Tiredness Detection for a Robotic Mobility Aid

Chuqiao Dong*
UBTECH NA RESEARCH & DEV CTR CORP
Pasadena, USA
chuqiao.dong@ubtrobot.com

James Kennedy*
Futronics NA Corporation
Pasadena, USA
james.kennedy@ubtrobot.com

Matthew K.X.J. Pan
Electrical and Computer Engineering Department
Queen's University
Kingston, Canada
matthew.pan@queensu.ca

Zhen Xiu
Futronics NA Corporation
Pasadena, USA
eddie.xiu@ubtrobot.com

Chengkun Zhang
Futronics NA Corporation
Pasadena, USA
chengkun.zhang@ubtrobot.com

Huan Tan
Futronics NA Corporation
Pasadena, USA
huan.tan@ubtrobot.com

***Abstract*—The proportion of elderly people in society is predicted to continue to rise in the coming decades. Mobility is a key aspect of many daily activities, but falls become an increasingly significant health risk with age. With the COVID-19 pandemic, many elderly users prefer or require assistive devices, rather than human support, in walking and carrying out daily tasks. However, prior work has shown that when using passive assistive mobility devices, fall risks can actually increase. This presents an opportunity for assistive robots to help maintain and improve the mobility of elderly users, with an additional emphasis on safety, made possible through sensing capabilities. In this paper, we present a computer vision system that detects the eye blink and face angle patterns for exhibiting signs of tiredness. In addition to the frame-based detection, we also introduce a time-window collation with a machine learning classifier. The system proposed here is critical in monitoring the user, performing real-time detection, and recommending they take a break if tiredness is detected. The overall system architecture and algorithmic details are presented, then a series of experiments are conducted to validate the performance of the approach.**

***Index Terms*—Human-robot interaction, assistive robotics, mobility aids, pandemic solution, computer vision**

I. Introduction

Throughout the coming decades, the proportion of elderly people in society is predicted to continue to rise [1]. This has the potential to cause societal issues, with the World Health Organization suggesting that one solution could be to help people maintain their capabilities as they age [2]. A key aspect of carrying out daily tasks is mobility. However, falls become an increasingly significant health risk with age [3]. Furthermore, with the COVID-19 pandemic, there are concerns over receiving human support, and preference or requirement to instead use assistive devices. It has been shown that using passive assistive mobility devices can actually lead to an increased fall risk [4]. This is therefore an opportunity for assistive robots, which can be equipped with sensors and software, enhancing safety features.

In this paper, we propose a system that uses a visual sensor (an RGB camera) to monitor the tiredness of a user while they are using a robotic mobility aid. Existing approaches in the field have primarily focused on monitoring the user gait in order to detect when they might be falling, e.g., [5]. With our approach, the goal is to instead monitor the user for signs of tiredness or fatigue, to suggest that they rest before they are so tired that they fall.

The work presented here takes inspiration from driver drowsiness detection and proposes a solution that is validated against the data from this domain. This paper contributes a novel system for detecting tiredness for a robotic mobility aid by taking advantage of the binary features of eye state and face angle. A series of machine learning classifiers are tested for the strongest approach. The overall goal of the work is to enhance the robustness of an algorithm for tiredness detection. While the results are promising, we also highlight the need for further exploration and data collection for this application.

II. Related Work

Various existing works have explored aspects of safety when using passive and robotic assistive mobility aids. For instance, [6] explores the height and width of the frame of a walker device to determine how this impacts balance, so that appropriate mechanical designs can be made. It has also been established that forearm support can be important for improved balance when using such devices [7].

When considering devices with increased sensing capabilities, a common research goal is to measure the gait state of the user. This can help with control of the device, with the user gait being used as an input for motor control algorithms [5]. These approaches tend to be somewhat reactive though, waiting for the gait state to begin falling, or for the user to have already released the device from their grip. In this paper we propose another solution that could augment these existing approaches to improve the safety of the user: a system that monitors the user for signs of tiredness and recommends a break before the problem becomes critical. This idea has overlaps with the concept of drowsy driving detection, studied with automobile users. Drowsiness has some distinctions from tiredness,

*These authors contributed equally to this work

Fig. 1. Example use case for the algorithm presented in this paper. A user walks with an assistive mobility robot. A camera is mounted on the robot that is intended to capture the head of the user.

