2021 43rd Annual International Conference of the
IEEE Engineering in Medicine & Biology Society (EMBC)
Oct 31 - Nov 4, 2021. Virtual Conference

Offline and Real-Time Implementation of a Terrain Classification Pipeline for Pushrim-Activated Power-Assisted Wheelchairs*

Mahsa Khalili, Kevin Ta, H.F. Machiel Van der Loos, Senior Member, IEEE, Jaimie F. Borisoff

Abstract— **Pushrim-activated power-assisted wheelchairs (PAPAWs) are assistive technologies that provide propulsion assist to wheelchair users and enable access to various indoor and outdoor terrains. Therefore, it is beneficial to use PAPAW controllers that adapt to different terrain conditions. To achieve this objective, terrain classification techniques can be used as an integral part of the control architecture. Previously, the feasibility of using learning-based terrain classification models was investigated for offline applications. In this paper, we examine the effects of three model parameters (i.e., feature characteristics, terrain types, and the length of data segments) on offline and real-time classification accuracy. Our findings revealed that Random Forest classifiers are computationally efficient and can be used effectively for real-time terrain classification. These classifiers have the highest performance accuracy when used with a combination of time- and frequency-domain features. Additionally, we found that increasing the number of data points used for terrain estimation improves the prediction accuracy. Finally, our results revealed that classification accuracy can be improved by considering terrains with similar characteristics under one umbrella group. These findings can contribute to the development of real-time adaptive controllers that enhance PAPAW usability on different terrains.**

I. Introduction

Pushrim-activated power-assisted wheelchairs (PAPAWs) are mobility assistive devices that can provide on-demand propulsion assistance to their users. PAPAW use can improve accessibility of environments that are commonly inaccessible or difficult to access for manual wheelchair users, such as uneven terrains or steep slopes [1–3]. Although commercially available PAPAWs can be used on a variety of indoor and outdoor terrains, their controllers are mainly insensitive to environmental changes (e.g., slipperiness or roughness of the terrain). Therefore, the lack of an adaptive control framework may affect PAPAWs' usability or safety on different terrains [2]. It is worth noting that similar concerns related to terrain-dependent driving performance have been reported among power wheelchair users (e.g., getting stuck on gravel) [4]. To address this limitation, terrain classification frameworks can be used in conjunction with adaptive controllers to improve PAPAW operation by adjusting torque, velocity, or acceleration/deceleration in different environments.

Terrain classification frameworks have been extensively studied in the context of mobile robot and planetary rover navigation [5–7]. Kinematic/vibration-, vision-, and acoustic-based classifiers are among the most commonly used terrain classification models. Although visual and acoustic features can provide useful information about terrain characteristics, they are also affected by other ambient changes, such as lighting and noise. Hence, kinematic-based measurements (e.g., time/frequency-domain features) can provide more robust terrain-specific characteristics for classification purposes. It should be noted that using multimodal sensory measurements (i.e., fusing vision, acoustic, and/or vibration data) could contribute to more reliable predictions compared to using individual modalities [8]. Machine learning algorithms are commonly used for terrain classification purposes [7,9,10]. Classifiers such as Random Forest (RF) and Support Vector Machine (SVM) are commonly used in conjunction with feature engineering methods to improve prediction performance. Time- and frequency-domain statistical measures, alongside power spectral density (PSD) and fast Fourier transform (FFT) characteristics, are among the most commonly used feature extraction methods [11].

In our previous work [12], we examined the feasibility of using kinematic data and machine learning algorithms to develop a terrain classification framework for PAPAWs. We used gyroscope and acceleration data that were collected on 4 outdoor terrains consisting of grass, gravel, asphalt, and sidewalk as well as 3 indoor terrains consisting of concrete, linoleum, and carpet. Although an overall prediction accuracy of 80.2% was achieved for a 7-class RF classifier, the rate of correct outdoor terrain classification was notably higher compared to indoor terrain predictions. The best classifier consisted of a combination of 25 time-, frequency-, PSD-, and FFT-based features, with time and frequency features having the highest contributions to terrain classification. Although the proposed framework was successfully tested offline, the feasibility of implementing the proposed model was not evaluated for real-time classification.

