Proceedings of the 2020 IISE Annual Conference
L. Cromarty, R. Shirwaiker, P. Wang, eds.

Recognition of driving distraction using driver's motion and deep learning

Abstract ID: 790170

Ziyang Xie, Li Li, Xu Xu
Edward P. Fitts Department of Industrial and Systems Engineering
North Carolina State University
Raleigh, North Carolina, 27695, U.S.A.

Abstract

In this study, we propose a new method for driving distraction recognition by integrating wearable sensing and deep neural networks. Twenty participants with wearable inertial sensors attached on their wrist performed five manual distraction tasks under audio instructions in a driving simulator. Based on the captured data, recurrent neural networks with long short-term memory (RNN-LSTM) were applied to classify the manual distraction tasks. Besides, for the aim of real-time distractions recognition, a wrist velocity-based action detector is designed. The results indicated that the overall F1-score for the distraction task classification is 0.85, which suggests that wearable sensing has good potentials for driving distraction recognition. The results also indicate that the wrist velocity threshold used in the action detector significantly affects the recall and precision.

Keywords

Driving safety, Machine learning, Kinematics analysis, Distraction detection.

Introduction

Driving distraction is a form of driver inattention defined as 'driver engagement in internalized thoughts' [1]. According to National Highway Traffic Safety Administration [2], more than three thousand people were killed by distracted driving in 2017 in the U.S. Effectively detecting and mitigating driving distraction can reduce the likelihood of fatal car accidents. In addition, in the era of automated driving, distraction detection is a key factor for the driving system to evaluate the timing of transfer between manual and auto-driving mode [3]. Therefore, driving distraction recognition has been an active research topic in transportation safety.

There are three major types of driving distractions: manual distraction, visual distraction, and cognitive distraction [4]. Several works related to visual and cognitive distraction recognition have been reported in recent years[5]. However, most of the studies used a fixed camera in front of drivers to record facial expressions and eye movement for distraction recognition, which could make drivers feel uncomfortable due to privacy concerns[6]. Therefore, researchers are seeking alternative ways to monitor driving behaviors. Manual distraction indicates that drivers take their hands off the wheel (e.g., drinking water and touching the control panel), which is typically a consequence of simultaneous visual and cognitive distraction [7]. Thus, monitoring a driver's manual behavior through non-camera-based methods in one possible way to recognize driving distraction without triggering privacy concerns.

To date, a variety of non-camera-based methods for monitoring manual distractions have been proposed. For example, current on-the-market driver assistance systems use torque sensors mounted on the steering wheel to detect 'hands-off the wheel' distractions[8]. However, only detecting 'hands-off the wheel' provides limited information regarding the specific type of manual distraction. This limitation can be solved by tracking drivers' body motions and classifying their body motions to specific distractive behaviors through the motion data [9]. Particularly, wearable inertial measurement units (IMUs), whose output is linear acceleration and angular velocity values on frames, show its advantages in driver's motion tracking: First, applying wearable IMUs does not require mounting specific devices in a car, so that vehicle retrofitting is not needed. Second, wearable sensors are individual-specific, not car-specific, so that wearable IMUs can measure the behaviors of those drivers who are driving multiple cars.

Different machine learning algorithms are widely used in wearable IMU-based human activity recognition (HAR) problems. Previous research applies K-nearest neighbor (KNN) and support vector machine (SVM) to classify human activities [10, 11]. KNN and SVM are suitable for the classifications that are based on the unique body motion features in the spatiotemporal domain (e.g., touching head vs. reaching a child in the back seat). However, they are not robust enough for classifying similar manual distraction behaviors (e.g., drinking vs. texting) [12] because the differences in the spatiotemporal domain among these behaviors are subtle. To improve recognition robustness, deep neural networks, such as convolutional neural networks (CNNs) and recursive neural networks (RNNs) have been recently applied to HAR problems [13, 14]. While CNNs are suitable to extract features from images, RNN performs better when the inputs are highly dynamic time-series [15]. In HAR problems, promising results based on RNNs are recently reported [16, 17], which revealed that RNNs could be a better choice when the input is the time-series collected from IMUs. In this study, we apply long short-term memory (LSTM), a commonly used RNN structure, to driving distraction recognition problems since LSTM is well known for its efficiency and classification accuracy in time series data [13, 18].

