Comparative Analysis of Environment Recognition Systems for Control of Lower-Limb Exoskeletons and Prostheses

Brock Laschowski, *Student Member, IEEE*, William McNally, *Student Member, IEEE*, Alexander Wong, *Senior Member, IEEE*, and John McPhee

***Abstract* – Environment recognition systems can facilitate the predictive control of lower-limb exoskeletons and prostheses by recognizing the oncoming walking environment prior to physical interactions. While many environment recognition systems have been developed using different wearable technology and classification algorithms, their relative operational performances have not been evaluated. Motivated to determine the state-of-the-science and propose future directions for research innovation, we conducted an extensive comparative analysis of the wearable technology, training datasets, and classification algorithms used for vision-based environment recognition. The advantages and drawbacks of different wearable cameras and training datasets were reviewed. Environment recognition systems using pattern recognition, machine learning, and convolutional neural networks for image classification were compared. We evaluated the performances of different deep learning networks using a novel balanced metric called "NetScore", which considers the image classification accuracy, and computational and memory storage requirements. Based on our analysis, future research in environment recognition systems for lower-limb exoskeletons and prostheses should consider developing 1) efficient deep convolutional neural networks for onboard classification, and 2) large-scale open-source datasets for training and comparing image classification algorithms from different researchers.**

I. Introduction

Most lower-limb exoskeletons and prostheses for assistive technology and movement rehabilitation have used biomechatronic control systems with hierarchical architectures [1]–[5]. The high-level controllers predict the upcoming locomotion mode (patient intention) using real-time sensor measurements and machine learning algorithms to automatically switch between locomotion modes. The mid-level controllers translate the predicted locomotion modes into mode-specific reference trajectories using finite-state machines with discrete position or mechanical impedance control parameters. The low-level controllers use conventional algorithms like proportional-integral-derivative (PID) control to calculate the error between the current and desired state trajectories and drive the onboard actuators to minimize the error using reference tracking and closed-loop feedback control [1]–[5].

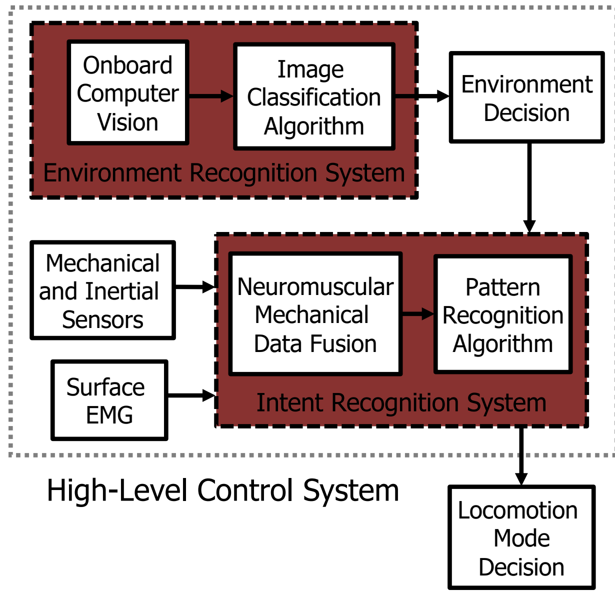

Fig 1. Schematic of an automated locomotion mode recognition system (high-level controller) for lower-limb exoskeletons and prostheses.

Drawing inspiration from human motor control, augmenting neuromuscular-mechanical data with information about the surrounding environment could improve the high-level control performance (Fig. 1). During locomotion, supraspinal levels of the central nervous system acquire state information via ascending pathways from biological sensors (e.g., visual system) and optimize/control the musculoskeletal system and human biomechanics through feedforward efferent commands [3]. However, these control loops are compromised in patients using lower-limb exoskeletons and prostheses owing to deficiencies in human-machine interfaces. Environment sensing using computer vision artificially restores these control loops between the walking mechanics and environment state, therein enabling more accurate and responsive locomotion mode transitioning. Environment recognition could also be used to modulate the actuator reference dynamics (e.g., increasing power

*This research was funded through the Natural Sciences and Engineering Research Council of Canada, John McPhee's Tier I Canada Research Chair in Biomechatronic System Dynamics, and Alexander Wong's Tier II Canada Research Chair in Artificial Intelligence and Medical Imaging.

