Journal of Science and Technology
ISSN: 2456-5660 Volume 9, Issue 4 (April -2024)
www.jst.org.in
DOI:https://doi.org/10.46243/jst.2024.v9.i4.pp1-9

Integration of Human Vision and Machine perception to Forecast the User's Desired Mode of Movement by Using Deep Learning Technique

Dr.Ch.Santhi Rani[1] |Ch.Lavanya[2] |Sk.Nagurbi[3]|V.Sahith sai[4]|B.Harshith[5]

[1]Dean Academics, Department of Electronics and Communication Engineering, Usha Rama College of Engineering and Technology,
[2]Department of Electronics and Communication Engineering, Usha Rama College of Engineering and Technology,
[3]Department of Electronics and Communication Engineering, Usha Rama College of Engineering and Technology,
[4]Department of Electronics and Communication Engineering, Usha Rama College of Engineering and Technology,
[5]Department of Electronics and Communication Engineering, Usha Rama College of Engineering and Technology,

A. To Cite this Article

Dr.Ch.Santhi Rani. Ch.Lavanya, Sk.Nagurbi, V.Sahith sai, B.Harshith, **"Integration of Human Vision and Machine perception to Forecast the User's Desired Mode of Movement by Using Deep Learning Technique"** ***Journal of Science and Technology, Vol. 09, Issue 04 - April 2024, pp1-9***

B. Article Info

Received: 28-02-2023 ***Revised: 05 -03-2024*** ***Accepted: 21-03-2024*** ***Published: 6-04-2024***

ABSTRACT

Wearable robot control relies on anticipating the user's preferred mode of locomotion to provide smooth transitions for the user when traversing different terrains. While machine perception has shown promise recently for detecting impending terrains in the trip path, current methods are unable to recognize human intent, which is necessary for coordinated wearable robot operation, and are instead restricted to environment perception. Therefore, the goal of this research is to create a new system that accurately forecasts the user's mode of movement by combining machine perception (which captures ambient data) with human vision (which represents user intent). The system can detect the user's intended path in a complicated setting with various terrains since it has multimodal visual information. Moreover, a fusion algorithm based on dynamic time-warping techniques To produce flexible judgments on the time of locomotion mode change for wearable robot control, a fusion technique was devised to align the temporal forecasting from individual modalities. Through the use of experimental data gathered from five people, the system's performance was verified. It demonstrated a high degree of intent detection accuracy (almost 96% on average) and dependable decision-making on locomotion transition with customizable lead time. These encouraging results show that combining machine perception and human vision may be used to identify lower limb wearable robots' intent to move.

KEYWORDS: deep learning, wearable robots, intent detection, machine perception, and human vision.

II. INTRODUCTION

IN NATURE, humans have developed locomotor abilities to adjust to shifting dynamics and functional needs in negotiating diverse settings. Many lower limb wearable robots, such as exoskeletons and robotic prostheses, have been developed to restore natural locomotion for populations with reduced mobility due to conditions like spinal cord injuries or limb amputations. However, these wearable robots are unable to adapt to their surroundings in a way that meets the needs of the user since they are not connected to the user's neurological control system. Thus, to enable lower-limb wearable robots to coordinate with user intent for environmentally adaptive locomotion, solutions are required**.** Finding the user's mode of locomotion (e.g., level ground walking, stair ascent/descent, ramp ascent/descent) has always been the first step in solving problems. mechanical sensors on board, like Wearable robots with inertial measurement units (IMU) and motion and force sensors have been used to categorize various gait patterns, thereby determining the locomotion mode being used. But typically, these mechanical sensors don't offer meaningful measurement alters until the manner of locomotion is changed. Wearable robot control to enable smooth terrain transitions is challenged by the mechanical sensors-based locomotion mode recognition system's delayed, reactive response. Electromyography (EMG) signals, or efferent neural control signals of limb mechanics, have been utilized alone or in combination with mechanical signals to determine the user's intended locomotion to forecast the locomotion mode changes. According to one study, combining mechanical and EMG sensors can increase the precision of identifying the locomotor mode. Utilizing cameras to recognize the terrain in front of the user is an additional strategy. By outfitting wearable robots with machine perception sensors, robotic limb control is made possible. suitable terrain transitions, as demonstrated by the categorization of simple RGB photos taken by a camera for terrain identification. However, because the 3D environment data is compressed, the simple RGB picture classification method is less able to differentiate between sloping terrains. To reconstruct the depth images and 3D point clouds before classifying them, efforts have been made to extract more depth information from the surroundings, hence increasing the range of applications for machine perception.