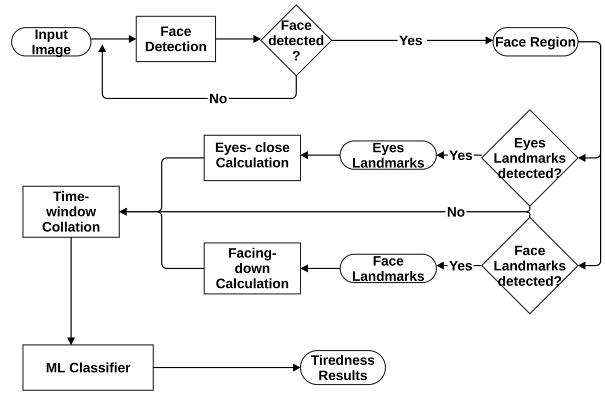

Fig. 2. Tiredness Detection Process

wherein a drowsy person may be physically capable, but not alert, whereas tiredness or fatigue includes physical tiredness. While we are primarily interested in the latter, there are clear overlaps in the concepts and possible detection methods which make drowsy detection algorithms of relevance.

Many drowsy driving detection algorithms rely on studying eye blinks as input features, then use the count of blinks over a minute to determine likely tiredness [8]–[10]. Machine learning has also been applied to drowsy driving detection through the use of classifiers processing action units (AU) from the facial action coding system (FACS) as feature inputs [11]. Given these prior works, it is clear that eye blinks and head position are important in the detection of drowsy behaviors, and have been used with promising results in driving scenarios.

III. Methodology

The use case for the system is that of a single user, controlling a robotic mobility device. The intended device is similar to a rollator or walker, as shown in the example in Fig. 1. The mobility device might vary in its specific design, but our system is based on having a frame that the user interacts with using the upper body. A camera can then be attached to this frame and focused on the user's face.

The Tiredness Detection Algorithm (TDA) takes an RGB image stream directly from a USB camera. The output of the TDA is a prediction of whether the user within the image(s) is showing signs of being tired or not. Given the use case targeting a mobility aid robot which assists one person at a time, an assumption is made that the largest face in the image will be regarded as the user and the TDA only processes this face. The approach takes inspiration from drowsy driving prior work [8], [10], [11] which finds that eye and head behavior are key indicators for tiredness detection.

Fig. 2 shows the detailed flow for the TDA. With the RGB image stream as an input, a face detection module is first used to determine whether the flow should continue. Only when a face is present will the algorithm be applied. Within the selected face region, an eyes-closed and a facing-down calculation are performed in parallel.

For the eyes-closed calculation, six landmark points are detected for each eye (Fig. 3A, B). The landmarks are detected using the dlib implementation of a 68-point facial landmark detection model [12], [13]. For each eye detected, an eye-height (D_h) is calculated based on the average distance between points 1 and 5, and points 2 and 4 ($(D_{p_{1,5}}+D_{p_{2,4}})/2$). An eye-width ($D_w$) is calculated from the distance between points 0 and 3 ($D_{p_{0,3}}$). The ratio between D_h and D_w indicates the eyes-closed ratio (R_{ec}), like the Eye-Apect Ration in [9]. When both eyes are detected, the eyes-closed ratio is the average value of both R_{ec}, like the Eye-Aspect Ratio in [9].

For calculating whether the person is facing down, six face landmarks are used (Fig. 3C, D). Similar to the eye detection, the landmarks are acquired from dlib with the same landmark detection model. After this, a projection from pixel-wise landmarks to a 3-D reconstruction is done through OpenCV [14]. Specifically, the pitch angle is used to represent A_{fd}, where facing downwards is negative.

R_{ec} and A_{fd} are both frame-based calculation results. However, the TDA is a real-time detection algorithm with an expected incoming frame rate between 15 and 30 frames per second (FPS). Thus, a time-window can be used to reduce the influence of noise and to measure behavior over time. For instance, a single frame may show the eye being closed, but this is not necessarily an indicator of tiredness. Better indicators include the eye being closed multiple times in a short time-frame, the eyes being shut for the entire period, or slow opening and closing eye speed, all of which often indicate tiredness [8], but can only be observed over time.

