01 Jan 2023

Patients Arms Segmentation And Gesture Identification Using Standalone 3D LiDAR Sensors

Omar Rinchi

Nathanael Nisbett

Ahmad Alsharoa
Missouri University of Science and Technology, aalsharoa@mst.edu

Recommended Citation

O. Rinchi et al., "Patients Arms Segmentation And Gesture Identification Using Standalone 3D LiDAR Sensors," *IEEE Sensors Letters*, Institute of Electrical and Electronics Engineers, Jan 2023.
The definitive version is available at https://doi.org/10.1109/LSENS.2023.3303081

Sensor applications

Patients Arms Segmentation and Gesture Identification Using Standalone 3-D LiDAR Sensors

Omar Rinchi*, Nathanael Nisbett, and Ahmad Alsharoa**

Electrical, Computer Engineering Department, Missouri University of Science, Technology, Rolla, MO 65409 USA
** Student Member, IEEE*
*** Senior Member, IEEE*

Manuscript received 3 June 2023; accepted 19 July 2023. Date of publication 7 August 2023; date of current version 18 August 2023.

Abstract—The intelligent and autonomous learning of patients' activities will lead to an incredible progression toward future smart e-health systems. With the recent advances in artificial intelligence, signal processing, and computational capabilities; light detection and ranging (LiDAR) technology can play a significant role in enhancing the current patients' activity recognition (PAR) systems. In this letter, we propose confidential and accurate patient arms behavior monitoring using a standalone 3-D LiDAR sensor. Due to the unavailability of LiDAR data, we use a computer-programmed 3-D simulator to generate virtual-LiDAR (VLiDAR) 3-D point cloud data that simulates real patient movements. These virtual data are used to train a multilayer-perception (MLP) model to segment the data points of the patient's body into arms versus not arms. We further propose a subsegmentation technique to segment patient's arms point cloud data into upper or lower arms. Finally, we demonstrate uses of arms gesture identification using the proposed scheme. The numerical results show that the proposed MLP model achieves a test accuracy of 90.8% and a cross-validation accuracy of 87.4%.

Index Terms—Sensor applications, arms segmentation, light detection and ranging (LiDAR), neural networks, multilayer perceptron (MLP).

I. INTRODUCTION

Continuous surveillance monitoring of patients has a profound effect on reducing in-patient hospital-acquired harm and in living environments, especially among the elderly. Due to these concerns, in the United States (U.S.) alone, around 65% and 90% of hospital and postacute care patients, respectively, are being checked with nursing staff frequently [1]. According to the Bureau of Labor Statistics, there is a lack in the registered nursing workforce in the U.S. [2] while the colleges across U.S. will not be able to fulfill this demand [3]. In addition, the nursing services are expensive with an average nursing pay is around $50/h [4]. The use of sensors for automatic monitoring is one solution to reduce nursing efforts. This solution is referred to human activity recognition (HAR) solution, a process of gathering sensor-based data and processing them to identify human behaviors, actions, and interactions. Patients activity recognition (PAR) as a special case of HAR is more concerned in monitoring patients' activities. Several sensors can be used to conduct PAR such as [5], [6], [7]:

1) video-cameras;
2) depth sensors;
3) wearable devices;
4) wireless sensors.

The red–green–blue (RGB)-based video cameras are subjected to several limitations. For example, video-cameras infringe upon the privacy of people, require proper lighting conditions, can be subjected to optical illusions, and finally, produce 2-D images that lead to complex background subtraction. The RGB-depth-based depth sensors are mostly rusticated to short distances and indoor environments due to their interference with the infrared bands in the natural sunlight (e.g., Microsoft Kinect 2010). Wearable devices are limited in the PAR capabilities and require to be wearable all the time. Finally, wireless sensors are subjected to wireless typical limitations, such as noise, interference, and fading.