Research has indicated that this methodology facilitates precise identification of the walking terrain in front of the user, which arrives between 0.6 and 4 seconds before the actual change in locomotion mode. Furthermore, this method does not call for the training of a user-specific model, making it more realistic and scalable. It has been difficult to directly use machine perception for wearable robot control, nevertheless. In essence, existing systems did not directly tap into human biological signals for intent recognition; instead, they relied only on environment perception to estimate human intent. The fundamental premise is that when walking terrain changes, the user's intention to switch locomotion modes will also vary accordingly. However, in a complicated setting with numerous terrains present at once, this assumption might not hold. Furthermore, the user's intention to switch between locomotion modes does not always coincide with the moment when the terrain changes are recognized. Terrain recognition alone is not enough for wearable robots to make an accurate control decision; instead, the user's purpose must be understood. In light of the increasing need to reliably capture user movement intent through scalability, utilizing data obtained from human vision has gained appeal as numerous research has indicated that human eyesight plays a crucial role in controlling movement. Human vision, a measure of visual attention, has been applied to decipher user intent in locomotion tasks such as route planning and foot placement. Furthermore, the scalability of human sight is conferred by its relatively stable behaviors across persons and daily settings. The gaze therefore presents itself as a potentially useful information source for determining user intent. Human vision, however, often occurs before the actual locomotor mode shifts, sometimes looking away from the approaching terrain and sometimes toward the world far ahead. As such, it is difficult to allow wearable robots to use gaze information alone to determine when to make timely transitions. As a result, we developed a novel approach in this study that combined human vision and ambient cues to forecast a user's desired mode of movement for lower-limb wearable robots. data that machine perception interprets. RGB visual images taken using eye tracker glasses included gaze information directed towards a certain terrain. In the meantime, point clouds of the surroundings provided information from a depth camera to accurately identify the terrain in front of the user. We used two networks to categorize the environmental and human vision data independently as they were carried by different modalities, and we implemented the idea of dynamic temporal warping (DTW) to combine the many forecasting types. Our fusion technique made continuous probabilistic judgments rather than immediately building deterministic ones and provided a flexible option to determine the timing for wearing robots to alter their method of mobility in accordance with probability in order to facilitate a seamless, secure, and easy transition for people. The following are this study's primary contributions: 1) By fusing human vision with machine perception, we suggested a new system to anticipate planned movement changes for wearable robots.2) To combine multimodal data and generate variable choices regarding the timing of locomotion transition, we created a DTW-based approach.3)Using early validations conducted in indoor trials (including participants who were amputees of both lower limbs and able-bodied individuals), we showcased the practicability and potential of the system for wearable robot control.

III. Methodology

A. Overview of the Pipeline

The suggested system forecasted the anticipated movement of the user options for wearable robot control prior to moving across different types of terrain. In order to achieve this, two categorization networks were used, as seen in Fig. 1(a).In particular, the PointNet forecasted the terrain directly in front of the user (i.e., 1.5 meters ahead) along the walking path, but the Gaze-Terrain Network (GT-Net) forecasted the terrain on which the user's gaze fixated regardless of the distance between them an integrated module also produced a combined probabilistic decision on the planned locomotion mode for robotic control of the subsequent step by aligning the temporal forecasting from both networks generating a combined probabilistic determination of the desired locomotion mode for robotic guidance in the subsequent phase The eye-tracker and depth camera provided the multimodal data that the system supplied into PointNet and GT-Net, respectively. While the depth camera provided synchronized scene images, the eye tracker recorded gaze positions. point clouds representing the perceived surroundings given the noisy nature of gaze location measurement, utilizing the measured gaze directly may necessitate precise knowledge of noise level (i.e., gaze tracking error), which is typically unknown to us. Therefore, as demonstrated, we used gaze information in an indirect manner by superimposing gaze locations on scene photos to create gaze-embedded scene (GES) images, which we then fed into the GT-Net in Figure 1(c). The network avoided the need for acquiring human attention by learning to anticipate human attention distributions and jointly attend terrains in an end-to-end manner, exact information about noise. The GT-Net architecture is displayed in Fig. 1(b). It was composed of three parts: a feature extractor, a gaze encoder, and a classification head. It was inspired by earlier work. The design of the feature extractor was derived from the 224 ×224 ×3 is the input size for MobileNetV1. In a single feedforward run, three sets of hierarchical features—designated as F1, F2, and F3—were retrieved. F1 had dimensions of 28 × 28 × 128 and F2 and F3 had dimensions of 14 × 14 × 256 and 7 × 7 × 512, in that order. We were able to obtain multi-scale features by such an extraction, which took into consideration variations in feature requirements across the jobs of forecasting attention and identification of terrain. The early-stage features F1 and F2, which contained more local visual information, were supplied into the gaze encoder in order to anticipate the attention distribution map z because the superimposed gaze was rather small. However, late-stage features F3, which provide more global visual information, were used for terrain recognition in order to comprehend the scene's context on a broader scale.