This paper, which presents a methodological extension of our previous study, focuses on (1) examining the effects of window size, feature subsets, and re-grouping terrains on classification performance; and (2) investigating the feasibility of real-time implementation of the proposed classification models. The rest of this paper is organized as follows. Offline and real-time classification pipelines are presented in Section II. The findings regarding the effects of window size, feature subsets, and terrain types on classification accuracy are presented and discussed in Sections III and IV, respectively. Finally, the main contributions of this work and the implications of our findings are presented in Section V.

II. Methods

A. Test setup and experimental protocol

Triaxial gyroscope (ω_x, ω_y, ω_z) and acceleration (a_x, a_y, a_z) data were collected using a wheelchair-mounted (Fig. 1)

*Research supported by the Natural Sciences and Engineering Research Council of Canada (NSERC).

M. Khalili is with the School of Biomedical Engineering and H.F.M. Van der Loos is with the Department of Mechanical Engineering at the University of British Columbia, Vancouver, Canada. K. Ta is with ETH Zürich, Switzerland. J.F. Borisoff is with the British Columbia Institute of Technology (BCIT). Corresponding author: mahsa.khalili@alumni.ubc.ca.

978-1-7281-1178-0/21/$31.00 ©2021 IEEE

MPU-6050 inertial measurement unit (IMU). IMU measurements were sampled at 300 Hz and transferred to a laptop in real-time. Study participants included 1 skilled and 3 experienced wheelchair users (1 Female and 3 Male; average weight: 136±15 lbs.). Participants performed a set of pre-defined wheelchair activities (e.g., moving straight forward and turning) at a self-selected speed on 7 terrains of grass, gravel, asphalt, sidewalk, concrete, linoleum, and low-pile carpet. These experiments were performed with the Twion™ powered wheels (Alber GmbH, Germany). More information about the test protocol can be found in our previous paper [12].

B. Offline model training and testing

Data preprocessing: First, gravitational acceleration was subtracted from acceleration measurements. Next, all kinematic measurements were filtered with a fourth-order Butterworth filter with a 20 Hz low-pass cut-off frequency. Time-series measurements were split into short overlapping segments that consisted of 512 or 1024 data points with 50% overlap. To prevent frequency leakage for PSD analysis, these segments were windowed with a Hanning function. Following PSD analysis, data associated with frequencies greater than 25 Hz were removed from all datasets. Finally, four label sets associated with different combinations of indoor and outdoor terrains were added to the datasets (Table I.)

Feature extraction: The following features were extracted from kinematic measurements: (a) “Time” features, including mean, standard deviation (std), norm, maximum (max), minimum (min), root mean square (rms), zero crossing rate (zcr) of each time window; and (b) “Frequency” features, including root mean square frequency (rmsf), frequency centre (fc), root variance frequency (rvf) of PSD signals. These features were calculated for all IMU measurements (i.e., X, Y, and Z axes of gyroscope and accelerometer). The “Time” and “Frequency” features consisted of 42- and 18-dimension data frames, respectively. Pairwise Pearson correlation coefficients were calculated between these features and each label set.

Classification pipeline: The train/validation dataset included 2/3 of the experimental measurements that were selected through a semi-random process to ensure data homogeneity (i.e., including equal subsets from all participants, all terrains, and all maneuvers). The rest of the measurements were used for testing and evaluation. The classification pipeline consisted of three steps, including (1) standardization: using standard scaler to remove the mean and to scale to unit variance; (2) feature selection: using a sequential feature selection algorithm to identify the most relevant features for classification; and (3) model training and validation: training RF classifiers with a grid-search process for hyperparameter tuning and with five-fold cross-validation.