Light detection and ranging (LiDAR) technology can enhance PAR by outperforming all the pre-mentioned limitations [8]. LiDARs have been briefly utilized in the HAR/PAR literature [9], [10], [11], [12]. However, the available solutions have the following limitations:

1) They require fusing the LiDAR with other instruments for either improving the performance or ground truth construction [9], [10];
2) They are limited to 2-D LiDAR solutions [11];
3) They suffer from biased non-LiDAR point cloud data [12].

To the best of the authors' knowledge, 1) efficient and accurate HAR/PAR using a standalone 3-D LiDAR is still lacking, and 2) non of the LiDAR-based HAR/PAR works have considered human body parts segmentation.

In this letter, we propose a conditional and accurate patient arms segmentation using a standalone 3-D LiDAR. We utilize a 3-D simulator to generate a clinical situation of interest to collect virtual point cloud data. We process these data using a multi-layer-perception (MLP) architecture that segments the patient point cloud into arms versus nonarms. The arms point cloud is further subsegmented into upper and lower arms. Finally, we utilize the proposed segmentation to conduct gesture identification use cases. The main contributions of this letter are summarized as follows.

1) We develop a novel conditional and accurate patient arms segmentation using a standalone 3-D LiDAR.
2) We propose solving LiDAR datasets absence problem for HAR/PAR applications by utilizing computer-based simulations.
3) We propose using a computationally efficient MLP model to segment the point cloud into arms versus nonarms.
4) We propose a novel subsegmentation process to further segment the arms data into upper arm and lower arm.
5) We illustrate several use cases for gesture identification.

II. METHODOLOGY

The goal is to conduct patient body parts segmentation and to use this segmentation to identify different scenarios of patient arm gestures. We

Corresponding author: Omar Rinchi (e-mail: omar.rinchi@mst.edu).
Associate Editor: S. Dhanekar.
Digital Object Identifier 10.1109/LSENS.2023.3303081

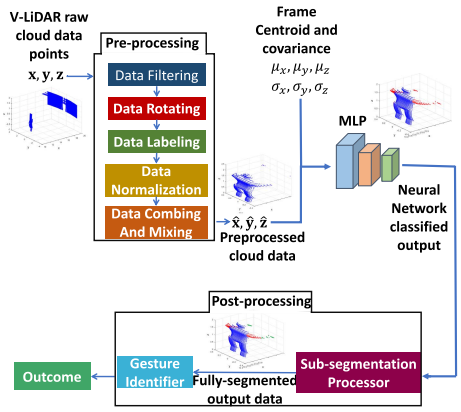

Fig. 1. Block diagram of the proposed solution.

simulate and program a desired clinical scenario in software where cloud data points are extracted using virtual-LiDAR (V-LiDAR). As shown in Fig. 1, the raw data undergo preprocessing before being classified into "arm" and "nonarm" by an MLP model. The "arm" data are then segmented into "upper" and "lower" arm using our processor. Finally, full-body gestures are identified.

A. Data Generating

One of the main challenges that face LiDAR-based HAR/PAR is the absence of LiDAR datasets especially datasets related to humans or clinical situations. On the contrary, generating 3-D cloud datasets for LiDARs is not an easy task due to the following reasons:

1) It can be a costly process;
2) Generating datasets can take a significant amount of effort;
3) Generating datasets is a time-consuming process;
4) Certain scenarios can be risky, complex, or hard to implement.

