A Feasibility Study of Depth Image Based Intent Recognition for Lower Limb Prostheses

Huseyin Atakan Varol, *Member, IEEE* and Yerzhan Massalin

***Abstract*— This paper presents our preliminary work on a depth camera based intent recognition system intended for future use in robotic prosthetic legs. The approach infers the activity mode of the subject for standing, walking, running, stair ascent and stair descent modes only using data from the depth camera. Depth difference images are also used to increase the performance of the approach by discriminating between static and dynamic instances. After confidence map based filtering, simple features such as mean, maximum, minimum and standard deviation are extracted from rectangular regions of the frames. A support vector machine with a cubic kernel is used for the classification task. The classification results are post-processed by a voting filter to increase the robustness of activity mode recognition. Experiments conducted with a healthy subject donning the depth camera to his lower leg showed the efficacy of the approach. Specifically, the depth camera based recognition system was able to identify 28 activity mode transitions successfully. The only case of incorrect mode switching was an intended run to stand transition, where an intermediate transition from run to walk was recognized before transitioning to the intended standing mode.**

I. Introduction

Due to the aging population and aggravating diabetes epidemic, the prevalence of lower limb loss is projected to increase significantly in the coming decades [1]. Functional replacement of lost limbs with robotic prostheses might improve the quality of life of the growing amputee population. This might be the motivation for the recent surge in robotic prosthesis research [2-4]. Powered lower limb prostheses were shown to reproduce characteristics of healthy gait kinematics in activities requiring net energy delivery to the joints such as upslope walking [5] and stair ascent [6].

The generalized control framework for the state-of-the-art powered lower limb prostheses can be represented as a three level hierarchy [7]. The high-level controller - sometimes referred as the intent estimator, supervisory controller or perception layer - is responsible for activity mode recognition and user volitional intent identification. The middle-level control takes the user intent and device state as inputs and maps these to the desired device state outputs. The low-level control consists of feedforward and feedback controllers, which force the actuators of the prosthesis to follow the desired position and/or torque references.

Various sensor modalities and their fusion were employed to realize the high-level control of robotic leg prostheses. Varol et al. utilized mechanical sensors of the device for the intent recognition problem of a powered knee and ankle prosthesis [8]. Electromyographic (EMG) and mechanical sensor signals were fused in order to improve the accuracy of locomotion-mode identification for neural control of robotic legs [9]. Hargove et al. decoded EMG signals from residual thigh muscles in real-time using pattern-recognition algorithms to change ambulation modes of a robotic prosthesis [10].

Sensory input for the high level controllers are either coming from a mechanical sensory interface on the prosthesis or from an EMG interface on the residual limb. The former implicitly measures the interaction of the device with the user and the environment. The latter obtains the information sent from the brain via efferent nerve fibers. The brain uses supraspinal level inputs such as the ones from the vestibular and visual systems extensively for decision making besides the afferent feedback from the musculoskeletal system. Present controllers try to infer these decisions either through mechanical interaction with the prosthesis or the muscle contractions on the residual limb, both of which are subject to different levels of delay. Presumably, decoding environmental information using a perception based sensor might decrease the intent recognition delays and also increase the reliability.

Optical sensors creating 3D representation of the environment with high reliability are one of the main reasons for the latest technical achievements in the area of autonomous robotic vehicles [11]. Point clouds obtained from depth images were also used for robot localization and navigation [12]. Recently collision avoidance in human-robot collaborative tasks was achieved using depth cameras [13]. Researchers were able to switch the locomotion mode of a hybrid mobile robot capable of moving with legs and/or wheels using time-of-flight depth sensors [14]. In the sphere of robotic prostheses, stairs were segmented using depth sensing and important parameters for stair gait were extracted for improving the control of robotic leg prostheses [15].

This work describes our preliminary study to evaluate the applicability of a depth camera based approach for activity mode intent recognition (see Figure 1). The rest of the paper is organized as follows. Our methodology and implementation are explained in Section II. The results of activity mode classification and switching are shown in Section III. Section IV concludes the paper.