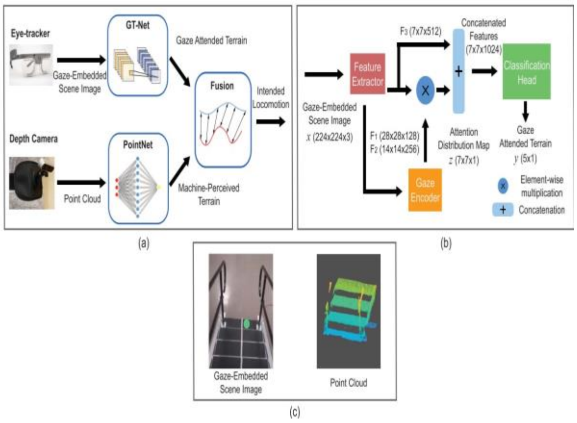


Fig1 shows a diagram of the suggested system. (a) The planned locomotion recognition pipeline fuses unimodal forecasting to provide choices on the locomotion mode. It receives as inputs gaze-embedded scene pictures and point clouds from the data gathering devices (i.e., eye-tracker and depth camera).

(b) The architecture of GT-Net, which consists of three parts (the feature extractor, gaze encoder, and classification head) uses five candidate classes (the upstairs, downstairs, up-ramp, down-ramp, and level-ground) to create a gaze-attended terrain.

(c) An illustration of a point cloud and a gaze-embedded scene picture. After that, the classification head may forecast gaze-attended terrains with a global average pooling layer and a completely linked softmax layer by merging the attention distribution map z and global features F3. We used the PointNet that was already in place in addition to GT-Net to employ point cloud data to differentiate perceived terrains in front of the user. Two fundamental elements (that is, the common to accomplish both permutation and transformation invariances, the network incorporates a transformer net and a multi-layer perceptron. But taking into account our tiny 5-class problem instead of the initial 40-class one result, we retrained each original layer with our problem-specific data after reducing the number of trainable parameters by a factor of 4 to make the architecture lighter and more appropriate for our use. Finally, the decision fusion module looks for the best method of aligning the mismatched multimodal forecasting by applying the DTW idea.In accordance with the alignment, a fused to operate wearable robots in the following stage, a probabilistic judgment on the desired locomotion mode is produced.

B. The Gaze-Terrain Network's Training Objective

We broke down the challenge of detecting planned terrain y given an input of by drawing inspiration from earlier research GES picture x into two smaller tasks: z is the estimate of human attention, and x is the forecasted attended terrain. Using the training objective Lobj to be minimized, which is as follows:

$$L_{obj}=-\log p(y|x,q(z|x))+D_{KL}q(z|x)||p(z|x)(1)$$

where the first element on the right-hand side is the terrain forecasting y negative log-likelihood (also known as a cross-entropy loss) given input x and the estimated attention distribution, q (z|x), and the second term is the distance between the estimated attention distribution (q(z|x) and the previous attention distribution (p(z|x)) which is known as the Kullback-Leibler (KL) divergence. For the sake of this investigation, we used a conventional 2D isotropic Gaussian tracking error and assumed that the previous attention distribution, p(z|x), was the same as the numerical distribution centered at the measured gaze point.

C. Point Cloud Egomotion Compensation

An unsorted collection of vectors that depicts an object's form in $\mathbb{R}^3$ spaces is called a point cloud. The object's original representation in the global coordinate $\mathbf{O}_g$ (e.g., ground) must be reconstructed by applying a homogeneous transformation, $^{g}\mathbf{T}l \in SE(3)$, to account for the relative orientation and position between the coordinates caused by ego motion. This is because the point cloud data captured by the depth camera is represented in the camera's local coordinate $\mathbf{O}_l$.

$$\begin{bmatrix} \mathbf{p}_g \\ 1 \end{bmatrix} = {}^{g}\mathbf{T}_l \begin{bmatrix} \mathbf{p}_l \\ 1 \end{bmatrix} = \begin{bmatrix} {}^{g}\mathbf{R}_l & {}^{g}\mathbf{d}_l \\ \mathbf{0}_{1\times 3} & 1 \end{bmatrix} \begin{bmatrix} \mathbf{p}_l \\ 1 \end{bmatrix}$$

where the same point is represented by $\mathbf{P}_l$ and $\mathbf{P}_g$ in the local and global coordinates, respectively, relative position is indicated by the notation relative position, and rotation matrix, $^{g}\mathbf{R}_l \in SO(3)$, given that $^{g}\mathbf{d}_l \in \mathbf{R}^3$. Readings from an IMU that is integrated into the depth camera allow us to determine the value of g$\mathbf{R}_l$. We just ignored the value of g$\mathbf{d}_l$ and set it to a zero vector during the ego-motion adjustment since the pose of the point cloud was more important to our application than its absolute position.

