitchen being perpendicular to the camera meant that figures in the view were often self-obfuscating. To avoid these issues, a camera should be installed so that a front-on view of the subject is more likely. Installation in a light-colored environment away from metal, sunlight, and heat sources is also optimal as these can hinder the clarity of the depth image. Encouraging subjects to avoid wearing black can also ensure that the depth sensor captures depth detail in their silhouette. Future data collection should seek to address these issues.

The process of saving depth images on-board the Foresite DS5 sensor is computationally taxing, limiting the frame rate to 7 fps rather than 30. Fewer frames limits the smoothness and intricacy of movement captured, making recognition more difficult. This issue can be addressed with faster hardware, or by saving exclusively skeletal data. With the HCN algorithm, skeletal-only action recognition is very feasible, making saving skeletal data a reasonable future adjustment.

While action recognition results on our data set are less accurate than lab data, we have shown that filters may be effective strategies for dealing with sub-optimal skeletal data. Our novel bone length metric seems to have a positive impact on action recognition performance, as does limiting input to upper body joints. While these filters are helpful, if applied too aggressively or in combination with each other they reduce the amount of training data available. We believe future studies with larger data and more aggressive filters may be able to achieve better results.

We believe the metrics gathered by our assessment system will be sufficient for an occupational therapist to use in creating a personalized rehabilitation plan. There is, however, information gathered by our system that is not entirely reflected in those metrics. One such metric is symmetry of hand usage. In future work, an action recognition algorithm will be trained to recognize which hand is being used for reaching, and return a count of how often each hand is used. We believe this will be a very telling metric for patients with hemiparesis. Our stroke patient reached almost exclusively with her left hand, limiting our ability to test such a system.

We are confident that further research will enable a fully automatic in-home monitoring system. With more research into data collection and action recognition, a complete in-home system is possible. Such a system will enable occupational therapists to give more personalized rehabilitation plans, and ultimately improve the lives of stroke patients.

REFERENCES

[1] National Center for Chronic Disease Prevention and Health Promotion, Division for Heart Disease and Stroke Prevention, “Stroke Fact Sheet,” cdc.gov, para. 4, Sept. 1, 2018.

[2] American Heart Association, “Heart Disease and Stroke Statistics—2012 Update,” *American Heart Association*, 125:e2-e220, 2011.

[3] J. Collins, J. Warren, M. Ma, R. Proffitt and M. Skubic, “Stroke patient daily activity observation system,” *IEEE Intl. Conf. on Bioinformatics and Biomedicine*, Kansas City, 2017, pp. 844-848.

[4] C. Li, Q. Zhhong, D. Xie, and S. Pu, “Co-occurrence Feature Learning from Skeleton Data for Action Recognition and Detection with Hierarchical Aggregation,”' *CoRR*, 2018.

[5] M. Ma, B. J. Meyer, L. Lin, R. Proffitt and M. Skubic, “VicoVR-Based Wireless Daily Activity Recognition and Assessment System for Stroke Rehabilitation,” *IEEE Intl. Conf. on Bioinformatics and Biomedicine*, Madrid, 2018, pp. 1117-1121.

[6] Z. Zhang, “Microsoft Kinect Sensor and Its Effect,” in *IEEE MultiMedia*, vol. 19, no. 2, pp. 4-10, Feb. 2012.

[7] N. Dalal and B. Triggs, “Histograms of oriented gradients for human detection,” *IEEE Computer Society Conf. on Computer Vision and Pattern Recognition*, San Diego, 2005, pp. 886-893 vol. 1.

[8] O. Oreifej and Z. Liu, “HON4D: Histogram of Oriented 4D Normals for Activity Recognition from Depth Sequences,” *IEEE Conf. on Computer Vision and Pattern Recognition*, Portland, 2013, pp. 716-723.

[9] C. Chen, K. Liu, and N. Kehtarnavaz, “Real-Time Human Action Recognition Based on Depth Motion Maps,” *Journal of Real-Time Image Processing*, vol. 12, no. 1, pp. 155–163, 2013.

[10] I. Kapsouras and N. Nikolaidis, “Action Recognition by Fusing Depth Video and Skeletal Data Information,” *Multimedia Tools and Applications*, Volume 78, Issue 2, pp 1971-1998, Jan. 2019.

[11] R. Proffitt and B. Lange, “Considerations in the Efficacy and Effectiveness of Virtual Reality Interventions for Stroke Rehabilitation: Moving the Field Forward,” *Physical Therapy*, Volume 95, Issue 3, pp. 41-448, March 2015. [Online]. Available: Oxford Academic, academic.oup.com. [Accessed June 19, 2019].

[12] M. Ma, R. Proffitt, M. Skubic. “Validation of a Kinect V2 based rehabilitation game,” in *PLoS One*, vol. 13, no. 8, Aug. 2018.

[13] M. E. Nixon, A. M. Howard and Y. Chen, “Quantitative evaluation of the Microsoft KinectTM for use in an upper extremity virtual rehabilitation environment,” *2013 International Conference on Virtual Rehabilitation (ICVR)*, Philadelphia, PA, 2013, pp. 222-228.

[14] A. Wu, W. Zheng and J. Lai, “Robust Depth-Based Person Re-Identification,” in *IEEE Transactions on Image Processing*, vol. 26, no. 6, pp. 2588-2603, June 2017.

[15] D. Kingma and J. Ba, “Adam: A Method for Stochastic Optimization,” in *International Conference on Learning Representations*, May 2015

TABLE V
ACTION ASSESSMENT OF LEFT AND RIGHT HAND MOVEMENTS

Action Label[a]	**Extent 3D (mm)**[b]		**Extent Volume (cm^3)**[c]		**Max Velocity (mm/s)**		**Mean Velocity (mm/s)**		**Mean / Max Velocity**		**Normalized Jerk**	
	L	***R***	***L***	***R***	***L***	***R***	***L***	***R***	***L***	***R***	***L***	***R***
Left Hand Up	662.14	671.99	2962	1766	1041.42	498.08	315.26	105.31	0.30	0.21	352.18	108.48
Right Hand Forward	527.73	573.98	5104	2121	725.00	653.69	459.26	55.29	0.63	0.269	29.29	59.28
Manipulation	608.15	632.85	144	1154	205.45	333.88	55.29	172.15	0.27	0.52	371.64	761.88

[a] While the action labels specify the hand performing the action, we analyze motion for both the left and right hands in the the *L* and *R* columns respectively.
[b] Extent 3D indicates the furthest distance the hand is from the midpoint of the two shoulders.
[c] Extent Volume indicates the 3D volume a cube surrounding their extent of reach would have.