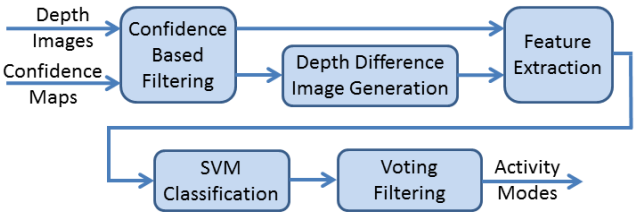

Figure 1. Depth image based activity mode recognition framework.

The authors are with the Department of Robotics and Mechatronics, Nazarbayev University, 53, Kabanbay Batyr Ave, Z05H0P9 Astana, Kazakhstan. Emails: {ahvarol, yerzhan.massalin}@nu.edu.kz. Corresponding author: Huseyin Atakan Varol.

978-1-4577-0220-4/16/$31.00 ©2016 IEEE

II. Depth Image Based Intent Recognition

The activity mode recognizer is intended to be incorporated to a lower limb prosthetic device as the high-level controller. An SVM based classifier is at the heart of the activity mode recognizer. A database of depth images belonging to different activity modes is used to train this classifier. To encompass the dynamic aspects of gait, depth difference images are generated by subtracting a filtered depth image from a previous one. Simple features of rectangular sub-regions of depth and depth difference images are extracted for classifier training. The last part of the activity mode recognizer is the voting filter, which is used to increase the robustness of the activity mode switching in real-time operation. This section describes the components of the activity mode recognizer in detail.

A. Depth Image Acquisition and Database Generation

The experimental procedures involving human subjects described in this paper were approved by the Research Committee of the School of Science and Technology of the Nazarbayev University. Figure 2 shows a 3D printed adapter housing DepthSense DS 325 depth camera [16] donned by a healthy 27-year-old male subject (1.75 m and 64 kg). The depth camera protruded from the upper part of the lower leg making a 45 degree angle with the ground normal. The camera is capable of providing three data streams: RGB images, depth images and confidence maps. Depth data for a pixel is the distance value in millimeters to the nearest object in the field of view of the camera corresponding to that particular pixel coordinate. It is represented as a 16 bit unsigned integer. The confidence map contains the amplitude values of the returning infrared light at each pixel and can be used as a measure of reliability for each depth value. In this work depth images of 320 × 240 pixel resolution and corresponding confidence maps were acquired to a notebook computer (carried in a backpack by the subject) at 30 Hz sampling rate via USB port. The accuracy of depth images decrease with increasing object distance. Therefore, we saturated distances exceeding 1 m as a preprocessing step.

The subject participated in ten data collection sessions. For each trial the subject was instructed to ambulate freely between two and three minutes engaging in all five activity modes (standing, walking, running, stair ascent and stair descent). The data collection experiments were conducted in various buildings and the atrium of the Nazarbayev University campus under natural, artificial and mixed lighting conditions. Trials were also recorded with a chest-strapped action camera (GoPro HERO4 Black). Class labels were assigned to the depth images with the help of the action camera record. Out of these ten sessions, five were used for classifier training and testing. Two sessions were used for voting filter parameter selection. The remaining three sessions were used to test the efficacy of our approach for real-time operation scenarios.

B. Depth Image Filtering Using Confidence Measures

Time-of-flight depth cameras emit modulated infrared light and determine the distance by measuring the phase difference of the light reflected from the objects. Low infrared reflectivity of surfaces and multiple reflections are some of the factors which introduce noise to depth images. It is recommended to filter/smooth the depth images as a preprocessing step.

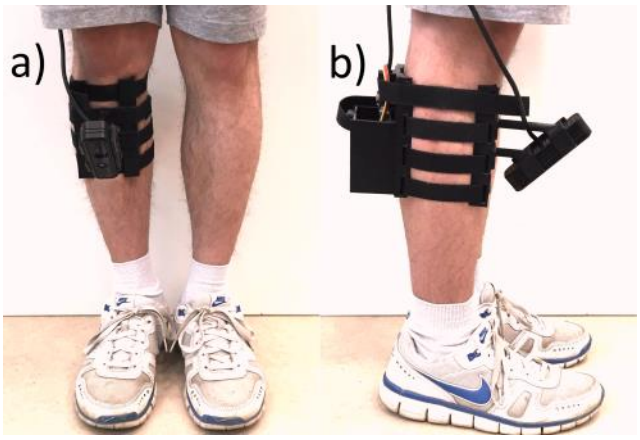

Figure 2. Attachment of the depth camera to the leg of the experimental subject: a) front and b) side views.