Prior literature was used to determine an appropriate time window. Eye closure speed is a strong predictor of fatigue, with alert drivers closing their eyes in less than 0.5 seconds [15]. The time window should be longer to capture multiple instances of this phenomenon and account for the difference in age (and likely movement speed) between those from drowsy driving studies and the intended user for the mobility aid. Consequently, the time window is selected to be 3 seconds long. With the eye-closed and face-down features calculated for each frame, a classifier can be trained over the time window to produce a prediction about the overall tiredness state.

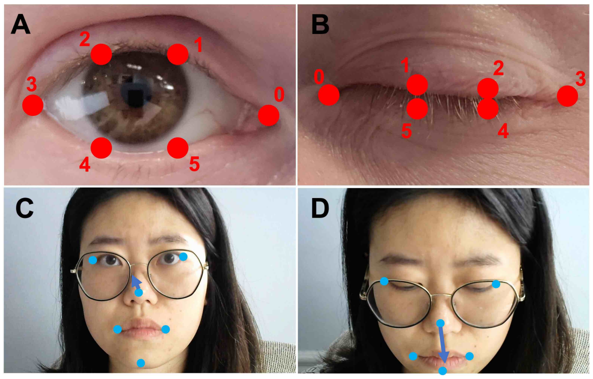

Fig. 3. (A-B) Landmark representation for each open eye and closed eye. (C-D) Landmark representation for not facing down and facing down.

TABLE I
(A). VALIDATION OF TDA FEATURES. (B). PERFORMANCE METRICS FOR CLASSIFIERS AT PREDICTING TIREDNESS.
TOTAL FRAMES (*total*), TRUE POSITIVE (*TP*), FALSE NEGATIVE (*FN*), FALSE POSITIVE (*FP*), RECALL (*Re*), PRECISION (*Pr*) AND F1 SCORE (*F1*) STATISTICS. SUPPORT VECTOR MACHINE (*SVM*), MULTI-LAYER PERCEPTRON (*MLP*), RANDOM FOREST (*RF*) CLASSIFIERS.

(A)	Total	TP	FN	FP	Re	Pr	F1
T_{eye}	2423	913	90	211	0.91	0.81	0.86
T_{face}	3656	484	62	280	0.89	0.63	0.74

(B)	Total	TP	FN	FP	Re	Pr	F1
SVM	92	36	12	1	0.75	0.97	0.85
MLP	92	35	13	1	0.73	0.97	0.83
RF	92	34	14	1	0.71	0.97	0.82

IV. EXPERIMENTS

A. Datasets

For validating the approach, the first step is to verify that the face detection and landmark detection can be used in conjunction with the devised algorithms to provide accurate predictions for the eye open/closed state and the face down/not state. Only one dataset was found that had explicit labels for the eye being open or closed that also shows the full face: the Closed Eyes in the Wild (CEW) dataset [16]. This dataset has 2423 images in total, with 1192 closed eye images and 1231 open eye images. Almost all images are of different people.

For the face angle, the BIWI dataset was selected as this contains a large number of images and annotations for calculating the head angle [17]. From the head angle, the ground truth label of whether the head is facing down can be calculated through OpenCV reconstruction [14] using the same head angle threshold as for prediction.

After validation of the component parts, the overall tiredness prediction algorithm can be tested. For this full evaluation, The University of Texas at Arlington Real-Life Drowsiness Dataset (UTA-RLDD) was selected as it contains a large volume of data, with natural (not acted) displays of drowsiness [10]. Currently, no dataset exists that serves the exact use-case of the work presented here: elderly people using mobility devices with varying tiredness levels. This dataset was determined to be the closest publicly available alternative. However, UTA-RLDD contains multiple drowsiness levels, which do not all fit the use case investigated in this work. Furthermore, the drowsiness levels of the dataset are self-reported, only achieving 57.8% accuracy from observers watching the videos. To address these issues, a subset of the videos were used for the analysis here. Only the first fold of the data was used as this had the highest human judgment accuracy [10]. The videos were then annotated for segments that were considered to be extremely drowsy. A second-coder verified these annotations; any segments with disagreement were rejected. The resulting data consists of videos from 11 individuals, ranging from 4 seconds to 62 seconds in length, for a total of about 14 minutes (25369 frames). This same time length was extracted randomly from the 'alert' videos, resulting in 24492 frames, to create an approximately balanced set of drowsy vs. alert data.