D.FusionProcess

A subsequence variation of DTW was used to align and fuse multimodal forecasting since the sequences of forecasting between modalities were typically misaligned. The Appendix contains implementation details for the DTW, which was first suggested, Additionally, we utilized Jensen-Shannon divergence to quantify the similarity of each pair of forecasting for DTW alignment costs because both GT-Net and PointNet provided forecasting of a probability distribution over all locomotion modes.

By simply averaging the probability distribution over all forecasting inside the window, a fused probabilistic decision may be obtained from the best-matched forecasting sequences of both modalities within the window of interest. Additionally, a transition threshold must be reached in order to change the probabilistic prosthetic control choice into a deterministic one to be configured and utilized as follows:

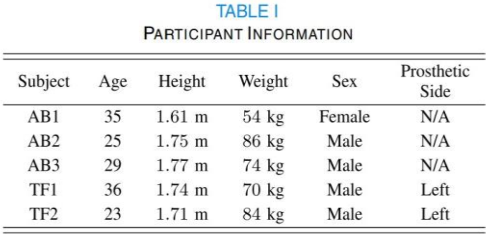


TABLE I
PARTICIPANT INFORMATION

Subject	Age	Height	Weight	Sex	Prosthetic Side
AB1	35	1.61 m	54 kg	Female	N/A
AB2	25	1.75 m	86 kg	Male	N/A
AB3	29	1.77 m	74 kg	Male	N/A
TF1	36	1.74 m	70 kg	Male	Left
TF2	23	1.71 m	84 kg	Male	Left

$$D_t = \begin{cases} i*, & \boldsymbol{Pr}_t[i*] > \epsilon_i \\ D_{t-1}, & \text{otherwise} \end{cases}$$

where Dt−1, Dt, and $\mathbf{Pr}_t$[i] stand for the probability of the ith locomotion mode in the current fused probabilistic decision, the previous fused deterministic decision, and the current fused deterministic decision, respectively. The locomotor mode with the highest probability in the present fused probabilistic choice is indicated by the index i
∗. The inference task determines which threshold should be used; further information is included in Section IV-C below.

IV. RESULTS

A. Stand Test Intent Forcastion AccuracyThe forcastion accuracy for gaze-attended terrains in the naïve models' standing task is displayed in Fig. 3(a). As expected, the 10 naive-designed models performed the worst in terms of average accuracy (52.45% ± 2.39%) in capturing the user's intents when gaze information was not used. This suggests that unprocessed photos that just show visual situations might not be adequate

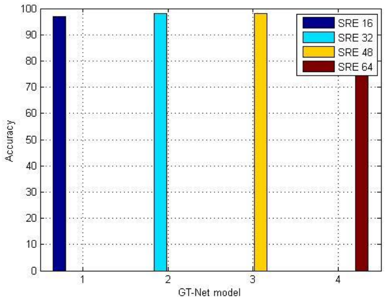
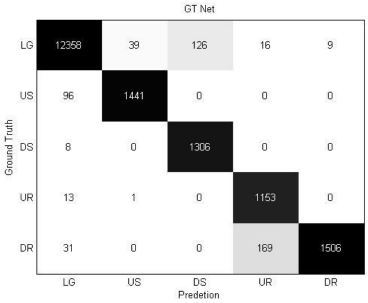


Fig. 3 GT-Net Model

More specifically, the GT-Net design produced the best accuracy of 96.86% when averaged over all 40 models for each design. Furthermore, we saw that the accuracy performance across the three designs was also impacted by the manner in which the gaze information was used. With respect to changing SRE values for the numerical gaze distribution, the GT-Net models demonstrated the greatest insensitivity while maintaining a comparable level of accuracy. With an SRE of 32 pixels, the maximum accuracy was 97.27%, while with an SRE of 64 pixels, the lowest accuracy was 96.12%. The data also suggest that in order to guarantee reliable user intent identification and insensitivity to SRE settings during model training, our developed gaze encoder in the GT-Net is crucial.



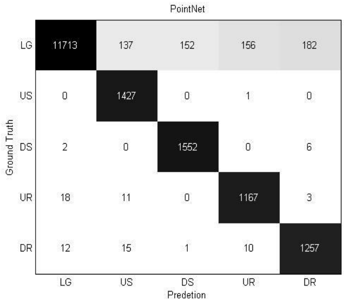
GT Net

Ground Truth	LG	US	DS	UR	DR
LG	12358	39	126	16	9
US	96	1441	0	0	0
DS	8	0	1306	0	0
UR	13	1	0	1153	0
DR	31	0	0	169	1506

Predetion

Fig 4: Confusion matrix for unimodal predicting, with the left column and bottom row summarizing the sensitivity and precision values, respectively. (a) GES pictures obtained by the eye-tracker combined with GT-Net forecasting.