In this paper, we proposed a method for driving distraction recognition by integrating wearable sensing and recursive neural networks with long short-term memory (RNN-LSTM). The proposed method includes data collection, data post-processing, network training, and real-time distraction detection and classification.

Method

Participants

We have recruited 20 participants (8 females and 12 males, all with valid driver licenses) aged from 25 to 55 years old. Among these 20 participants, data from 15 participants (9 males and 6 females) were used for training the RNN-LSTM model, and the data from the other five participants (3 males and 2 females) were used for model validation.

Apparatus

Experiments were conducted in an RTI driving simulator (Ann Arbor, MI), which is a mock-up vehicle cockpit including a steering wheel, a brake, seats, etc. The simulator is fixed-base with three plasma displays placed approximately 2 meters away from the driver (Figure 1). The driving environments and traffic scenarios generated by RTI SimCreator and SimVista software are displayed in front of the driver. The simulated driving scenarios include both city roads and highway. City roads consist of straight road, intersections, stop with a traffic light, curved road, while highway consists of curved and straight road. A cellphone, a touch panel, markers, and a cup are placed near the seat as distraction tools. IMUs (MVN, Xsens technologies, the Netherlands) are attached to participants' right wrist. The raw output of the IMU includes linear acceleration and angular velocity in 3D at 100 Hz. Three camcorders synchronized with the IMUs were placed around the driving simulator to record participants' driving behavior during the experiment for the need to code to the specific manual distraction behaviors later manually.

Figure 1: Screenshots from the recorded video from the back, side, and front camera. The participant is driving on the highway on the RTI Simulator.

Distraction tasks

Each participant performed five common driving distractions: talking on a cellphone (Phone), drinking water (Drink), taking a marker from the passenger's seat, and placing it into a cup holder (Marker), touching the infotainment system screen (TouchScreen), texting (Texting). These five distraction tasks have been previously reported as those among the most common distraction tasks [19].

Experiment protocol

Before the experiment, participants needed to complete informed consent as well as a demographic questionnaire. Each participant was assigned for six driving trials. Each driving trial lasted for approximately 15 minutes. In five of six driving trials, verbal instructions were given by a speaker notifying participants to perform different distractions. A verbal instruction was given in about every 30 seconds. In the rest driving trial, no verbal instruction was provided, and the participants' behavior in this trial served as the baseline.

The structure of RNN-LSTM

RNN-LSTM was configured by Keras (ver.2.3) in Python 3 (ver.3.6). Raw data collected from the IMU mounted on the right wrist first went through the LSTM layer, where its features and essential information were captured (Figure 2). The drop out layer next to the LSTM layer is to avoid overfitting problems and to reduce redundancy. The drop out rate in the drop out layer determines the rate that this neural network randomly drops its data. In hidden layers, a drop out rate closer to 0.5 is more efficient as reported in prior research[20]. After the dropout layer, the data flows into a fully connected dense layer where classifications are made. LSTM dimension is typically between the number of classes and the size of data [21]. After preliminary tests, the dimension of 100 is adopted for balancing the learning efficiency and classification robustness. This neural network structure takes wrist acceleration values in x,y,z directions as input. The output is the classified manual distraction behavior. This RNN-LSTM was then trained by the ground-truth training data.

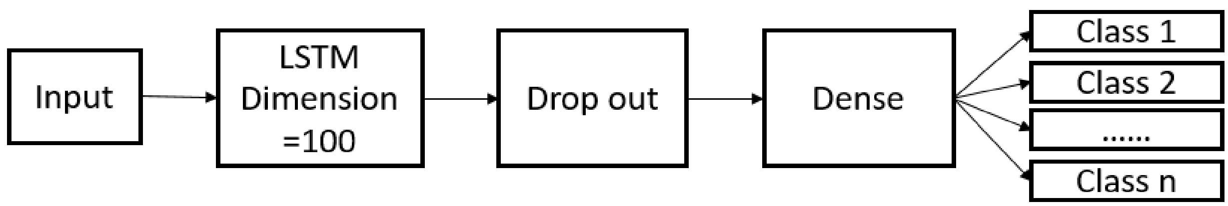

Figure 2: Structure of the proposed RNN-LSTM for driving distraction recognition

Data segmentation for RNN-LSTM training

For LSTM model training, we first set the ground-truth by manually segment the wrist motion during driving distraction tasks using the recorded video. Particularly, the frame when a driver's hand detached from the steering wheel was set as 'start frame' (S.F.), and the frame that ten seconds after S.F. was set as 'end frame' (E.F.). There were 1679 driving distraction tasks observed and extracted manually (Figure 3). All the wrist motion data during various distraction tasks were then interpolated to 300 frames to increase the learning rate and recognition efficiency.