As described in [17] and implemented for a mobile robot application in [14], we used a 3 by 3 averaging filter which uses corresponding normalized confidence map values as weighting factors for each pixel. Accordingly, the filtered pixel at the coordinate (i,j) of the filtered depth image F was computed as $F(i,j) = \frac{\sum_{a=-1}^{1}\sum_{b=-1}^{1} I(i+a,j+b)\cdot A(i+a,j+b)}{\sum_{a=-1}^{1}\sum_{b=-1}^{1} A(i+a,j+b)}$, where I and A are the original depth image and the confidence map, respectively.

C. Feature Extraction from Rectangular Sub-regions

The depth images from different activity modes frequently look indiscernible. As an example, left column of the Figure 3 shows one frame from the standing and one frame from the stance phase of walking. For this reason, using only depth images for activity mode recognition is analogous to recognizing activity modes using only instantaneous position-based features for a mechanosensory intent recognizer. In order to embed motion information to the features, we created depth difference images by subtracting the depth image of five samples before from the current depth image (resulting in a time interval of 165 ms). For standing mode, this difference is minute as reflected by the uniformly colored depth difference image (see Figure 3). On the contrary, depth difference image for the stance phase of walking reflects the motion of the shank during this activity. These differences are shown on the right column of Figure 3.

To ensure real-time applicability of the approach, computationally efficient features containing rich motion information are needed. We used a variant of the features used in [14, 18]. Specifically, we divided the filtered depth image into rectangular sub-regions of the same size iteratively (see Figure 4). Level 1 is the full image. Level 2 consists of four quadrants of the full image. Each of these quadrants is again divided to four parts in Level 3 and so on. Four iterations of this process yielded 81 sub-regions. Four statistical features (mean, minimum, maximum and standard deviation) were computed for each region. With the repetition of the same procedure with the depth difference

image, a total of 680 features (85 × 4 × 2) were generated from each sample.

D. SVM Classification

SVM classification method tries to find optimal hyperplanes separating samples belonging to different classes [19]. SVM leverages the "kernel trick" which allows linear classification algorithms to learn nonlinear mappings without their explicit representation in a computationally efficient manner.

In this work, Matlab implementation of the SVM algorithm was employed. We used a 5 fold cross-validation scheme leaving features from one trial out at each fold for testing. Cubic SVM kernel and one versus one option for multiclass classification were chosen. The features were normalized to the range [0, 1] before SVM training. The SVM model belonging to the fold with the highest accuracy and the normalization parameters were saved to be used in the further experiments.

E. Voting Scheme for Real-time Mode Switching

Especially during the transitions between activity modes, a few instances can be misclassified. If activity mode switching is executed based on these raw classification decisions, the controller might switch between different activity modes erroneously. In a powered prosthesis this might decrease the user comfort. In the worst case, the user might fall and sustain an injury. To alleviate this problem, a voting scheme similar to the one in [8] was implemented.

For the voting filter, last *n* classifier decisions are stored in a vector and mode switching is executed only if a certain number of the classification results belong to one specific class. The increased robustness comes with the price of additional mode switching delay. Therefore, careful parametrization of the voting filter is needed. The parameters of the voting filter were determined by trial and error. According to these parameters, a mode switching occurs if more than 60 percent of the classification results belong to one specific class in a voting vector containing the latest 15 classifier decisions.