One may argue that the advantageous characteristics stem from the simultaneous end-to-end learning of terrain identification and human attention. Under the supervision of both gaze and landscape data, the gaze encoder could efficiently determine what visual data



PointNet

Ground Truth	LG	US	DS	UR	DR
LG	11713	137	152	156	182
US	0	1427	0	1	0
DS	2	0	1552	0	6
UR	18	11	0	1167	3
DR	12	15	1	10	1257

Predetion

(b) PointNet forecasts using the depth camera's point clouds. allowed for the adaptive development of attention distributions, and it was more relevant and informative. Furthermore, since the GT-Net is a stand-alone module in our suggested system, the encouraging outcomes imply that many additional applications may use it to identify the visual attention and intent of humans.

B. Unimodal Forecasts for Mobility Task

The confusion matrices of the best-trained PointNet and GT-Net models throughout the locomotion tasks are displayed in Fig. 4. The results in Fig.4(a) reveal that, in comparison to other gaze-attended terrains, the GTNet found it more difficult to identify the UR and DR, which demonstrated the lowest accuracy (86.2%) and sensitivity (84.1%), respectively. A portion of the comparatively poorer performance on ramps might be attributed to the eye-tracker's intrinsic limitations with RGB pictures. Sloped surfaces may not be distinguishable in pure RGB photos without additional depth and motion information.

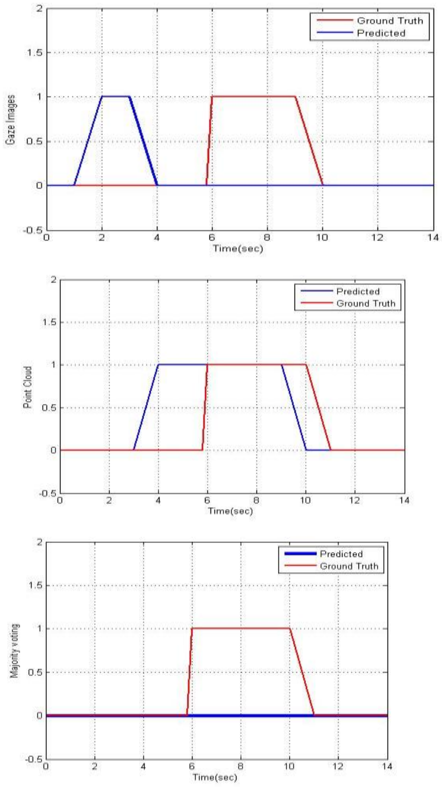


Fig5: A comparison of the performance of mean probability fusion, DTW-based fusion, and majority voting fusion for a scenario including sequential mode switches in locomotion. (a)Forcastions were made using majority voting fusion with GES point clouds and picture runs, especially if the photos were taken locally, directly above the runs, and did not provide the broad spatial

information that side views may provide. However, as Fig. 4(b) illustrates, the pointNet exhibited a generally acceptable sensitivity but a relatively poor accuracy (i.e., below 90%) on the US (89.7%), UR (87.5%), and DR (86.8%), with a tendency to misclassify LG to them. Biased

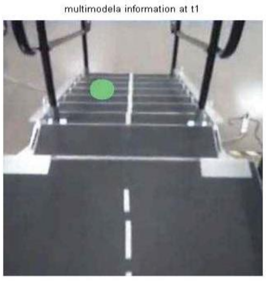
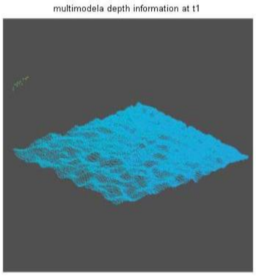
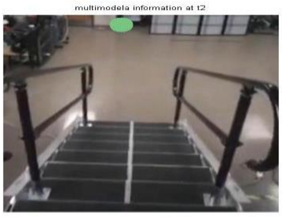


Fig 5 (b)





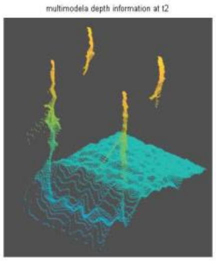


Fig 5(b–c) Point clouds and GES pictures taken at times t1 and t2, respectively

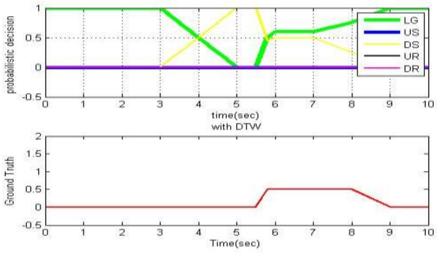
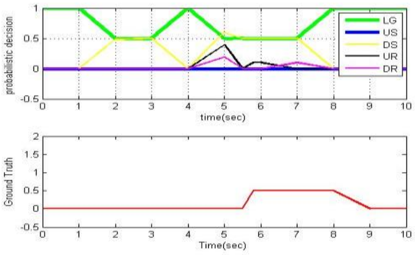