Evaluation: The test dataset was normalized using the scalar values from the training step and transformed to include the top selected features. The output of the classification pipeline (i.e., the optimal trained model) was used to evaluate the classification performance for the isolated test dataset. Each classifier’s performance was evaluated using prediction accuracy and confusion matrix. We used Python 3.7 and scikit-learn for data preprocessing, training/validating/testing the classifiers. Additionally, we used the Cochran's Q test and McNemar's test from the Mlxtend Python library (Raschka, S. 2016) to compare the performance of different classifiers.

Figure 1. Left: IMU attached to the wheelchair frame; Right: the coordinate system of the IMU

TABLE I. LABEL SETS USED FOR CLASSIFICATION

Label set	List of labels
8-Class	grass, gravel, asphalt, sidewalk, concrete, linoleum, carpet, no-motion[a]
6-Class	grass, gravel, asphalt, sidewalk, indoor, no-motion
5-Class	grass, gravel, asphalt-sidewalk[b], indoor, no-motion
3-Class	outdoor, indoor, no-motion

a. “no-motion” data is associated with the stationary state of the wheelchair recorded at the start/end of data acquisition. These data are significant for real-time terrain classification applications.

b. Asphalt & sidewalk datapoints are grouped together as a single terrain, labelled asphalt-sidewalk.

C. Real-time model implementation and testing

A simulated acquisition process was used to stream data through the real-time pipeline at 300 Hz, mimicking the hardware acquisition pipeline. These data were collected in a rolling window of the most recent N (512 or 1024) kinematic data points. A separate process took the data window and ran through the same preprocessing, feature extraction, and classification steps as the offline process, except over the single updating data window. The real-time testing pipeline continuously classified the most recently captured information in the data window. Subsequent windows used in classification contained shifted overlaps with previous windows and newly acquired data. In contrast to the offline classification, the parallel acquisition and classification processes introduced asynchronicity for the shift length across each iteration. Therefore, with different subsets of data, the offline and online accuracy may exhibit marginally different values. The real-time performance was evaluated in terms of prediction accuracy and computational time.

III. RESULTS

A total of 10,517 and 5,197 data segments were used for 512- and 1024-window datasets, respectively. The breakdown of different labels for the 512-window dataset is as follows {no-motion: 762; concrete: 1,380; carpet: 1,352; linoleum: 1,302; asphalt: 1,357; sidewalk: 1,429; grass: 1,503; gravel: 1,432}.

A. Offline classification: training and testing models

We trained a total of 24 classifiers to examine the effects of window size (512 vs. 1024 data points in each window), available feature subsets (“Frequency”, “Time”, “Combined”), and terrain types (3-Class, 5-Class, 6-Class, 8-Class) on classification performance (Fig. 2). Our results revealed that higher performance accuracy is achieved by increasing the data points in each time window. Additionally, our findings suggest that classification performance decreases by increasing the number of classes, meaning that differentiating indoor vs. outdoor terrains can be done with higher accuracy compared to differentiating individual indoor and outdoor terrains. We also found that using a combination of time and frequency features contribute to higher classification accuracy compared to using only time or only frequency features. In all cases, regardless of the window size or the number of classes, significantly higher accuracy is achieved when using "Combined" features compared to using "Frequency" features. However, the difference between the classifiers using "Combined" or "Time" features are not statistically significant in all cases (e.g., Fig. 2: 3-Class Combined-512 vs. 3-Class Time-512).

We performed further analysis to examine the effects of number and types of features as well as the number of classes on classification performance (Fig. 3). These findings provide further evidence that classifiers based on "Combined" features have consistently higher performance accuracy regardless of the number of features, number of classes, or the window size. It is worth noting that we found moderate or strong correlation between the majority of the top selected features and associated label sets. Examples of these values for a "3-Class Combined-1024" classifier are presented in Table II.

Confusion matrices were used to examine the classification accuracy for different terrain types and to identify misclassified labels (Fig. 4). These figures show which terrains are more difficult to distinguish. As an example, linoleum, concrete, and carpet in 8-Class models were often misclassified due to similarity in their surface characteristics. Visualization of the predicted labels for one of the experimental maneuvers on sidewalk is shown in Fig. 5. Confusion matrices also display the effects of regrouping similar terrains (e.g., 8-Class vs. 6-Class) on the overall classification performance.