Due to the nature of LiDAR data, particularly its reliance on geometric information over the texture of the environment, various authors in the literature rely on simulations rather than experimental evaluation as a solution to the aforementioned limitations. For instance, Espineira et al. [13] utilized a computer 3-D simulator (e.g., Unreal Engine 4) to generate a scenario of LiDAR working against weather conditions. As a result, we propose the utilization of computer-programmed software to generate confidential V-LiDAR 3-D cloud data.[1] We use Webots [14] a 3-D robot simulator software to generate simulated scenarios of patients walking in the environment. We consider all the possible combinations of flexion and extension of the three main joints on both arms, which include: the glenohumeral joint, the elbow joint, and the radiocarpal joint. These movements capture a wide range of actions, including reaching, lifting, pushing, pulling, and other fundamental arm motions. Our focus on the isolated arm movements is motivated by its significant value as it has been reported in the healthcare literature, e.g., [15]. To guarantee the best generalization, we generate four different datasets where every dataset has a patient that walks in a different configuration. In dataset-1, the patient walks in an opposite U-shape where his back and front are facing the V-LiDAR. In dataset-2, the patient walks in a horizontal manner where his sides face the V-LiDAR. In dataset-3, the patient walks in a diagonal shape where he faces the V-LiDAR with an angle, the angle changes with time depending on his location. In dataset-4, the patient walks in an opposite-diagonal shape where he faces the V-LiDAR with a different angle, the angle again changes with time depending on his location. Each configuration might contain not only the patient's corresponding data but also the surroundings, such as walls, ground, and furniture. We adjust the V-LiDAR to make a sampling rate of 0.3 s between any frame and its following frame. A more detailed description of the data is given in Table 1.

[1]LiDARs do not collect personally identifiable information, such as facial features or textual data; hence, we describe it as confidential.

TABLE 1. Details of Each Data Set

data set	data set-1	data set-2	data set-3	data set-4	Total
Number of clouddata points	620,680	711,818	602,187	1,095,186	3,029,871
SimulationTime [s]	16.4130	25.5397	21.0187	39.0788	102.0502

B. Data Preprocessing

Data preprocessing is required to make the data in a suitable format. We describe the preprocessing in this section.

1) Data Filtering: The measured data contain not only the data corresponding to the patient, but also the surrounding data, such as walls, ground, and furniture. Therefore, to remove the unwanted data, we use region of interest (ROI) filtering. We define the ROI to be any data point with its x-coordinate less than xth and its z-coordinate above zth, such as

$$\begin{aligned} \text{if}(x_i > x\text{th}) &\rightarrow \text{delete the point} \\ \text{if}(z_i < z\text{th}) &\rightarrow \text{delete the point} \end{aligned} \tag{1}$$

where x_i and z_i refer to point i in the vectors $\mathbf{x}$ and $\mathbf{z}$ that contains the x-coordinate and the z-coordinate data of any frame.

2) Data Rotating: For the datasets with the patient moving diagonally (i.e., dataset-3 and dataset-4), easier data labeling can be done if the patient faces the V-LiDAR either in parallel or perpendicular. To do that, we define the following rotation matrix (RM) to rotate the data:

$$\begin{bmatrix} \cos(\theta_r) & \sin(\theta_r) & 0 \\ \sin(\theta_r) & \cos(\theta_r) & 0 \\ 0 & 0 & 1 \end{bmatrix} \tag{2}$$

where θ_r is the angle of rotation along the z-axis. Note that we rotate the data only for labeling purposes. After labeling, we restore the data to its original status.

3) Data Labeling: For training purposes, every data point must be labeled. We label every data point as an arm or nonarm. For this reason, we define a binary classifier, such as binary-1 refers to data that belongs to the arm, and binary-0 refers to data that does not belong to the arm. We define the following rule as follows:

$$\begin{aligned} \text{if}(v_i < (\mu_v - k_1\sigma_v)) &\rightarrow \text{binary-1} \\ \text{if}(v_i > (\mu_v + k_2\sigma_v))) &\rightarrow \text{binary-1} \\ \text{else} &\rightarrow \text{binary-0} \end{aligned} \tag{3}$$

where v_i is the ith point in the vectors $\mathbf{v}$ after filtering, and rotating. The vector $\mathbf{v}$ can be either the vector $\mathbf{x}$ or $\mathbf{y}$ depending on the dataset and time instant, μ_v is the mean value of all the data along the vector $\mathbf{v}$ in a specific frame, σ_v is the standard deviation of that vector, and k_1 and k_2 are factors calculated experimentally. The factors in each frame are chosen case dependent.