Fig .5. (d) Grey-shaded regions represent time periods where the probability of LG exceeds the probability of DS. These judgments are based on probabilistic analysis of mean probability fusion and associated ground truth locomotion modes.



e) Probabilistic choices of DTW-based fusion and related ground truth locomotion modes; the pink region indicates the relevant time frame for deciding whether to alter the control mode for wearable robots.

runs, especially if the photos were taken locally, directly above the runs, and do not provide the broad spatial information that side views may provide. However, as Fig. 4(b) illustrates, the pointNet exhibited a generally acceptable sensitivity but a relatively poor accuracy (i.e., below 90%) on the US (89.7%), UR (87.5%), and DR (86.8%), with a tendency to misclassify LG to them. biased forecasts of transition zones from LG to non-LG terrains contributed to the low accuracy, the reason for the higher likelihood of misclassification of these locations as non-LG terrains was the presence of feature points in the recorded point clouds that did not correspond to a clean LG point cloud, which is usually a horizontal plane.

C. Fusion's Impact on the Transition of Locomotion

A sample of the data from the continuous locomotion challenge is displayed in Fig. 5. By combining forecasting of each separate modality in the sequence, three methods (majority voting fusion, mean probability fusion, and DTW-based fusion) were evaluated for their capacity to produce precise and fast outcomes. The majority voting fusion failed to detect the change from LG to DS, as Fig. 5(a) illustrates. The temporal misalignment of predicting sequences, where no overlapping decision of DS was identified between modalities, caused the missing transition. Humans typically focus on the chosen terrain three to five steps ahead of the actual switch to the locomotion mode, and they stop using their gaze after that. Point clouds, on the other hand, need a specific amount of space between the camera and the terrain in order to accurately anticipate; as a result, their forecasting of terrain information is made after the gaze-based user intent. The participant in Fig. 5(b) focused on the DS at time t1 and saw the point cloud as LG due to their distance from the stairs. However, once the user prepared themselves on the DS, The focus shifted away from the LG at the top of the stairs at time t2, even though point clouds indicated the presence of DS (Fig. 5(c)). This discovery is consistent with the gaze pattern described in prior research, which found that we periodically pay attention to glance toward the end of a pathway when we approach the transition region when descending stairs. The mean probability fusion strategy showed an unstable pattern of decision-making, albeit somewhat reducing the number of missed detections. The fused decision was able to identify the DS (the yellow line, which has the highest probability in a brief amount of time just before the first grey region), as seen in Fig. 5(d). However, the user's choice fluctuated between DS (the region in between) and LG (the first and second grey sections) as they got closer to the transition and started down the stairs. The accuracy and dependability of the decisions are questionable since this technique evaluates the forecasting from both modalities equally, regardless of the temporal link. Even while mean probability fusion showed some slight instabilities, our DTW-based fusion shows promise in resolving this problem and improving the accuracy of fused judgments.

It provided an accurate prediction and a valid time (roughly 1.3 seconds) for wearable robots to switch from control mode to DS for the same data, as indicated by the highlighted pink area in Fig. 5(e). This period began when the probability of DS becoming the greatest of all locomotion modes and ended when the user actually stepped on the DS (ground truth). This implies that even while the DTW-based fusion added a reasonable amount of latency, the forecasting nature of human intent and machine perception to direct the transition between locomotion modes meant that

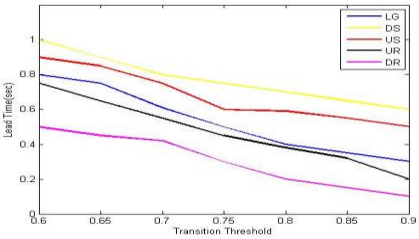
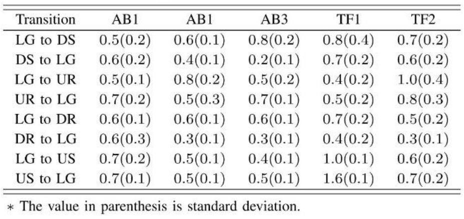


Figure 6: shows how changing the transition threshold affects the lead time for every mode of locomotion. Green square marks indicate levels producing false alarms, whereas red diamond markings indicate thresholds producing miss detection.