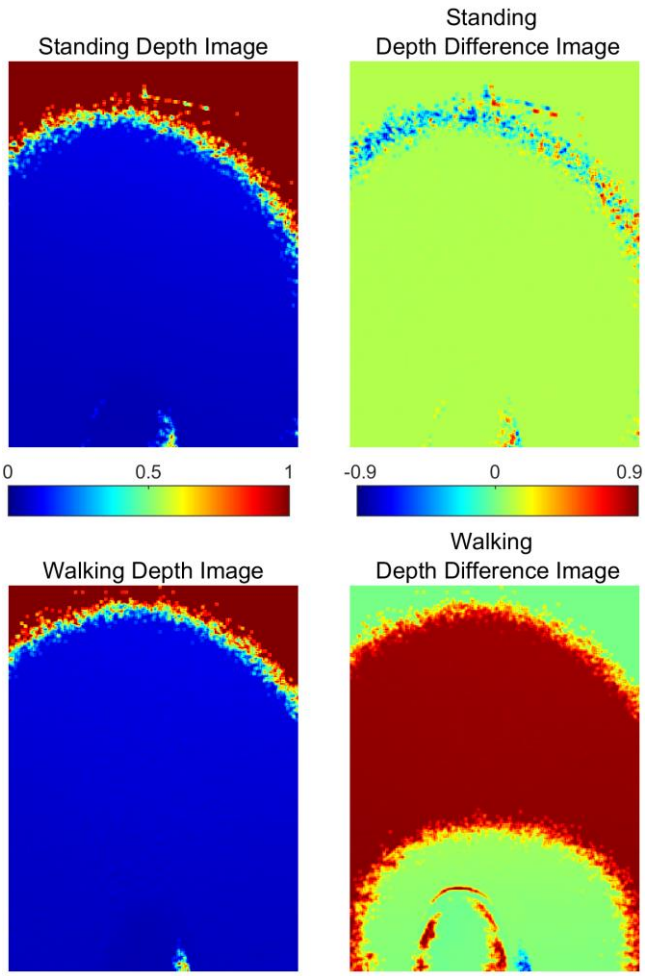

Figure 3. Standing and walking depth images and the corresponding difference images obtained by subtracting the depth images from the depth images five sample before.

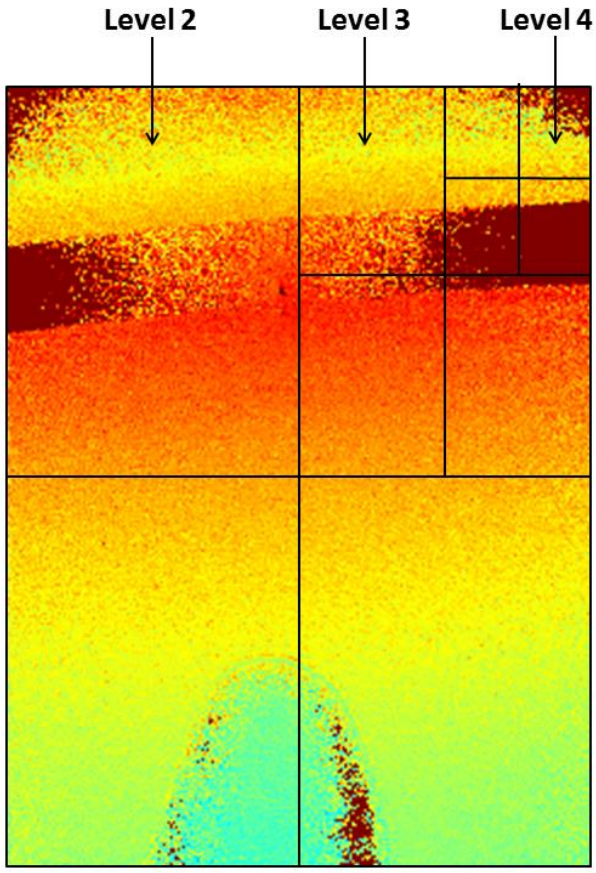

Figure 4. Depth image while approaching a set of stairs illustrating the rectangular regions of different recursive levels used for feature extraction. Level 1 is the whole depth image.