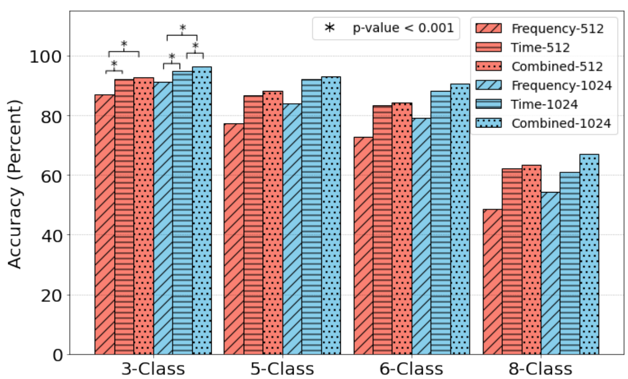

Figure 2. Classification performance (to avoid excessive annotations, significant differences are shown for 3-Class models only)

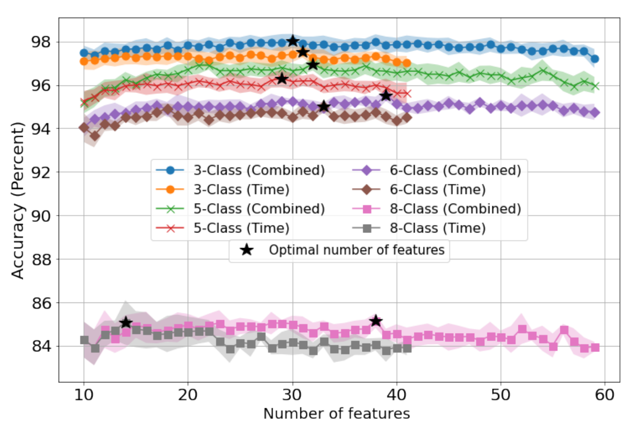

Figure 3. Classification accuracy vs. number of selected features for 1024-window datasets (due to notably poor performances, classifiers based on "Frequency" features are not shown here)

TABLE II. TOP SELECTED FEATURES (3-CLASS COMBINED-1024)

Features with correlation coefficient greater than 0.5			Features with correlation coefficient between 0.3 and 0.5		
norm ω_x	zcr ω_y	min a_y	zcr ω_x	rvf ω_y	norm a_y
rms ω_y	min a_x	rmsf a_y	rvf ω_x	zcr ω_z	rvf a_z
max ω_y	std a_x	zcr a_z	fc ω_x	fc a_x	-
std ω_y	rms a_x	std a_z	rmsf ω_y	rms a_y	-

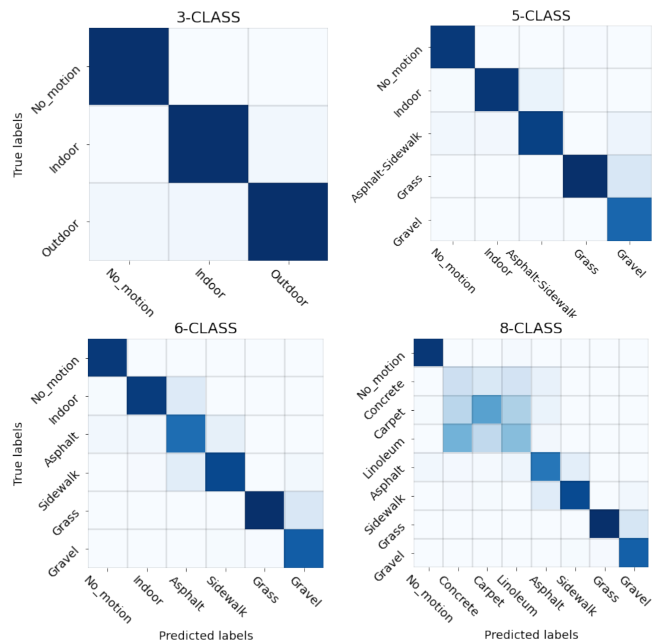

Figure 4. Confusion matrices for different label sets (1024-window)