TABLE II

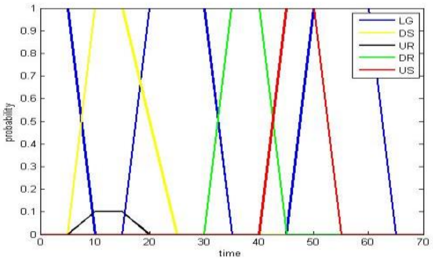
MEAN LEAD TIME (SECONDS) OF TRANSITION DECISIONS



Transition	AB1	AB1	AB3	TF1	TF2
LG to DS	0.5(0.2)	0.6(0.1)	0.8(0.2)	0.8(0.4)	0.7(0.2)
DS to LG	0.6(0.2)	0.4(0.1)	0.2(0.1)	0.7(0.2)	0.6(0.2)
LG to UR	0.5(0.1)	0.8(0.2)	0.5(0.2)	0.4(0.2)	1.0(0.4)
UR to LG	0.7(0.2)	0.5(0.3)	0.7(0.1)	0.5(0.2)	0.8(0.3)
LG to DR	0.6(0.1)	0.6(0.1)	0.6(0.1)	0.7(0.2)	0.5(0.2)
DR to LG	0.6(0.3)	0.3(0.1)	0.3(0.1)	0.4(0.2)	0.3(0.1)
LG to US	0.7(0.2)	0.5(0.1)	0.4(0.1)	1.0(0.1)	0.6(0.2)
US to LG	0.7(0.1)	0.5(0.1)	0.5(0.1)	1.6(0.1)	0.7(0.2)

∗ The value in parenthesis is standard deviation.

The choice was still made before wearable and body robots moved. Consequently, a crucial unanswered challenge is how to decide when wearable robots should change their locomotion method of control. Ideally, the control choice should occur before the transition phase. We may create a variable transition threshold that adjusts the decision time for locomotion mode transition in accordance with the requirements since the DTW outputs the fused probability for each locomotion mode. By averaging over all participants, Fig. 6 illustrates how the transition threshold affects decision fusion performance for each locomotor mode. For every mode, there was a definite trend showing that as the threshold rose, the decision lead time shrank. Setting a higher threshold was equivalent to having a filter with a longer window length since the transition threshold functioned as a filter, resulting in a longer delay (i.e., reduced lead time). Individual modes, however, differed in the amount of delay. In particular, the mean lead time dropped from 0.58 to 0.03 for LG, from 0.98 to 0.43 for US, from 1.10 to 0.63 for DS, from 0.79 to 0.07 for UR, and from 0.91 to 0.03 for US with the threshold ranging from 0.6 to 0.9. Furthermore, we saw that some mode transitions—such as those to LG with thresholds over 0.7 and DR with thresholds over 0.9—saw miss detection when the threshold was set high. On the other hand, shifting the threshold in the other direction—for example, to 0.6 for transitions to UR—might result in a false alert issue. Therefore, establishing the ideal lead time is necessary for managing prosthesis transitions, since neither a lengthy nor a short lead time is realistically adequate. The typical walking tempo, according to earlier research, is around 100 steps per minute, or 0.6 seconds for each step. Thus, it makes sense to extend the lead time.



An example series of DTW-based choices made throughout every transition condition in a continuous locomotion task is shown in Figure 7. (a) Probabilistic choices.

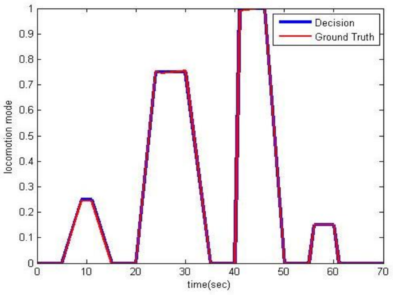


Fig 7(b) Deterministic choices following thresholding and the associated ground truth

of choice as near to 0.6 seconds as feasible to ensure a seamless and prompt transition for a prosthesis. By carefully changing the threshold, we were able to manage the lead time to sit within a zone of around 0.6 seconds, given the encouraging findings shown in Fig. 6. For this reason, in this study, we explicitly established the criterion at 0.6 for LG, 0.8 for US, 0.9 for DS, 0.7 for UR, and 0.8 for DR. Table II provides an overview of the average lead time of decisions for all participants and transition situations. It should be noted that participant TF1's lead time (1.6 seconds) to get from the US to LG was much longer than the intended lead time (0.6 seconds). This is because the TF1 participant used a step-by-step gait strategy instead of a step-over-step technique, which is how people with able bodies normally ascend stairs. This meant that the transition took longer to complete, resulting in a longer lead time. Therefore, we propose that while adjusting the transition threshold, it is important to take into account not only the locomotion mode but also the individual's gait pattern. When our study's created method is applied to the control of wearable assistive devices in various patient groups, it will make for fascinating future research. An example series of continuous judgments made in the continuous locomotion test utilizing the suggested DTW-based fusion with the aforementioned thresholds is displayed in Fig. 7. It is evident that the suggested fusion correctly and error-free identified every transition in the sample sequence. For the remaining trials, the same outcomes were likewise obtained.