TABLE I. 5-FOLD CROSS-VALIDATION CONFUSION MATRIX OF THE SVM CLASSIFIER FOR FIVE ACTIVITY MODES

		Predicted Class				
		Run	**Walk**	**Stand**	**Upstairs**	**Downstairs**
Actual Class	**Run**	1724	35	35	4	6
	Walk	5	7213	29	3	34
	Stand	0	2	11646	4	0
	Upstairs	3	19	3	1146	0
	Downstairs	3	33	3	0	982

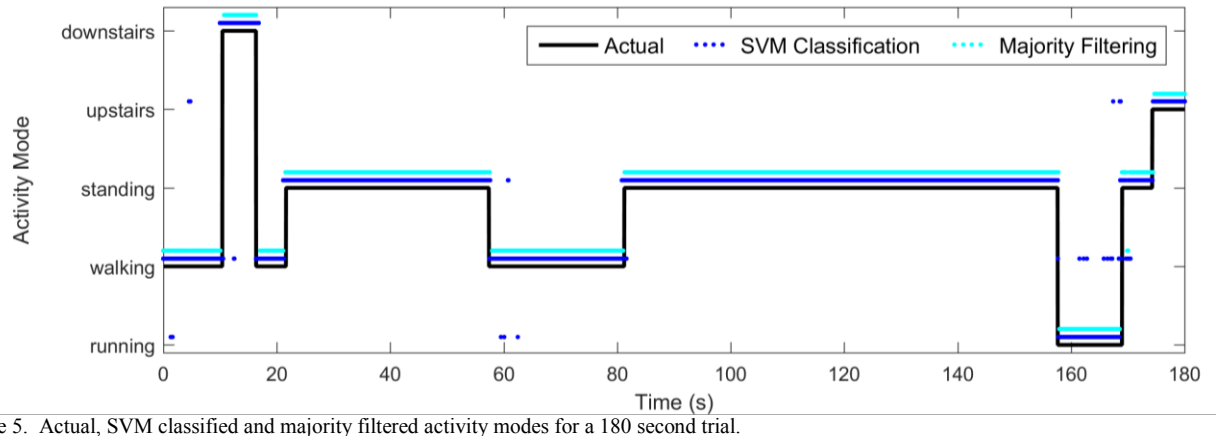

Figure 5. Actual, SVM classified and majority filtered activity modes for a 180 second trial.

III. Results

The confusion matrix combining the results of the 5-fold cross-validation for the SVM classifier is given in Table I. 22711 of the 22932 samples are classified correctly reaching an accuracy of 99.0 percent. The accuracies for walking and standing are higher compared to the running, stair ascent and stair descent. Walking and standing modes also have significantly more samples. This suggests that the expansion of the database with more running, stair ascent and stair descent samples could improve the performance.

Figure 5 shows the effects of the voting scheme to activity mode switching. In simulations conducted with the three trials set aside for assessing voting scheme performance, the depth camera based recognition system was able to identify all 28 activity mode transitions successfully. The only case of incorrect switching was an intended run to stand transition. An intermediate transition from run to walk was recognized before transitioning to the intended standing mode.

IV. Conclusion

This work describes an activity mode recognition framework, which classifies images obtained from a time-of-flight depth camera using SVM algorithm. Classification accuracy of 99 percent for 5-fold cross-validation is achieved. A voting filter is used to increase the reliability of activity mode switching for real-time operation scenarios. The results show that depth image based activity mode recognition is a promising option for the supervisory control of powered lower limb prostheses. Future work includes testing of the framework with multiple subjects and incorporating the depth image based recognizer to the high level control system of a powered lower limb prosthesis.

References

[1] K. Ziegler-Graham, E. J. MacKenzie, P. L. Ephraim, T. G. Travison, and R. Brookmeyer, "Estimating the prevalence of limb loss in the United States: 2005 to 2050," *Archives of Physical Medicine and Rehabilitation,* vol. 89, pp. 422-429, Mar 2008.

[2] F. Sup, H. A. Varol, J. Mitchell, T. J. Withrow, and M. Goldfarb, "Preliminary evaluations of a self-contained anthropomorphic transfemoral prosthesis," *IEEE-ASME Transactions on Mechatronics,* vol. 14, pp. 667-676, Dec 2009.