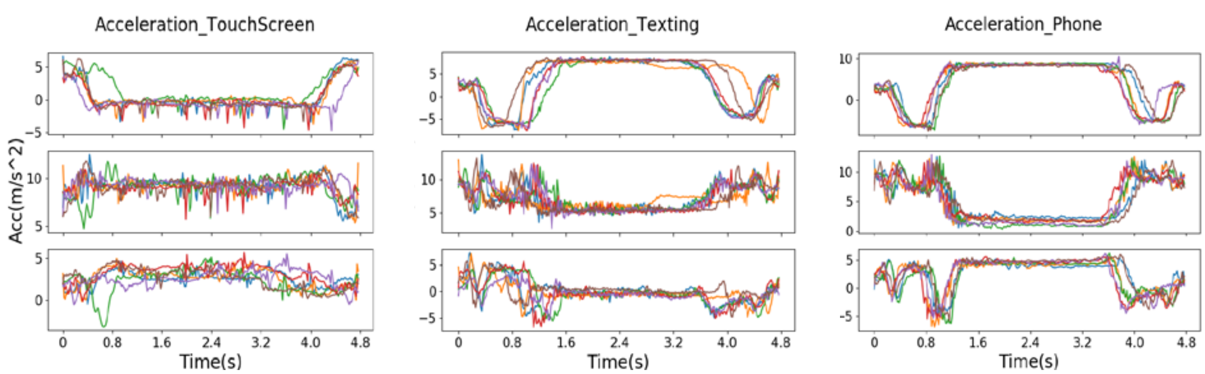

Figure 3: An example of measured wrist linear acceleration of the right wrist during distracted driving from a participant. The curves in different colors indicate seven repetitions of a distraction task; three of five distractions are shown in the figure.

Real-time distraction detection

Note that the purpose of this research is to recognize driving distractions in real-time. Thus, we needed to develop a method to automatically segment the real-time IMU data stream and send this segmented data into the trained RNN-LSTM. Given the participants move their hands off the wheel at the beginning of a manual distraction task with an increased wrist linear velocity, we could use wrist velocity as an indicator for detecting potential manual distraction.

The linear wrist velocity can be acquired from calculating the integral of the linear wrist acceleration, which is the raw output of the wrist-mounted IMU.

Based on the linear velocity, an action detector was designed as a 10-second moving window. First, the average value of the first 50 frames was subtracted from all the data points within this moving window (to eliminate drifting noise). The action detector then made an integration within the window length to get the velocity change within 1000 frames (10s). From the preliminary test, a velocity of 0.3 m/s to 0.5 m/s is an appropriate threshold since most of the distraction velocities are higher than 0.5 m/s, and normal motion velocities are lower than 0.3 m/s. The threshold was set and compared with the velocity change. If the velocity change is larger than the threshold, the detector will report a potential distraction, and send the raw data in the moving window to RNN-LSTM for distraction task classification.

Results and discussion

When the trained RNN-LSTM is applied to the manually segmented data in the validation dataset, the overall F1-score for different distracted driving tasks ranges between 0.92 and 0.99. Precision, recall, and F1-score are calculated [22, 23] and shown in Table 1. When the action detector is combined with the trained RNN-LSTM for real-time driving distraction recognition, the overall F1-score for different distracted driving tasks ranges between 0.75 to 0.95 (Table 2). Specifically, greater recall in distraction recognition (as shown in Table 1 and 2) indicates that more actions are correctly predicted and a greater precision indicates less false predictions are made.