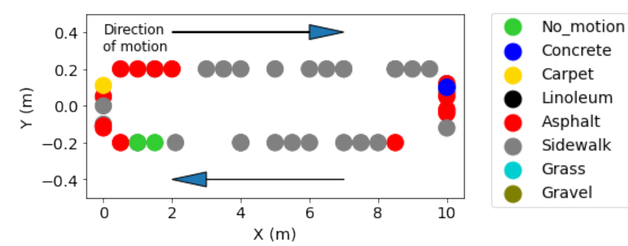

Figure 5. Predicted labels for a single maneuver on "sidewalk" for "8-Class Combined-1024" classifier (this maneuver consisted of 10 m straight propulsion followed by a 180° turn and repeated 8 times).

B. *Real-time terrain classification*

We found that real-time classification models had similar prediction accuracy to the offline classifiers. A summary of real-time classification performance of 8 models is presented in Fig. 6. Classification time for these classifiers for a single sample was measured on a computer with an Intel® Core™ i7-6560U CPU and 8 GB RAM. While we hypothesized using "Combined" features would increase the classification time, we found inconclusive evidence in the implementation of the acquisition-classification pipeline. Similarly, we did not find any direct relation between the prediction time and the length of window segments.

IV. DISCUSSION

The prediction accuracy of 24 terrain classification models was evaluated offline and in real-time, and similar performance was observed for both implementations. Larger

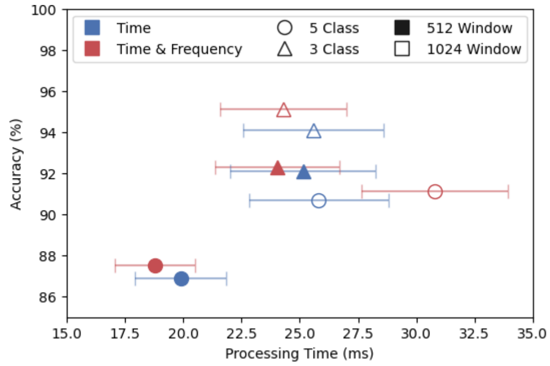

Figure 6. Real-time terrain classification performance

window segments were associated with higher prediction accuracy. Similar findings were reported in the context of terrain classification for mobile robots [7]. However, the trade-off comes in the form of classification latency, where the larger window sizes correspond to greater temporal history. During transitions between terrains, a classification with a longer time window would be slower at identifying the new terrain. This identification lag could be compensated by incorporating a reduced number of predicted classes, thus reducing the number of terrain transitions seen by the PAPAW.

We can see in Fig. 5 that several datapoints were misclassified as asphalt as opposed to sidewalk. This could be associated with significant overlap between kinematic characteristics on these two terrains, which is specifically highlighted when changing direction/speed. Reducing the number of predicted classes has a positive effect on prediction accuracy. For instance, higher prediction accuracy was achieved when grouping asphalt and sidewalk as one terrain type. However, this improvement in prediction accuracy comes at the cost of reduced granularity for terrain identification. For controller design, greater granularity would allow for more specific tuning for performance, whereas the greater accuracy may improve consistency for the user. The optimized trade-off between granularity and clinical relevance can be further explored through clinical testing. Although the prediction accuracy of terrain classification was reported to be affected by the linear velocity of the mobile platform in previous studies [11,13,14] (e.g., the prediction accuracy of grass vs. gravel was shown to be higher at higher speeds), classifiers developed in our study are mainly insensitive to velocity variability for a range of common wheelchair activities (including linear and circular motion).

A combination of time and frequency features exhibit better prediction accuracy compared to the use of only time or only frequency features. This is similar to what has been reported regarding the importance of combining different data sources to improve classification accuracy [15]. In contrast to the findings of Mei et al. [7], we found that time features have higher contributions to classification performance compared to frequency features (i.e., time-features exhibit higher predictive power). Similar to previous reports [16], we found that gyroscope measurements provide relevant information for terrain classification purposes. This is of significant importance since many of the existing terrain classification models rely on acceleration data only.