Brock Laschowski is with the Department of Systems Design Engineering at the University of Waterloo, ON M5S 3G9, Canada (phone: 519-888-4567; email: blaschow@uwaterloo.ca).

William McNally is with the Department of Systems Design Engineering at the University of Waterloo, ON, Canada (email: wmcnally@uwaterloo.ca).

Alexander Wong is with the Department of Systems Design Engineering at University of Waterloo, Canada (email: alexander.wong@uwaterloo.ca).

John McPhee is with the Department of Systems Design Engineering at the University of Waterloo, ON, Canada (email: mcphee@uwaterloo.ca).

to assist with steeper ramps) and optimal path planning (e.g., detecting opportunities for energy regeneration) [6-7].

Huang's research group originally combined environment information with neuromuscular-mechanical data to improve locomotion mode recognition [8]–[12]. Their environments were modelled as prior probabilities using maximum entropy and incorporated into the discriminant functions of linear discriminate analyses (LDA). Equal prior probabilities are typically used for locomotion mode recognition with LDA classification. Huang simulated different walking environments by adjusting the prior probabilities of each class, therein allowing their locomotion mode recognition system to adapt to different environments. Using adaptive prior probabilities significantly outperformed (i.e., 95.5% classification accuracy) the locomotion mode recognition system without environment data (i.e., 90.6% accuracy) [10]. These seminal contributions have since given rise to environment recognition systems using different wearable technology and artificial intelligence algorithms for image classification. However, their relative performances across operational metrics for onboard control have not been evaluated. To determine the state-of-the-science and propose future directions for research advancement, we compared the practical advantages and disadvantages of different wearable technology, training datasets, and image classification algorithms used for recognizing human walking environments.

II. Environment Recognition Systems

A. Experimental Data Collection

Different wearable sensors have been implemented for environment recognition in robotic lower-limb exoskeletons and protheses, including radar and laser rangefinders [10]-[13], RGB cameras [14]–[21], and 3D depth-sensing cameras [22]–[29]. Radar detectors can uniquely measure distances through non-conducting materials and are invariant to lighting conditions and surface texture. Huang prototyped a wearable terrain recognition system using a laser rangefinder and inertial sensor to approximate the geometry of the oncoming walking environment [10]-[12]. Compared to radar and laser rangefinders, camera-based systems provide information about the complete field-of-view and can recognize physical obstacles and terrain changes in peripheral locations. For these reasons, this analysis focused on camera-based environment recognition systems (examples shown in Fig 2).

Most researchers have used RGB cameras for environment recognition [14]–[21]. Each RGB pixel contains light intensity information. Unless multiple cameras are used at different angles, RGB cameras cannot provide distance measurements. Conversely, depth cameras can provide 3D information about the sampled field-of-view. Depth pixels contain both distance measurements and light intensity information. The environment recognition systems from Hargrove [22], [29], Fu [26]–[28], and Varol [23], [24] used depth cameras, which calculate distances by measuring the infrared light time-of-flight. Depth measurement accuracies typically degrade with increasing distance. Consequently, environment recognition systems using depth cameras have had limited capture volumes (i.e., ~1-2 m of maximum range imaging) [22]–[24], [29]. Implementation of depth cameras would also require lower-limb exoskeletons and prostheses to have high computing power microcomputers with low power consumption. The current embedded systems would need significant modifications to support onboard microcomputers capable of processing depth images [24].

Table 1 summarizes the experimental data collection of vision-based environment recognition systems. Most wearable cameras have been mounted on the chest [16], [17], [19], [20], [22], waist [18], [28], [29], and lower-limbs [21], [23], [24], [27]. Limited studies have adopted head-mounted cameras for biomimicry [14], [15]. Unlike lower-limb systems, chest and

Fig 2. Examples of wearable camera systems used for environment sensing. The top and bottom photographs were provided by Dan Simon (Cleveland State University, USA) and Huseyin Atakan Varol (Nazarbayev University, Kazakhstan), respectively.

Table 1. Comparison of experimental data collection for vision-based environment recognition systems.