4) Data Normalizing: Normalizing inputs for an artificial Neural Network is a common pre-processing step. We normalize each axis of the data using the following rule:

$$v_i = v_i / \max(\mathbf{v}) \tag{4}$$

where $\max(.)$ is a function that computes the maximum element of the input vector. We do this rule to all vectors and datasets.

5) Data Combining and Mixing: After conducting all the previous preprocessing steps 1–4, we combine the data from all the datasets into one large dataset. The size of this combined dataset is 437 745 measurement and the full simulation time is 102.0502 s. After combining the data, we mix all the frames randomly to gain the advantages of data shuffling preprocessing (e.g., to reduce variance and overfitting, to ensure fair distribution, and others).

TABLE 2. MLP Parameters

Feature	Description
Learning rate factor	0.03
Momentum factor	0.4
Number of layers	50
Number of neurons at each layer	[15 10 10 10 5]
Maximum number of epochs	5000
Performance function	Mean square error
MSE threshold	10^{-5}
Training/ Validation/ Testing data	70% / 15%/ 15%
Size of data	437745×9

C. Neural Network Architecture

Processing 3-D cloud data points for computer vision is not straightforward. The conventional models that work properly on 2-D RGB pixel images, such as convolutional neural network (CNN) or you only look once (YOLO), fail to accomplish these tasks. This is because LiDAR output 3-D cloud data points are:

1) Sparse;
2) Distributed in a varying density;
3) Unordered;
4) Could not be analyzed isolated [16].

The current utilized approaches to tackle these problems are to either voxelize the 3-D cloud data points into multiple cells (voxels) and apply 3-D convolutions or to analyze the features of individual points, we contribute to the latter approaches as it can yield higher accuracy. However, the latter approaches (e.g., PointNet [17]) suffer from computational complexity due to the complexity in the architecture and the requirement of the whole frame to be fed as an input. As a result, we propose a less complex MLP model that requires only nine features:

1) the point of interest (i.e., $\hat{\mathbf{x}}$, $\hat{\mathbf{y}}$, and $\hat{\mathbf{z}}$);
2) the centroid of the human frame (i.e., μ_x, μ_y, and μ_z);
3) the standard deviation of the human frame (i.e., σ_x, σ_y, and σ_z).

We will show in Section III that it is possible to conduct computationally efficient binary segmentation by learning the relationship between the input parameters. Table 2 summarizes the utilized model parameters.

D. Subsegmentation Processor

Motivated by anthropometric science, humans' upper and lower arms lengths are constrained statistically by length ratio η. We use this ratio to subsegment the output data that has been classified as an arm using the MLP model into either the upper arm or lower arm. More specifically, we first project the data with the outcome of binary-1 per arm into the plane that corresponds to the maximum standard deviation of binary-1 data, more precisely

$$\begin{aligned} &\text{if}(\sigma = \hat{\sigma}_x) \rightarrow \text{project the data into the XZ plane} \\ &\text{if}(\sigma = \hat{\sigma}_y) \rightarrow \text{project the data into the YZ plane} \end{aligned} \tag{5}$$

where $\sigma = \max(\hat{\sigma}_x, \hat{\sigma}_y)$, while $\hat{\sigma}_x$ and $\hat{\sigma}_y$ corresponds to the standard deviation of the data with outcome of binary-1 in a certain frame along the x-axis and the y-axis, respectively. If the whole projection can be curve fitted into a line, we set all the data points that length low than the η threshold into the upper arm; otherwise, we set the rest of the points into the lower arm. If the whole projection cannot be curve fitted into a line, we take the shorter segment as an upper arm while the taller segment is considered as a lower arm.