B. Feature Algorithm Validation

With the entire workflow (Fig. 2) implemented on the assistive robot using an NVIDIA Jetson NX board, TDA is able to perform real-time detection. Specifically, TDA uses 11.67% of the CPU computing power with a detection rate of 27 FPS. Thus, it is a light algorithm that is able to quickly respond when the user shows evidence of tiredness while using the assistive robot. The performance of TDA is first validated using frame-wise tiredness detection results T_{eye} and T_{face}. Fig. 3 show representative results for eye and face landmark detection, respectively, with the results summarized in Table I(A). The goal is to establish the accuracy of the algorithms that are being used to generate the features for the later classification stage.

The results for the eye open/close validation, T_{eye}, are calculated based on the CEW dataset. No face is detected in 234 (9%) of the frames. For the remaining frames, the recall and precision are satisfactory, confirming that the eye state can be predicted somewhat reliably using the proposed algorithm.

Facing down detection T_{face} has been validated based on a total of 3656 frames (766 facing down and 2890 not). No face was detected in 845 (23%) frames. In the validation test dataset, facing down data accounts for 21% of the data. An ideal dataset for validation would consist of images explicitly and accurately labeled with 'facing down' or not, but no publicly available dataset was found matching this criteria. Nevertheless, the F1 score of 0.74 indicates that there is a strong agreement between these two algorithms, which provides validation for using this implementation.

C. Tiredness Detection Results

With validation of the component features, we now explore the application of these features to the tiredness detection classification. As described previously, no ideal dataset was found so this section considers a manually crafted subset of the UTA-RLDD. First, the T_{eye} and T_{face} features were extracted for each frame of the dataset. Then, a non-overlapping 3 second time-window was applied. Note that in the real-time implementation, a sliding window is used instead. This results in a total of 221 videos in the drowsy class and 235 videos in the alert class.

Given the relatively small amount of data, a number of simple classifiers were considered: Multi-Layer Perceptron [18], Support Vector Machine (SVM) [19] and Random Forest [20]. Each algorithm was implemented with grid search to find optimal hyper-parameters. A random training/test split of 80/20 was applied. Table I(B) shows the performance of the different classifiers. While all classifiers have similar performance, the SVM has the best performance overall, achieving an F1 score of 0.85 and an accuracy of 86%. This compares favorably to the accuracy that would be achieved with a majority class baseline of 52%. The final hyper-parameters for the SVM were as follows: Gaussian radial basis function (RBF) kernel, $\gamma = 0.001$, $C = 100$.

The raw data (i.e., before thresholds were applied to determine a binary eye open/closed or face down/not state) from the T_{eye} and T_{face} calculations were also tested. The performance using these raw values was lower in all cases, both with and without scaling applied. This provides support for the utility of the T_{eye} and T_{face} algorithms as designed in Sec. III.

V. Discussion

The results from the previous section provide validation for the proposed eye-closed and facing-down algorithms and indicate that satisfactory performance can be achieved on the overall classification of tired behavior. However, there are several limitations of the work here that must be addressed to further advance the possibility of using such an algorithm in a real-world, safety critical environment. Firstly, the data used for the experiments contain mostly younger participants. While this provides a starting point for performing experiments and developing an algorithmic approach, there are open questions surrounding how well the findings would transfer to the behavior and facial features of the elderly.

A key next step for this research is to collect data of the elderly population using assistive mobility devices and test the approach to see whether it remains robust. This is not an easy task given the challenge in collecting high-quality data. Designing a data collection to intentionally cause fatigue would also have many possible risks.

The system introduced here focused on computer vision based detection. In real-life applications, more system information can be collected through sensor fusion, e.g., devices such as IMUs could be used to detect user activity. During known tasks, and with repeated use from the same user, expectations about walking speed and other motions could be established. These factors could then be incorporated into the classifier that is proposed in this paper to help improve the performance.

In this paper, we take inspiration from drowsy driving detection to propose and test an algorithm for tiredness detection for a mobility aid robot. It is found that the proposed algorithm achieves satisfactory performance and enhances the robustness of the predictions made from the raw features. However, further testing is needed with the targeted elderly population before real-world use. We discuss the challenges in collecting appropriate data for this purpose but highlight the promise that has been shown in applying computer vision and machine learning algorithms to this application, which would have a positive impact on people's lives.