Reference	Sensor	Position	Dataset	Quality	Classes
[14]	RGB Camera	Head	5	928x620	2
[15]	RGB Camera	Head	40,743	128x128	2
[16]	RGB Camera	Chest	34,254	224x224	3
[17]	RGB Camera	Chest	940,000	1280x720	12
[18]	RGB Camera	Waist	7,284	224x224	3
[19]	RGB Camera	Chest	12,383	32x32 Grayscale	17
[20]	RGB Camera	Chest	3,669	32x32 Grayscale	12
[21]	RGB Camera	Lower Limb	3,992	1080x1920	6
[22]	Depth Camera	Chest	170	80x60	2
[23]	Depth Camera	Lower Limb	22,932	320x240	5
[24]	Depth Camera	Lower Limb	402,403	320x240	5
[26], [27]	Depth Camera	Lower Limb	7,500	224x171	5
[28]	Depth Camera	Waist	4,016	2048 Point Cloud	3
[29]	Depth Camera	Waist	4,000	171x224	5

waist-mounted cameras provide more stable image capturing and allow patients to wear pants and dresses without obstructing the sampled field-of-view. Future research should consider investigating different camera positions for optimal system design. To date, the largest dataset has contained ~940,000 RGB images [17], followed by 402,000 depth images from Varol's research group [24]. Interestingly, the research from Villarreal [15] simultaneously sampled and labelled environment images using a portable keyboard during data collection. Environment recognition systems have focused on classifying level-ground environments [8]–[12], [15], [16], [18]–[20], [23], [24], [26]–[30], incline stairs [8]–[12], [15]–[20], [23], [24], [26]–[30], and decline stairs [8]–[12], [15]–[20], [23], [24], [26]–[30]. Several investigations [11], [17] have incorporated an "others" class to improve classification accuracy when confronted with obscure environments. Classification accuracy (C_a) is defined as the percentage of correct classifications from the total number of classifications:

$$C_a = \frac{\text{Correct Classifications}}{\text{Total Classifications}} \times 100\% \qquad (1)$$

Whereas most environment recognition systems have been limited to controlled indoor environments and/or prearranged walking circuits [10]–[13], [18], [26]–[28], [30], Laschowski and colleagues [16], [17] had participants walk about unknown outdoor and indoor real-world environments; the images were collected throughout the summer, fall, and winter seasons to capture different weathered surfaces like snow, multicolored leaves, and grass. Apart from these studies [16], [17][1], none of the aforementioned datasets were made open-source, making comparisons between different wearable technology and classification algorithms challenging, as subsequently discussed.

B. *Image Classification Algorithms*

Environment recognition systems for lower-limb exoskeletons and prostheses have traditionally used statistical pattern recognition and machine learning for image classification [14], [21]–[25], [29] (Table 2). These algorithms require meticulous hand-engineering to develop feature extractors (e.g., edge detectors) that transform raw images into suitable representations or feature vectors. Manually-extracted features serve as inputs to the image classification (e.g., machine learning or threshold based algorithms). Hough transforms, along with Canny edge detectors or Gabor filtering, have been used for edge detection [14], [19], [29]. Canny edge detection typically includes Sobel operators, which identify gradients across pixels using convolutional filters. Hargrove [14], [22] used standard image pre-

Table 2. Comparison of environment recognition systems using statistical pattern recognition and machine learning for image classification. Note that these algorithms were developed and evaluated using individual private datasets (see Table 1).

Reference	Feature Extractor and/or Classifier	Hardware	Classification Accuracy	Computation Time
[14]	Hough Transform with Gabor Filter or Canny Edge Detector	Intel Core i5	Not Specified	8 seconds
[19]	Gabor Barcodes and Hamming Distances	Intel Core i7 (3.60GHz)	88.5%	0.15 seconds
[21]	SURF Features and Bag of Words	Intel Core i7-2600 CPU (3.40GHz)	86%	Not Specified
[22]	Heuristic Thresholding and Edge Detector	Intel Core i5	98.8%	0.2 seconds
[23]	Support Vector Machine	Intel Core i7-2640M (2.8GHz)	99.0%	14.9 ms
[24]	Cubic Kernel Support Vector Machine	Intel Core i7-2640M (2.8GHz)	94.1%	14.9 ms
[29]	Regions-of-Interest and Linear Discriminate Analysis	Intel Core i7-8750H (2.2GHz)	Not Specified	Not Specified

[1] https://ieee-dataport.org/open-access/exonet-database-wearable-camera-images-human-locomotion-environments

processing and rule-based thresholds to recognize convex and concave edges for stair recognition. While their system performance was satisfactory, the computations were time-consuming (i.e., 8 seconds/frame) and the system was evaluated using only five sampled images [14].