[3] S. K. Au and H. M. Herr, "Powered ankle-foot prosthesis - The importance of series and parallel motor elasticity," *IEEE Robotics & Automation Magazine,* vol. 15, pp. 52-59, Sep 2008.

[4] (15 Feb 2016). *Ossur Power Knee*. Available: http://www.ossur.com/prosthetic-solutions/products/dynamic-solutions/power-knee

[5] F. Sup, H. A. Varol, and M. Goldfarb, "Upslope walking with a powered knee and ankle prosthesis: Initial results with an amputee subject," *IEEE Transactions on Neural Systems and Rehabilitation Engineering,* vol. 19, pp. 71-78, Feb 2011.

[6] B. E. Lawson, H. A. Varol, A. Huff, E. Erdemir, and M. Goldfarb, "Control of stair ascent and descent with a powered transfemoral prosthesis," *IEEE Transactions on Neural Systems and Rehabilitation Engineering,* vol. 21, pp. 466-473, May 2013.

[7] M. R. Tucker, J. Olivier, A. Pagel, H. Bleuler, M. Bouri, O. Lambercy, *et al.*, "Control strategies for active lower extremity prosthetics and orthotics: A review," *Journal of Neuroengineering and Rehabilitation,* vol. 12, Jan 2015.

[8] H. A. Varol, F. Sup, and M. Goldfarb, "Multiclass real-time intent recognition of a powered lower limb prosthesis," *IEEE Transactions on Biomedical Engineering,* vol. 57, pp. 542-551, Mar 2010.

[9] H. Huang, F. Zhang, L. J. Hargrove, Z. Dou, D. R. Rogers, and K. B. Englehart, "Continuous locomotion-mode identification for prosthetic legs based on neuromuscular-mechanical fusion," *IEEE Transactions on Biomedical Engineering,* vol. 58, pp. 2867-2875, Oct 2011.

[10] L. J. Hargrove, A. M. Simon, A. J. Young, R. D. Lipschutz, S. B. Finucane, D. G. Smith, *et al.*, "Robotic leg control with EMG decoding in an amputee with nerve transfers," *New England Journal of Medicine,* vol. 369, pp. 1237-1242, Sep 26 2013.

[11] A. Geiger, P. Lenz, and R. Urtasun, "Are we ready for sutonomous driving? The KITTI Vision Benchmark Suite," in *2012 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2012, pp. 3354-3361.

[12] J. Biswas and M. Veloso, "Depth camera based indoor mobile robot localization and navigation," in *2012 IEEE International Conference on Robotics and Automation (ICRA)*, 2012, pp. 1697-1702.

[13] B. Schmidt and L. H. Wang, "Depth camera based collision avoidance via active robot control," *Journal of Manufacturing Systems,* vol. 33, pp. 711-718, Oct 2014.

[14] A. Saudabayev, F. Kungozhin, D. Nurseitov, and H. A. Varol, "Locomotion strategy selection for a hybrid mobile robot using time of flight depth sensor," *Journal of Sensors,* 2015.

[15] N. E. Krausz, T. Lenzi, and L. J. Hargrove, "Depth sensing for improved control of lower limb prostheses," *IEEE Transactions on Biomedical Engineering,* vol. 62, pp. 2576-2587, Nov 2015.

[16] (9 March 2016). *SoftKinetic DepthSense Cameras*. Available: http://www.softkinetic.com/Products/DepthSenseCameras

[17] M. Frank, M. Plaue, and F. A. Hamprecht, "Denoising of continuous-wave time-of-flight depth images using confidence measures," *Optical Engineering,* vol. 48, Jul 2009.

[18] E. Ugur, M. R. Dogar, M. Cakmak, and E. Sahin, "The learning and use of traversability affordance using range images on a mobile robot," in *IEEE International Conference on Robotics and Automation (ICRA)*, Roma, 2008, pp. 1721-1726.

[19] R. O. Duda, P. E. Hart, and S. D.G., *Pattern Classification*, 2nd ed.: Wiley-Interscience, 2000.