E. Gesture Identification

After the subsegmentation processor, we propose a simple gesture identification technique that can identify the gesture into one of the following states: starching forward, starching upward, starching downward, starching with 45°, composited starching (i.e., upper arms are stretching with 45° while the lower arms are stretching upward), or neither.[2] Our methodology is based on calculating the slopes L_u and L_l of the upper and lower arm segments, respectively, which have been identified with respect to the projection plane in the subsegmentation step, as is illustrated in Fig. 2(b). In addition, we calculate the mean value of the whole full arm segment $\hat{\mu}_a$ and compare it with the mean value along the z-axis of that particular frame to disengage between the upward and downward starching. Table 3 is used to identify the gesture. In Table 3, D.C. and V.H.V. refer to do not care condition, and very high value, respectively.

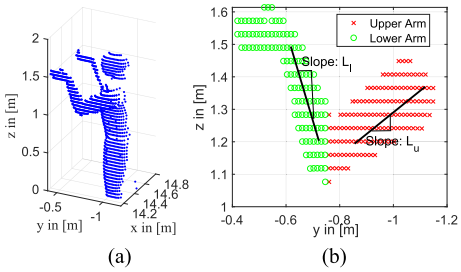

Fig. 2. Computing the slopes of the upper and lower arm segments in (b) from the patient body frame in (a).

TABLE 3. Gesture Identification Lockup Table

L_u	L_l	is $(\hat{\mu}_a > \mu_z)$	gesture
0	0	D.C.	searching forward
V.H.V.	V.H.V.	Yes	searching upward
V.H.V.	V.H.V.	No	starching downward
≈ 1	≈ 1	D.C.	starching with 45°
≈ 1	V.H.V	D.C.	Composited stretching
D.C.	D.C.	D.C.	neither

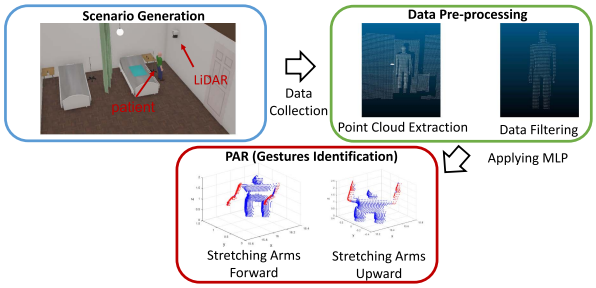

Fig. 3. Summary of the proposed generation and analysis of the LiDAR 3-D point cloud data for PAR.

Fig. 3 summarizes the proposed confidential and accurate patient arms behavior monitoring using a standalone 3-D LiDAR sensor. We first use Webots to construct a model that simulates a PAR scenario of interest. Then, we use a V-LiDAR to capture the raw data that will be preprocessed according to the methodology described in Section II-B. Finally, with the MLP model described in Section II-C and the subsegmentation processor described in Section II-D, the patient arm's status was identified according to Table 3.

III. NUMERICAL RESULTS

In this section, we present the simulation parameters that were used to generate selected numerical results to evaluate the proposed solution. Further, we illustrate the validation criteria that were used to validate the proposed MLP model. We set the data filtering thresholds xth and zth to be $x\text{th} = 16.5$ and $z\text{th} = 0.1$; the data rotation preprocessing step has been applied on dataset-3 with $\theta_r = 51.34°$

[2]Without loss in generalization the same methodology can be applied on more complex scenarios.

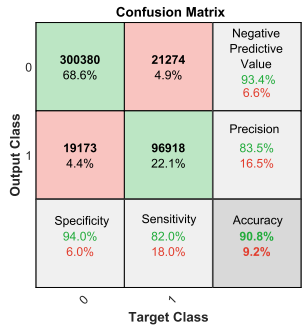

Fig. 4. Confusion matrix of the trained model.