Varol's research group leveraged support vector machines (SVMs) for classifying depth images of human locomotion environments [23], [24]. Support vector machines are supervised machine learning algorithms that map extracted features into a high-dimensional feature space and separate samples into different classes by way of constructing optimal hyperplanes with maximum margins. Varol compared different dimensionality reductions and SVM model designs. Their environment recognition system achieved ~94.1% image classification accuracy using a cubic kernel SVM and no dimension reduction [24]. Though SVM models are effective in high-dimensional space and offer good generalization (i.e., robustness to overfitting), these algorithms require manual selection of kernel functions and statistical features from regions-of-interest [23] [24]. Deep learning replaces manually-extracted features with multilayer networks that automatically learn optimal image features from training data. Convolutional neural networks (CNNs) typically outperform SVMs for image classification [31] and thus have become the standard in computer vision.

The latest cohort of environment recognition systems have used CNNs for classification [15], [16], [18], [20], [26]–[28] (Fig. 3). Multiple convolution and pooling layers are stacked with decreasing resolutions and increasing number of feature maps. The convolutional layers perform convolutions between the input and convolutional filters. The resulting feature maps are passed through nonlinear activation functions. The pooling layers spatially downsample the feature maps by aggregating neighboring elements using maximum values. CNNs conclude with fully-connected layer(s) and softmax, which estimate the probability distribution of each class. Environment recognition systems have typically involved supervised learning, wherein the differences between the predicted and labelled class scores are calculated, and the network parameters are updated to minimize the error through backpropagation. Table 3 compares the performances of different environment recognition systems using CNNs for image classification. The neural network from Simon's group [18] achieved the highest image classification accuracy (i.e., ~99%) albeit included the largest computational and memory storage requirements, with 27 million parameters and 7.7 billion multiply-accumulate operations (MACs).

Few environment recognition systems [18], [23], [24], [26] have included temporal information to minimize the occurrences of periodic misclassifications. Fu's research group [26] compared recurrent neural networks (RNNs) and hidden Markov models for environment recognition using sequential data. RNNs process sequences of inputs while maintaining internal hidden state vectors that contain time history information and that feedback on themselves recurrently. Although RNNs are designed to learn long-term dependencies, theoretical and empirical evidence has suggested that they struggle with storing sequential information over extended periods [31]. Therefore, recurrent neural networks are often supplemented with explicit memory modules such as long short-term memory (LSTM) networks. Fu [26] showed that hidden Markov models outperformed (i.e., ~97.3% accuracy) both recurrent neural networks (i.e., ~96.9% accuracy) and LSTM networks (i.e., ~96.8% accuracy) for human locomotion environment recognition.

III. Discussion

Environment recognition systems can improve the control of robotic lower-limb exoskeletons and prostheses by predicting the upcoming locomotion mode prior to execution, therein enabling more accurate and responsive transitioning between locomotion modes. These systems have used different wearable technology and classification algorithms for environment recognition. However, their relative operational performances for onboard real-time control (e.g., inference times, classification accuracies, and computational and memory requirements) have not been previously considered, which motivated the current analysis. We compared the practical advantages and drawbacks of different wearable technology, training datasets, and classification algorithms used for recognizing human walking