The proposed terrain classification frameworks in this work can be adopted for different PAPAW control applications. Subsequently, different model parameters can be determined based on user preferences or usability requirements (e.g., low-latency vs. higher accuracy terrain classification). Our future work includes collecting more data on a variety of terrains (including uneven, soft, wet surfaces) and when transitioning between different indoor and outdoor terrains. Additionally, we will compare the performance of the developed classification pipeline with existing neural network-based terrain classification frameworks.

V. Conclusion

We developed kinematic-based, velocity-independent, and cost-effective terrain classification frameworks that were successfully implemented offline and in real-time. These classification models are modular and can be customized for different PAPAW control applications. The output of these models can be used to adjust control parameters for different terrain conditions. Alternatively, the information about terrain type can be used in the future development of PAPAW controllers to improve traction or avoid wheel slip. Adoption of these control techniques may provide a more efficient and safe experience for PAPAW users.

References

[1] Algood SD,et al. Effect of a pushrim-activated power-assist wheelchair on the functional capabilities of persons with tetraplegia. Arch Phys Med Rehabil. 2005;86(3):380–6.

[2] Giesbrecht EM, et al. Experiences with Using a Pushrim-Activated Power-Assisted Wheelchair for Community-Based Occupations: A Qualitative Exploration. Can J Occup Ther. 2011;78(2):127–36.

[3] Guillon B, et al. Evaluation of 3 pushrim-activated power-assisted wheelchairs in patients with spinal cord injury. Arch Phys Med Rehabil. 2015;96(5):894–904.

[4] Wang H. Development and evaluation of an advanced real-time electrical powered wheelchair controller. 2012.

[5] Brooks CA, et al. Vibration-based terrain classification for planetary exploration rovers. IEEE Trans Robot. 2005;21(6):1185–90.

[6] Zhao K, et al. A New Terrain Classification Framework Using Proprioceptive Sensors for Mobile Robots. Math Probl Eng. 2017.

[7] Mei M, et al. Comparative Study of Different Methods in Vibration-Based Terrain Classification for Wheeled Robots with Shock Absorbers. Sensors. 2019;19(5).

[8] Weiss C, et al. A combination of vision- and vibration-based terrain classification. In: International Conference on Intelligent Robots and Systems, IROS. IEEE; 2008. p. 2204–9.

[9] Vulpi F, Milella A, Marani R, Reina G. Recurrent and convolutional neural networks for deep terrain classification by autonomous robots. J Terramechanics. 2021;96:119–31.

[10] DuPont EM, Roberts RG, Selekwa MF, Moore CA, Collins EG. Online Terrain Classification for Mobile Robots. ASME Int Mech Eng Congr Expo. 2005;42169:1643–8.

[11] Weiss C, Fechner N, Stark M, Zell A. Comparison of Different Approaches to Vibration-based Terrain Classification. 3rd Eur Conf Mob Robot. 2007;7–12.

[12] Khalili M, et al., Development of A Learning-Based Terrain Classification Framework for Pushrim-Activated Power-Assisted Wheelchairs. In: International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS. 2020. p. 4762–5.

[13] Dupont EM, et al. The Identification of Terrains for Mobile Robots Using Eigenspace and Neural Network Methods. In: Florida Conference on Recent Advances in Robotics. Miami, 2006. p. 1–5.

[14] Coyle E, et al. Vibration-based terrain classification for electric powered wheelchairs. In: Proceedings of the IASTED International Conference on Telehealth/Assistive Technologies. 2008. p. 139–44.

[15] Weiss C, Stark M, Zell A. SVMs for Vibration-Based Terrain Classification. In: Autonome Mobile Systeme. 2007.

[16] Ojeda L, et al. Terrain characterization and classification with a mobile robot. J F Robot. 2006;23(2):103–22.