IV. DISCUSSION

In this study, we have built a unique multimodal approach for lower-limb wearable robots to identify their intended mobility by the user. By including an educational human vision, the system was able to comprehend the user's intention beyond mere environmental perception. In order to optimize gaze integration's advantages, we investigated three distinct models. The findings demonstrated that rather than employing neural networks to directly filter visual input, they might provide more reliable and accurate forecasting by having them learn the attention distribution adaptively through the supervision of measured gaze information. Furthermore, a fusion approach based on DTW was implemented. By synchronizing temporal forecasting amongst several senses, this technique may provide dependable judgments regarding locomotion mode changes. additionally, the data analysis showed that the fusion provided a means of varying the lead time of transition decisions by a threshold, therefore granting robot control the adaptability to choose the appropriate timing of transitions in accordance with requirements. These encouraging results point to a great deal of promise for the future development of locomotion intent recognition for lower limb wearable robot control, combining human vision with ambient data. The developed system can be fully validated by real-time evaluations and comparisons with other existing methods, and it may be included in the control system for future work. Meanwhile, further on-board data (such as prosthesis kinematics and ground reaction force) may be included in the fusion-based locomotion mode identification system to give precise timing of transition for control. Finally, more practical situations including outdoor settings and halls with other pedestrians may be evaluated for reliable system validation in the future.

References

[1] Minhan Li, Graduate Student Member, IEEE, Boxuan Zhong, Edgar Lobaton, Member, IEEE, and He Huang, Senior Member, IEEE, "Fusion of Human Vision and Machine Perception for Forecasting Intended Locomotion Mode", IEEE TRANSACTIONS ON NEURAL SYSTEMS AND REHABILITATION ENGINEERING, VOL. 30, 2022

[2] V.Srinivas,Dr.Ch.Santhi Rani and Dr. T. Madhu, "Neural Network based Classification for Speaker Identification" International Journal of Signal Processing.Image Processing and Pattern Recognition., vol 7, No.1.pp.109-120.2014.

[3] B. Zhong, R.L.D.Silva, M., H. Huang, and E.Lobaton, "Environmental context forecasting for lower limb prostheses with uncertainty quantification," IEEE Trans.Autom.Sci.Eng., vol.18, no.2,pp.458–470,Apr.2021.

[4] K.Zhangetal, "Foot placement forecasting for assistive walking by fusing sequential 3D gaze and environmental context," IEEE Robot.Autom. Lett., vol.6, no.2,pp. 2509–2516, Apr. 2021.

[5] K.Min and J.J.Corso, "Integrating human vision into attention for egocentric activity recognition," in Proc. IEEE/CVF Winter Conf. Appl. Comput.Vis., Jan. 2021, pp. 1069–1078.

[6] Krishan Bhakta, Jonathan Camargo, Luke Donovan, Kinsey Herrin, and AaronYoung "Machine Learning Model Comparisons of User Independent & Dependent Intent Recognition Systems for Powered Prostheses," IEEE ROBOTICS AND AUTOMATION LETTERS, VOL.5, NO. 4, pp. 5393-5400, OCTOBER 2020

[7] Sarah Hood, Lukas Gabert, and TommasoLenzi, "Powered Knee and Ankle Prosthesis With Adaptive Control Enables Climbing Stairs With Different Stair Heights, Cadences, and Gait Patterns IEEE TRANSACTIONS ON ROBOTICS, VOL. 38, pp. 1430-1441 NO. 3, JUNE 2022

[8] K. Zhang, "A subvision system for enhancing the environmental adaptability of the powered transfemoral prosthesis," trans.Cybern.,2020,doi: 10.1109/TCYB.2020.2978216.

[9] B. Zhong, H. Huang, and E.Lobaton, "Reliable vision-based grasping target recognition for upper-limb prostheses," trans.Cybern., early access, Jun. 10, 2020,doi:10.1109/TCYB.2020.2996960.

[10] Kuangen Zhang, JingWang, Clarence W. de Silva, Life Fellow, IEEE, and Chenglong Fu "Unsupervised Cross-Subject Adaptation for Forcasting Human Locomotion Intent", IEEE TRANSACTIONS ON NEURAL SYSTEMS AND REHABILITATION ENGINEERING, VOL. 28, NO. 3, pp.646-657, MARCH 2020

[11] Yi-Xing Liu, Ruoli Wang, and Elena M. Gutierrez-Farewik, "A Muscle Synergy-Inspired Method of Detecting Human Movement Intentions Based on Wearable Sensor Fusion", IEEE TRANSACTIONS ON NEURAL SYSTEMS AND REHABILITATION ENGINEERING, VOL. 29, pp.1089-1098 Mar 2021.

[12] Ruijie Quan, Linchao Zhu, Member, IEEE, Yu Wu, Student Member, IEEE, and YiYang, "Holistic LSTM for Pedestrian Trajectory Forcastion", IEEE TRANSACTIONS ON IMAGE PROCESSING, VOL. 30, pp. 3229-3239, 2021.