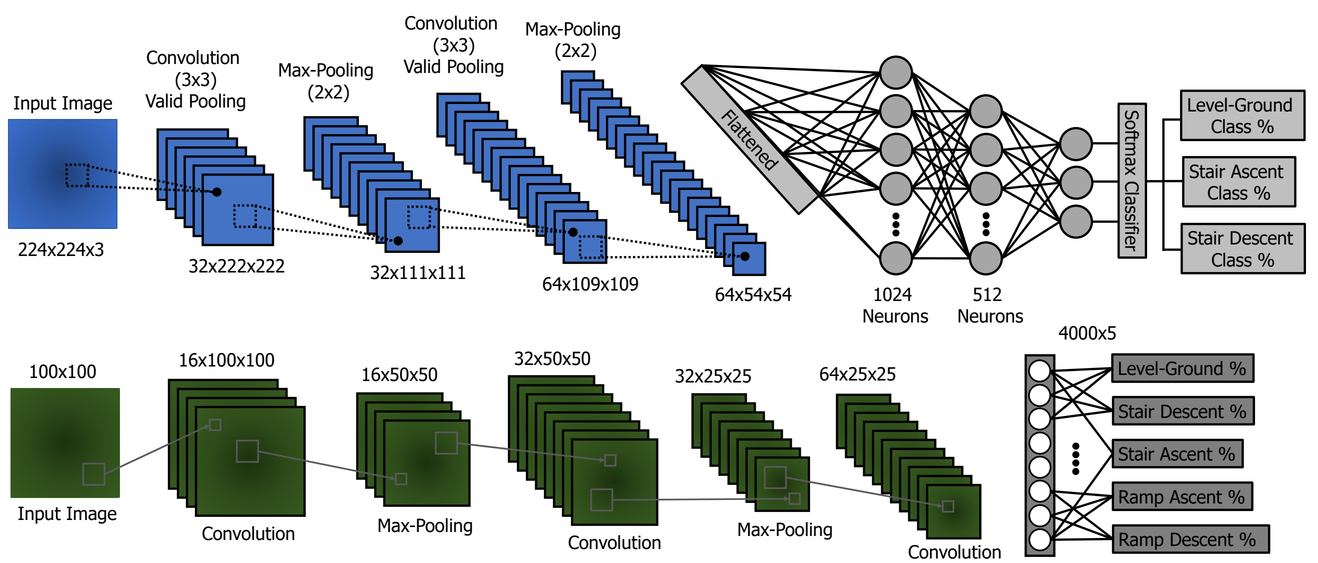

Fig. 3 Examples of convolutional neural networks used for environment recognition. The top and bottom schematics were recreated from graphics provided by Dan Simon (Cleveland State University, USA) and Chenglong Fu (Southern University of Science and Technology, China), respectively.

Table 3. Comparison of environment recognition systems using convolutional neural networks for image classification. Note that these networks were trained and evaluated on individual private datasets (see Table 1).

Reference	Numerical Operations (MACs)	Network Parameters	Hardware	Classification Accuracy	Inference Time	NetScore [32]
[15]	0.001125 billion	1.13 million	NVIDIA Geforce GTX 965M	90%	5.5 ms	107.1
[16]	1.285 billion	4.73 million	NVIDIA TITAN Xp	94.9%	0.85 ms	71.2
[18]	7.7 billion	27 million	NVIDIA Titan X	99.6%	50 ms	56.8
[20]	0.00199 billion	0.0388 million	Intel Core i7 (3.60GHz)	92%	Not Specified	119.7
[26]	0.013 billion	0.223 million	NVIDIA GeForce GTX 1050 Ti	96.8%	3.1 ms	104.8
[27]	0.013 billion	0.223 million	NVIDIA Quadro P400	98.9%	3 ms	105.2
[28]	0.0215 billion	0.05 million	NVIDIA GeForce GTX 1050 Ti	98%	2 ms	109.3

environments. Several researchers have explored using laser rangefinders [10]–[12] and radar detector [13] for environment sensing. However, most systems have involved RGB cameras [14]–[21] and depth cameras [22]–[28] mounted on the chest, waist, and lower-limbs (see Table 1). Environment recognition systems have used statistical pattern recognition and machine learning for image classification [14], [19], [21]–[25], [29] (Table 2). Unlike these algorithms that require manual feature engineering, convolutional neural networks automatically learn optimal image features from training data. This design feature, along with achieving higher image classification accuracy [31], could explain the growing implementation of deep learning networks for human locomotion environment classification [15], [16], [18], [26]–[28].