Table 1: RNN-LSTM Model’s performance on manually segmented data evaluated by precision, F1-score, and number of samples

Distraction task	Recall	Precision	F1-score	Number of samples
Phone	0.88	0.98	0.93	109
Text	0.98	0.99	0.99	105
Drink	0.95	0.96	0.96	107
TouchScreen	0.96	0.89	0.92	108
Marker	0.94	0.90	0.92	109

Table 2: RNN-LSTM Model’s performance on real-time data evaluated by precision, F1 score, and number of samples

Distraction Task	Recall	Precision	F1-score	Number of samples
Phone	0.97	0.93	0.95	109
Texting	0.94	0.67	0.78	105
TouchScreen	0.66	0.88	0.75	107
Drink	0.94	0.89	0.92	108
Marker	0.92	0.70	0.80	109

The results in Table 1 and 2 are the averages across all participants in the validation dataset. Table 3 further reveals the classification accuracy of the proposed method among each participant. The results indicate that F1-score varies from one participant to another. One potential reason for this inter-participant variability is that a simple wrist velocity-based action detector is not suitable for people with different driving characters. For example, the threshold of the action detection is set to 0.4 m/s based on suitable-for-most principle. However, participant #2 moved much faster than other participants when turning the steering wheel. A faster turning generates a greater wrist velocity, which triggers the action detector to send a non-distracted behavior into the neural networks for classification. A follow-up analysis showed that these non-distracted behaviors were mainly categorized to ‘Texting’ and ‘Marker’, which resulted in the low F1-score for these two categories in Table 2.

Table 3: Prediction rates in different participants.

Participant # in the validation dataset	Recall	Precision	F1-score	Number of samples
1	96%	87%	0.91	116
2	83%	74%	0.78	114
3	93%	90%	0.92	29
4	98%	82%	0.90	109
5	77%	80%	0.79	140

Particularly, the threshold of the action detector affects both recall and precision. In Figure 4, recognition results under different threshold values are summarized. With an increased threshold value, the precision increases while the recall decreases, which is consistent with our expectation that detectors with a higher threshold make more conservative predictions while missing more of the ground-truth distractions. In addition, the F1 score, which indicates the prediction accuracy in terms of both recall and precision, fluctuates in an acceptable range between 0.80 and 0.85.

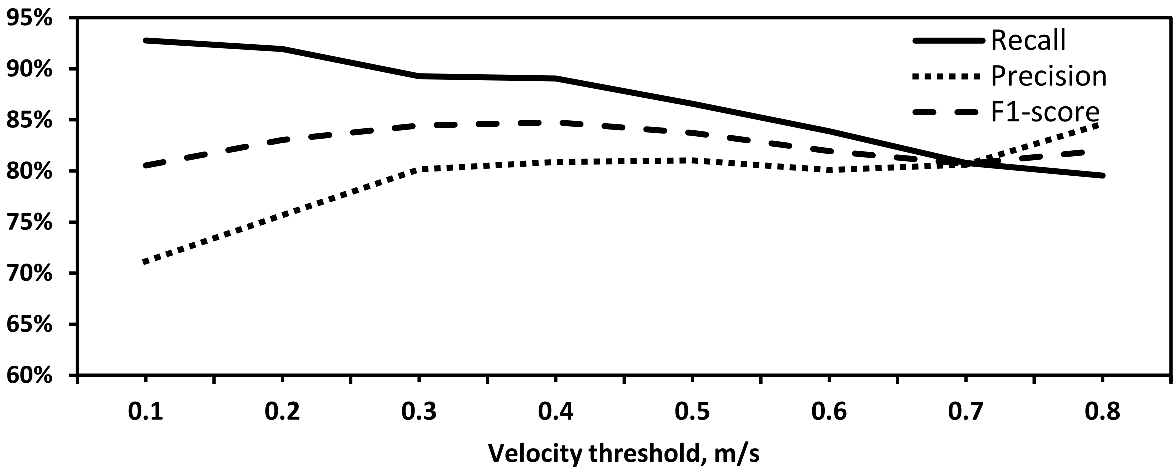

Figure 4: Overall recall, precision, and F1-score of the prediction under different velocity values.

Limitations and Future work

In summary, the results show a great potential of using and RNN-LSTM to perform real-time driving distraction recognition. However, some limitations of this study need to be addressed. First, the recall on 'TouchScreen' is not ideal, which indicates that the current method tends to miss many 'TouchScreen' behaviors. Therefore, instead of using linear velocity, which is small during 'TouchScreen' distractions, as the detection trigger, harnessing other kinematics data such as angular velocity may be necessary for detecting more 'TouchScreen' behaviors. Second, our current method performs better on manually segmented data (Table 1) than real-time detecting data (Table 2), which suggests that real-time data segmentation should be further studied to improve the overall robustness.