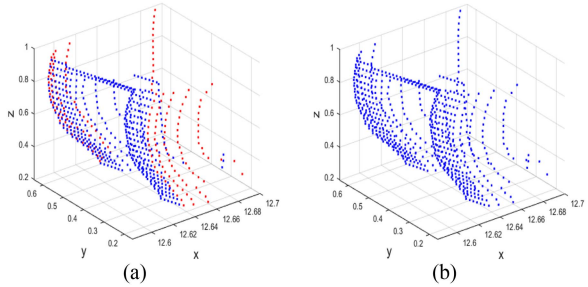

Fig. 5. (a) Wrong labeling. (b) MLP model corrects these errors.

TABLE 4. Performance Metrics of the Trained Model

	Forward	Upward	Downward	45°	Composited	Neither
Data set-1	92.5%	86.0%	87.4%	88.0%	86.5%	88.2%
Data set-2	89.5%	85.5%	83.6%	86.5.5%	85.1%	91.1%
Data set-3	88.3%	81.4%	83.5%	81.4%	81.8%	87.5%
Data set-4	86.4%	80.8%	82.0%	80.8%	79.4%	89.1%

and on dataset-4 with $\theta_r = 38.66°$, and all these parameters have been selected experimentally. For the proposed subsegmentation step, the ratio between the upper and lower arms η can vary between males and females and also between people of different nationalities, and these values can be found in the different anthropometric data reports [18]. We adjusted the patient in Webots simulation to get $\eta = 43.4\%$ so that we match the lengths of the upper and lower arms for male adults in the U.S. We use MATLAB to conduct all the processing of the raw data.

In Fig. 4, we present the confusion matrix of the trained model. The figure shows that the model was able to successfully classify the input data points into arm or nonarm for most of the time with a promising accuracy of 90.8%. The figure also illustrates other performance metrics. We use the k-folds cross-validation to validate the performance. We split the data into threefolds such that each fold is considered as test data and the remaining ones are the training data. The final accuracy result is obtained by computing the average over the three folds iterations. The obtained cross-validation accuracy is 87.4%. Moreover, Fig. 5 shows that some of the frames include data points that belong to the legs of the patient are mistakenly labeled as binary-1. For these cases, the trained MLP model was able to generalize and correct these errors to be binary-0.

In Table 4, we test the proposed gesture identification methodology using the generated datasets. The results represent the accuracy of each class. The results show that dataset-1 and dataset-2 have higher accuracy compared with dataset-3 and dataset-4. This is because the latter datasets include walking with an angle with respect to the V-LiDAR in comparison to the first two datasets that include the patient walking in parallel or perpendicular to the V-LiDAR. On the other hand, the results show that the detection of the forward class has higher accuracy compared with the others, a possible reason for that is outliers that might result in the main segmentation step. These outliers could affect the parameters that are used to identify the gestures (i.e., slopes L_u and L_l and the mean value $\hat{\mu}_a$) that might lead to a nonproper identification.

IV. CONCLUSION

In this letter, we proposed a confidential and accurate patient arms behavior monitoring using a standalone 3-D LiDAR. The data have been generated using computer-based simulations while we proposed an MLP architecture for training followed by subsegmentation step and a gesture identifier. The results showed that the proposed model can achieve a test accuracy of 90.8% and a cross-validation accuracy of 87.4%. This work can be extended in the future to include more practical PAR scenarios with more body parts, experimental data, and more flexible movements.

REFERENCES

[1] "The need for continuous monitoring in today's healthcare system," Managed Healthcare Executive, Mar. 2020. [Online]. Available: https://www.managedhealthcareexecutive.com/view/need-continuous-monitoring-todays-healthcare-system

[2] "Employment projections data for registered nurses," U.S. Bureau of Labor Statistics, Sep. 2022. [Online]. Available: https://www.bls.gov/ooh/healthcare/registered-nurses.htm#tab-6

[3] "Nursing shortage," The American Association of Colleges of Nursing (AACN), Oct. 2022. [Online]. Available: https://www.aacnnursing.org/news-information/fact-sheets/nursing-shortage

[4] "Hourly nurse pay for all 50 states | 2023," Becker's Behavioral Health, Apr. 2023. [Online]. Available: https://www.beckersbehavioralhealth.com/behavioral-health-news/hourly-nurse-pay-for-all-50-states-2023.html

[5] M. Ehatisham-Ul-Haq et al., "Robust human activity recognition using multimodal feature-level fusion," *IEEE Access*, vol. 7, pp. 60736–60751, 2019.