Environment recognition systems have focused on improving image classification accuracy, leading to more accurate yet inefficient algorithms with greater computational and memory storage requirements (see Table 3). These design features can be problematic for mobile and embedded vision applications (e.g., lower-limb prostheses) with limited operating resources. While the newly-developed graphics processing units (GPUs) have many core processors with high memory bandwidth [31], the current microcomputers used in lower-limb prostheses and exoskeletons cannot effectively support the architectural and computational complexities of traditional convolutional neural networks [24]. To enable onboard deep learning for real-time control, the ideal image classification algorithm would achieve high accuracy with minimal parameters, numerical operations, and inference time. Motivated by these design principles, we evaluated the performances of different convolutional neural networks ($\mathcal{N}$) for human locomotion environment recognition using a novel balanced metric called NetScore [32]:

$$\Omega(\mathcal{N}) = 20\log\left(\frac{a(\mathcal{N})^{\alpha}}{p(\mathcal{N})^{\beta}m(\mathcal{N})^{\gamma}}\right) \tag{2}$$

where $a(\mathcal{N})$ is the classification accuracy, $p(\mathcal{N})$ is the network parameters, $m(\mathcal{N})$ is the numerical operations during inference, and α, β, and γ are coefficients that control the effects of accuracy, and the architectural and computational complexities, respectively. Similar to Wong [32], the NetScore coefficients were $\{\alpha = 2, \beta = 0.5, \gamma = 0.5\}$ to better emphasize classification accuracy, while partially considering the network parameters and operations. The number of operations and parameters are representative of the computational and memory requirements, respectively. The resulting NetScores are presented in Table 3. Despite having the highest image classification accuracy, the neural network from Simon's research group [18] achieved the lowest NetScore (i.e., 56.8) due to the disproportionately large number of parameters and operations. Conversely, the network from Tung's research group [20] attained the highest NetScore with approximately 119.7. It should be noted that these networks were trained and evaluated on different private datasets.

These NetScores demonstrate a fundamental limitation in comparing the performances of convolutional neural networks trained using different datasets such that the smaller database from Tung [20] allowed a smaller classifier to obtain relatively high accuracy. Researchers have each individually collected experimental data for training their environment recognition algorithms. These repetitive measurements are inefficient for the biomechatronics field and individual private datasets have prevented comparisons between classification algorithms from different researchers. The absence of an open-source database of human locomotion environment images has impeded the development of environment recognition systems for lower-limb exoskeletons and prostheses. Drawing inspiration from ImageNet [33], we recently published the "ExoNet" database through IEEE DataPort [17]. ExoNet represents the first open-source database of high-resolution wearable camera images of human locomotion environments. Aside from being the only open-source database, the large scale and diversity of ExoNet distinguishes itself from previous environment recognition systems, with more than 940,000 human-annotated images and 12 classes (Table 1) [17]. With ongoing advances in wearable technology and artificial intelligence, larger and more diverse training datasets are needed to develop next-generation image classification algorithms for environment-aware controls.

Our comparative analysis concluded that future research in environment recognition systems for lower-limb exoskeletons and prostheses should consider developing 1) efficient deep convolutional neural networks for onboard classification with limited operational resources, and 2) large-scale open-source datasets for training and comparing classification algorithms from different researchers.

References

[1] N. E. Krausz and L. J. Hargrove, "A Survey of Teleceptive Sensing for Wearable Assistive Robotic Devices," *Sensors*, vol. 19, no. 23, p. 5238, Nov. 2019, doi: 10.3390/s19235238.

[2] A. J. Young and D. P. Ferris, “State of the Art and Future Directions for Lower Limb Robotic Exoskeletons,” *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 25, no. 2, pp. 171–182, Feb. 2017, doi: 10.1109/TNSRE.2016.2521160.

[3] M. R. Tucker et al., “Control Strategies for Active Lower Extremity Prosthetics and Orthotics: A Review,” *J. NeuroEngineering Rehabil.*, vol. 12, no. 1, 2015, doi: 10.1186/1743-0003-12-1.

[4] B. Laschowski and J. Andrysek, “Electromechanical Design of Robotic Transfemoral Prostheses,” *in ASME International Design Engineering Technical Conferences and Computers and Information in Engineering Conference*, Quebec City, Quebec, Canada, 2018, p. V05AT07A054, doi: 10.1115/DETC2018-85234.

[5] K. Zhang, C. W. De Silva, and C. Fu, “Sensor Fusion for Predictive Control of Human-Prosthesis-Environment Dynamics in Assistive Walking: A Survey,” Mar. 2019, arXiv:1903.07674.