Acknowledgment

The authors would like to acknowledge the Industrial Design, Hardware and Testing teams of Futronics NA Corporation for supporting this work during development.

References

[1] World Health Organization, "Ageing and health," 10 2021.
[2] ——, "10 facts on ageing and health," 5 2017.
[3] M. E. Tinetti, M. Speechley, and S. F. Ginter, "Risk Factors for Falls among Elderly Persons Living in the Community," *New England Journal of Medicine*, vol. 319, no. 26, pp. 1701–1707, 12 1988.
[4] H. Bateni and B. E. Maki, "Assistive devices for balance and mobility: Benefits, demands, and adverse consequences," *Archives of Physical Medicine and Rehabilitation*, vol. 86, no. 1, pp. 134–145, 1 2005.
[5] D. Zhao, J. Yang, M. O. Okoye, and S. Wang, "Walking Assist Robot: A Novel Non-Contact Abnormal Gait Recognition Approach Based on Extended Set Membership Filter," *IEEE Access*, vol. 7, pp. 76 741–76 753, 2019.
[6] S. B. Thies, R. Russell, A. Al-Ani, T. Belet, A. Bates, E. Costamagna, L. Kenney, and D. Howard, "An investigation of the effects of walking frame height and width on walking stability," *Gait and Posture*, vol. 82, no. December 2019, pp. 248–253, 2020.
[7] C. Jayaraman, C. K. Mummidisetty, A. Loesch, S. Kaur, S. Hoppe-Ludwig, M. Staat, and A. Jayaraman, "Postural and Metabolic Benefits of Using a Forearm Support Walker in Older Adults With Impairments," *Archives of Physical Medicine and Rehabilitation*, vol. 100, no. 4, pp. 638–647, 2019.
[8] A. Islam, N. Rahaman, and M. A. R. Ahad, "A study on tiredness assessment by using eye blink detection," *Jurnal Kejuruteraan*, vol. 31, no. 2, pp. 209–214, 2019.
[9] T. Soukupova and J. Cech, "Eye blink detection using facial landmarks," in *21st computer vision winter workshop, Rimske Toplice, Slovenia*, 2016.
[10] R. Ghoddoosian, M. Galib, and V. Athitsos, "A realistic dataset and baseline temporal model for early drowsiness detection," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2019.
[11] E. Vural, M. Cetin, A. Ercil, G. Littlewort, M. Bartlett, and J. Movellan, "Drowsy driver detection through facial movement analysis," in *international workshop on human-computer interaction*. Springer, 2007, pp. 6–18.
[12] V. Kazemi and J. Sullivan, "One millisecond face alignment with an ensemble of regression trees," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2014, pp. 1867–1874.
[13] D. E. King, "Dlib-ml: A machine learning toolkit," *J. Mach. Learn. Res.*, vol. 10, p. 1755–1758, dec 2009.
[14] OpenCV, "Open source computer vision library," 2015.
[15] Q. Ji and X. Yang, "Real-time eye, gaze, and face pose tracking for monitoring driver vigilance," *Real-time imaging*, vol. 8, no. 5, pp. 357–377, 2002.
[16] F. Song, X. Tan, X. Liu, and S. Chen, "Eyes closeness detection from still images with multi-scale histograms of principal oriented gradients," *Pattern Recognition*, vol. 47, no. 9, pp. 2825–2838, 2014.
[17] G. Fanelli, M. Dantone, J. Gall, A. Fossati, and L. Van Gool, "Random forests for real time 3d face analysis," *Int. J. Comput. Vision*, vol. 101, no. 3, pp. 437–458, February 2013.
[18] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning internal representations by error propagation," California Univ San Diego La Jolla Inst for Cognitive Science, Tech. Rep., 1985.
[19] C. Cortes and V. Vapnik, "Support-vector networks," *Machine learning*, vol. 20, no. 3, pp. 273–297, 1995.
[20] T. K. Ho, "Random decision forests," in *Proceedings of 3rd international conference on document analysis and recognition*, vol. 1. IEEE, 1995, pp. 278–282.