[6] H. A. Imran, "UltaNet: An antithesis neural network for recognizing human activity using inertial sensors signals," *IEEE Sens. Lett.*, vol. 6, no. 1, Jan. 2022, Art. no. 7000304.

[7] R. Zhang and S. Cao, "Real-time human motion behavior detection via CNN using mmWave radar," *IEEE Sens. Lett.*, vol. 3, no. 2, pp. 1–4, Feb. 2019.

[8] O. Rinchi, H. Ghazzai, A. Alsharoa, and Y. Massoud, "LiDAR technology for human activity recognition: Outlooks and challenges," *IEEE Internet Things Mag.*, vol. 6, no. 2, pp. 143–150, Jun. 2023.

[9] M. A. U. Alam, F. Mazzoni, M. M. Rahman, and J. Widberg, "LAMAR: LiDAR based multi-inhabitant activity recognition," in *Proc. 17-th EAI Int. Conf. Mobile Ubiquitous Syst.: Comput., Netw. Serv.*, 2020, pp. 1–9.

[10] J. Roche, V. De-Silva, J. Hook, M. Moencks, and A. Kondoz, "A multimodal data processing system for LiDAR-based human activity recognition," *IEEE Trans. Cybern.*, vol. 52, no. 10, pp. 10027–10040, Oct. 2022.

[11] F. Luo, S. Poslad, and E. Bodanese, "Temporal convolutional networks for multiperson activity recognition using a 2-D LIDAR," *IEEE Internet Things J.*, vol. 7, no. 8, pp. 7432–7442, Aug. 2020.

[12] T. Ueshima, K. Hotta, S. Tokai, and C. Zhang, "Training PointNet for human point cloud segmentation with 3D meshes," in *Proc 15-th Int. Conf. Qual. Control by Artif. Vis.*, 2021, pp. 72–77.

[13] J. P. Espineira, J. Robinson, J. Groenewald, P. H. Chan, and V. Donzella, "Realistic LiDAR with noise model for real-time testing of automated vehicles in a virtual environment," *IEEE Sens. J.*, vol. 21, no. 8, pp. 9919–9926, Apr. 2021.

[14] O. Michel, "Cyberbotics ltd. webots™: Professional mobile robot simulation," *Int. J. Adv. Robot. Syst.*, vol. 1, no. 1, 2004, Art. no. 5.

[15] M. Panwar et al., "Rehab-Net: Deep learning framework for arm movement classification using wearable sensors for stroke rehabilitation," *IEEE. Trans. Biomed. Eng.*, vol. 66, no. 11, pp. 3026–3037, Nov. 2019.

[16] M. Simony, S. Milzy, K. Amendey, and H.-M. Gross, "Complex-YOLO: An Euler-region-proposal for real-time 3D object detection on point clouds," in *Proc. 5-th Eur. Conf. Comput. Vis. Workshops*, 2018.

[17] C. R. Qi, H. Su, K. Mo, and L. J. Guibas, "PointNet: Deep learning on point sets for 3D classification and segmentation," in *Proc. IEEE 30-th Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 652–660.

[18] C. D. Fryar, M. D. Carroll, Q. Gu, J. Afful, and C. L. Ogden, "Anthropometric reference data for children and adults: United States, 2015–2018," *Vital Health Stat.*, vol. 3, no. 46, p. 44, Jan. 2021.