[6] B. Laschowski, J. McPhee, and J. Andrysek, “Lower-Limb Prostheses and Exoskeletons with Energy Regeneration: Mechatronic Design and Optimization Review,” *ASME J. Mech. Robot*., vol. 11, no. 4, p. 040801, Aug. 2019, doi: 10.1115/1.4043460.

[7] B. Laschowski, R. S. Razavian, and J. McPhee, “Simulation of Stand-to-Sit Biomechanics for Design of Lower-Limb Exoskeletons and Prostheses with Energy Regeneration,” *bioRxiv*, 2020, doi: 10.1101/801258.

[8] L. Du, F. Zhang, M. Liu, and H. Huang, “Toward Design of an Environment-Aware Adaptive Locomotion-Mode-Recognition System,” *IEEE Trans. Biomed. Eng.*, vol. 59, no. 10, pp. 2716–2725, Oct. 2012, doi: 10.1109/TBME.2012.2208641.

[9] H. Huang, Z. Dou, F. Zheng, and M. J. Nunnery, “Improving the Performance of a Neural-Machine Interface for Artificial Legs Using Prior Knowledge of Walking Environment,” *in Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, Boston, MA, USA, 2011, pp. 4255–4258, doi: 10.1109/IEMBS.2011.6091056.

[10] D. Wang, L. Du, and H. Huang, “Terrain Recognition Improves the Performance of Neural-Machine Interface for Locomotion Mode Recognition,” *in IEEE International Conference on Computing, Networking and Communications (ICNC)*, San Diego, CA, USA, 2013, pp. 87–91, doi: 10.1109/ICCNC.2013.6504059.

[11] M. Liu, D. Wang, and H. Huang, “Development of an Environment-Aware Locomotion Mode Recognition System for Powered Lower Limb Prostheses,” *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 24, no. 4, pp. 434–443, Apr. 2016, doi: 10.1109/TNSRE.2015.2420539.

[12] F. Zhang, Z. Fang, M. Liu, and H. Huang, “Preliminary Design of a Terrain Recognition System,” *in Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, Boston, MA, USA, 2011, pp. 5452–5455, doi: 10.1109/IEMBS.2011.6091391.

[13] B. Kleiner, N. Ziegenspeck, R. Stolyarov, H. Herr, U. Schneider, and A. Verl, “A Radar-Based Terrain Mapping Approach for Stair Detection Towards Enhanced Prosthetic Foot Control,” *in IEEE International Conference on Biomedical Robotics and Biomechatronics (Biorob)*, Enschede, Netherlands, 2018, pp. 105–110, doi: 10.1109/BIOROB.2018.8487722.

[14] N. E. Krausz and L. J. Hargrove, “Recognition of Ascending Stairs from 2D Images for Control of Powered Lower Limb Prostheses,” *in International IEEE/EMBS Conference on Neural Engineering (NER)*, Montpellier, France, 2015, pp. 615–618, doi: 10.1109/NER.2015.7146698.

[15] L. Novo-Torres, J.-P. Ramirez-Paredes, and D. J. Villarreal, “Obstacle Recognition using Computer Vision and Convolutional Neural Networks for Powered Prosthetic Leg Applications,” *in Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, Berlin, Germany, 2019, pp. 3360–3363, doi: 10.1109/EMBC.2019.8857420.

[16] B. Laschowski, W. McNally, A. Wong, and J. McPhee, “Preliminary Design of an Environment Recognition System for Controlling Robotic Lower-Limb Prostheses and Exoskeletons,” *in IEEE International Conference on Rehabilitation Robotics (ICORR)*, Toronto, ON, Canada, 2019, pp. 868–873, doi: 10.1109/ICORR.2019.8779540.

[17] B. Laschowski, W. McNally, A. Wong, and J. McPhee, “ExoNet Database: Wearable Camera Images of Human Locomotion Environments,” *bioRxiv*, Oct. 2020, doi: 10.1101/2020.10.23.352054.

[18] G. Khademi and D. Simon, “Convolutional Neural Networks for Environmentally Aware Locomotion Mode Recognition of Lower-Limb Amputees,” *in ASME Dynamic Systems and Control Conference*, Park City, Utah, USA, 2019, p. 11, doi: 10.1115/DSCC2019-9180.