Acknowledgments

This manuscript is based upon work supported by the National Science Foundation under Grant No 1822477.

References

[1] Regan, M.A., C. Hallett, and C.P. Gordon, *Driver distraction and driver inattention: definition, relationship and taxonomy.* Accid Anal Prev, 2011. **43**(5): p. 1771-81.
[2] Administration, N.H.T.S. *Distracted Driving*. 2017; Available from: https://www.nhtsa.gov/risky-driving/distracted-driving.
[3] Nilsson, J., P. Falcone, and J. Vinter, *Safe Transitions From Automated to Manual Driving Using Driver Controllability Estimation.* IEEE Transactions on Intelligent Transportation Systems, 2015. **16**(4): p. 1806-1816.

[4] NHTSA, *Traffic Safety Facts Research Notes 2016: Distracted Driving*. 2016.
[5] Li, N. and C. Busso. *Analysis of facial features of drivers under cognitive and visual distractions*. IEEE.
[6] Eckhoff, D. and C. Sommer, *Driving for big data? Privacy concerns in vehicular networking.* IEEE Security & Privacy, 2014. **12**(1): p. 77-79.
[7] Regan, M.A., C. Hallett, and C.P. Gordon, *Driver distraction and driver inattention: Definition, relationship and taxonomy.* Accident Analysis & Prevention, 2011. **43**(5): p. 1771-1781.
[8] Lombrozo, P.C., *Detecting driver grip on steering wheel*. 2014, Google Patents.
[9] Prayudi, I. and D. Kim. *Design and implementation of IMU-based human arm motion capture system*. in *2012 IEEE International Conference on Mechatronics and Automation*. 2012. IEEE.
[10] Kaghyan, S. and H. Sarukhanyan, *Activity recognition using k-nearest neighbor algorithm on smartphone with tri-axial accelerometer.* International Journal of Informatics Models and Analysis (I.J.I.M.A.), I.T.H.E.A. International Scientific Society, Bulgaria, 2012. **1**: p. 146-156.
[11] Jiang, L., et al., *Safedrive: Detecting distracted driving behaviors using wrist-worn devices.* Proceedings of the A.C.M. on Interactive, Mobile, Wearable and Ubiquitous Technologies, 2018. **1**(4): p. 144.
[12] Kotsiantis, S.B., I. Zaharakis, and P. Pintelas, *Supervised machine learning: A review of classification techniques.* Emerging artificial intelligence applications in computer engineering, 2007. **160**: p. 3-24.
[13] Hammerla, N.Y., S. Halloran, and T. Plötz, *Deep, convolutional, and recurrent models for human activity recognition using wearables.* arXiv preprint arXiv:1604.08880, 2016.
[14] Zeng, M., et al. *Convolutional neural networks for human activity recognition using mobile sensors*. in *6th International Conference on Mobile Computing, Applications and Services*. 2014. IEEE.
[15] Ordóñez, F. and D. Roggen, *Deep convolutional and lstm recurrent neural networks for multimodal wearable activity recognition.* Sensors, 2016. **16**(1): p. 115.
[16] Liu, J., et al. *Spatio-temporal lstm with trust gates for 3d human action recognition*. in *European Conference on Computer Vision*. 2016. Springer.
[17] Guan, Y. and T. Plötz, *Ensembles of deep lstm learners for activity recognition using wearables.* Proceedings of the A.C.M. on Interactive, Mobile, Wearable and Ubiquitous Technologies, 2017. **1**(2): p. 11.
[18] Hochreiter, S. and J. Schmidhuber, *Long short-term memory.* Neural computation, 1997. **9**(8): p. 1735-1780.
[19] Stutts, J., et al., *Driver's exposure to distractions in their natural driving environment.* 2005. **37**(6): p. 1093-1101.
[20] Srivastava, N., et al., *Dropout: a simple way to prevent neural networks from overfitting.* The journal of machine learning research, 2014. **15**(1): p. 1929-1958.
[21] Heaton, J., *Introduction to neural networks with Java*. 2008: Heaton Research, Inc.
[22] Buckland, M. and F. Gey, *The relationship between recall and precision.* Journal of the American society for information science, 1994. **45**(1): p. 12-19.
[23] Olson, D.L. and D. Delen, *Advanced data mining techniques*. 2008: Springer Science & Business Media.