[19] M. Nouredanesh, A. McCormick, S. Kukreja, and J. Y. Tung, “Wearable Vision Detection of Environmental Fall Risk Using Gabor Barcodes,” *in IEEE International Conference on Biomedical Robotics and Biomechatronics (BioRob)*, Singapore, Singapore, 2016, pp. 956–956, doi: 10.1109/BIOROB.2016.7523751.

[20] M. Nouredanesh, A. McCormick, S. Kukreja, and J. Y. Tung, “Wearable Vision Detection of Environmental Fall Risks Using Convolutional Neural Networks,” Nov. 2016, arXiv:1611.00684.

[21] J. P. Diaz, R. L. da Silva, B. Zhong, H. H. Huang, and E. Lobaton, “Visual Terrain Identification and Surface Inclination Estimation for Improving Human Locomotion with a Lower-Limb Prosthetic,” *in Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, Honolulu, HI, 2018, pp. 1817–1820, doi: 10.1109/EMBC.2018.8512614.

[22] N. E. Krausz, T. Lenzi, and L. J. Hargrove, “Depth Sensing for Improved Control of Lower Limb Prostheses,” *IEEE Trans. Biomed. Eng.*, vol. 62, no. 11, pp. 2576–2587, Nov. 2015, doi: 10.1109/TBME.2015.2448457.

[23] H. A. Varol and Y. Massalin, “A Feasibility Study of Depth Image Based Intent Recognition for Lower Limb Prostheses,” *in Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, Orlando, FL, USA, 2016, pp. 5055–5058, doi: 10.1109/EMBC.2016.7591863.

[24] Y. Massalin, M. Abdrakhmanova, and H. A. Varol, “User-Independent Intent Recognition for Lower Limb Prostheses Using Depth Sensing,” *IEEE Trans. Biomed. Eng.*, vol. 65, no. 8, pp. 1759–1770, Aug. 2018, doi: 10.1109/TBME.2017.2776157.

[25] B. H. Hu, N. E. Krausz, and L. J. Hargrove, “A Novel Method for Bilateral Gait Segmentation Using a Single Thigh-Mounted Depth Sensor and IMU,” *in IEEE International Conference on Biomedical Robotics and Biomechatronics (Biorob)*, Enschede, Netherlands, 2018, pp. 807–812, doi: 10.1109/BIOROB.2018.8487806.

[26] K. Zhang, W. Zhang, W. Xiao, H. Liu, C. W. De Silva, and C. Fu, “Sequential Decision Fusion for Environmental Classification in Assistive Walking,” *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 27, no. 9, pp. 1780–1790, Sep. 2019, doi: 10.1109/TNSRE.2019.2935765.

[27] K. Zhang et al., “Environmental Features Recognition for Lower Limb Prostheses Toward Predictive Walking,” *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 27, no. 3, pp. 465–476, Mar. 2019, doi: 10.1109/TNSRE.2019.2895221.

[28] K. Zhang, J. Wang, and C. Fu, “Directional PointNet: 3D Environmental Classification for Wearable Robotics,” 2019, arXiv:1903.06846.

[29] N. E. Krausz, B. H. Hu, and L. J. Hargrove, “Subject- and Environment-Based Sensor Variability for Wearable Lower-Limb Assistive Devices,” *Sensors*, vol. 19, no. 22, p. 4887, Nov. 2019, doi: 10.3390/s19224887.

[30] V. Rai and E. Rombokas, “Evaluation of a Visual Localization System for Environment Awareness in Assistive Devices,” *in Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, Honolulu, HI, 2018, pp. 5135–5141, doi: 10.1109/EMBC.2018.8513442.

[31] Y. LeCun, Y. Bengio, and G. Hinton, “Deep Learning,” *Nature*, vol. 521, no. 7553, pp. 436–444, May 2015, doi: 10.1038/nature14539.

[32] A. Wong, “NetScore: Towards Universal Metrics for Large-scale Performance Analysis of Deep Neural Networks for Practical On-Device Edge Usage,” Aug. 2018, arXiv:1806.05512.

[33] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, “ImageNet: A Large-Scale Hierarchical Image Database,” *in IEEE Conference on Computer Vision and Pattern Recognition*, Miami, FL, USA, 2009, pp. 248–255, doi: 10.1109/CVPR.2009.5206848.