

Analysis of yawning behaviour in spontaneous expressions of drowsy drivers

Zhuoni Jie*, Marwa Mahmoud†, Quentin Stafford-Fraser‡, Peter Robinson§, Eduardo Dias¶ and Lee Skrypchuk||

**Department of Industrial Engineering
Tsinghua University, Beijing, China*

Email: jiezn14@mails.tsinghua.edu.cn

† ‡ §*The Computer Laboratory*

University of Cambridge

Cambridge, United Kingdom

Email: {marwa.mahmoud, peter.robinson}@cl.cam.ac.uk, quentin@pobox.com

¶*Jaguar Land Rover*

Coventry, United Kingdom

Email: {edias, lskrypch}@jaguarlandrover.com

Abstract—Driver fatigue is one of the main causes of road accidents. It is essential to develop a reliable driver drowsiness detection system which can alert drivers without disturbing them and is robust to environmental changes. This paper explores yawning behaviour as a sign of drowsiness in spontaneous expressions of drowsy drivers in simulated driving scenarios. We analyse a labelled dataset of videos of sleep-deprived versus alert drivers and demonstrate the correlation between hand-over-face touches, face occlusions and yawning. We propose that face touches can be used as a novel cue in automated drowsiness detection alongside yawning and eye behaviour. Moreover, we present an automatic approach to detect yawning based on extracting geometric and appearance features of both mouth and eye regions. Our approach successfully detects both hand-covered and uncovered yawns with an accuracy of 95%. Ultimately, our goal is to use these results in designing a hybrid drowsiness-detection system.

Keywords-yawning detection; driver drowsiness; facial expression; hand touches

I. INTRODUCTION

The increasing number of traffic accidents due to fatigued driving has become a serious social problem. According to the U.S. National Highway Traffic Safety Administration, for the 5-year time period between 1989 and 1993, 100,000 crashes per year (1.6% of 6.3 million) were reported by the police with drowsiness identified as the leading cause [1]. Data for 2010 from the UK Department of Transport shows that fatigue contributed to 20% of the total number of road accidents [2]. Drowsiness is known to impair drivers' abilities of reaction, awareness, information processing and short-term memory. Therefore, developing a system for monitoring a driver's level of vigilance, alerting them when in a drowsy state and suggesting a rest when necessary could bring a significant reduction in such road accidents.

Traditionally, various car driving parameters obtained

from the Controller Area Network (CAN) bus (longitudinal and lateral acceleration, steering wheel angle, ... etc.) and various physiological signals (Electroencephalography, Electrocardiography and Electrooculography) have been used to detect driver drowsiness level [3]. However, integration of these devices into vehicles is sometimes intrusive due to the nature of measuring physiological signals. It can also be expensive, inaccurate, even impractical in response to complicated driving behaviours.

Eye movement patterns and yawning are two important cues to develop a natural and non-intrusive driver fatigue detection system. Most previous research has focused on eye behaviour alone to detect driver drowsiness. Yawn detection is challenging in natural driving scenarios because of varying illumination levels, inter-personal differences, and complicated expressions [4]. Previous research on yawning mostly measures and classifies the mouth opening states, which can fail when the hand covers the mouth during yawning. Most of the previous work was also developed using datasets of posed yawns, which are different from spontaneous yawns.

In this paper, we analyse yawning behaviour in a dataset of spontaneous expressions of drowsy drivers, and propose a new method to detect yawning that combines geometric and appearance features of the mouth and eye areas. The main contributions of this paper can be summarized as follows:

1. We present a spontaneous dataset of videos of drivers collected in simulated driving scenarios, recruiting both real sleep-deprived and alert volunteers. We present the labelling of different cues related to yawning, namely: yawning (both mouth covered and uncovered), speaking, and other face touches.

2. We propose that face touches can be used as a novel modality in drowsiness detection by presenting the correlation between hand-over-face touches, yawning and

drowsiness.

3. We propose a vision-based approach fusing both geometric and appearance features to detect mouth-covered yawns and uncovered yawns in natural expressions.

II. RELATED WORK

Yawning is an involuntary behaviour that is symbolized by wide open mouths. There have been several attempts to detect yawning, most of which focus on the measure of mouth opening. For example, Qiang Ji et al. [5] detected yawning using the openness of the mouth represented by the ratio of mouth height to width. Similarly, Tiesheng Wang et al. [6] also used this method to define mouth openness, and detected yawning if the ratio is above 0.5 in more than 20 frames in videos with a frame rate of 30 frames per second. Shabnam Abtani et al. [7] detected yawning by first detecting the yawn component in the face independent of the mouth location, and then using mouth location to verify the validity of the detected component. Wang et al. [8] extracted the mouth region of interest using color analysis, then segmented skin and lip pixels, and got lip features by connected component analysis. Then they took the mouth region geometric features to make up an eigenvector as the input of a backpropagation neural network, acquiring the output of three mouth states: normal, yawning, or talking. Lu et al. [9] located the driver's face region using the difference between two images, and then located the midpoint of the nostrils using a directional integral projection method. The yawn is then detected by calculating the vertical distance between the midpoint of nostril and the chin. However, these algorithms use geometric features only, which make them difficult to differentiate between normal mouth opening (e.g. speaking, smiling, coughing) and yawning. There are also some other disadvantages in geometric-based methods. First, they are highly dependent on the geometry of drivers' faces, making them sensitive to different face poses and their results are very much dependent on their individual implementations. Second, mouth geometric features can be inaccurate in some cases due to different head orientations, illumination conditions, existence of a beard or mustache, or acute lip movement. Third, most of these methods are not able to detect hand-covered yawning since the hand covers most of the mouth corners.

There is also some work in which other specific features are extracted to train machine learning algorithms. Saradadevi et al. [10] used a cascade of classifiers to detect the mouth, then trained on yawning images using an support vector machine (SVM). Fan et al. [11] proposed a method using a Gravity-Center template to detect the face, and then used Gabor wavelets of mouth corners and Linear Discriminant Analysis to detect yawning. Azim et al. [12] presented a system in which the yawning detection method used Viola-Jones face detection to locate the face, extracted the mouth window, and then searched for the lips through



Figure 1. Sample images of yawns from our spontaneous video dataset.

spatial fuzzy c-means clustering. Ibrahim et al. [13] proposed a method to detect yawning based on mouth opening, mouth covering and facial feature distortions. Then classification of local binary patterns (LBP) features extracted from the mouth when covered by a hand is used for mouth-covered detection. However, these previously-mentioned algorithms using appearance-based features have complex algorithms and classifiers, making them unrealistic to implement with a camera system inside the car with low processing power, and require training on datasets with large samples.

III. DATASET

There are a few publicly available datasets on yawning. However, most of them are not suitable for our problem of detecting spontaneous yawns in a car driving scenario. Some of them [14] contain only still images of yawns, which are not as reliable as videos because a single frame of an open mouth can be due to either speaking or yawning. Some published datasets such as yawDD [15] have the advantages of changing lighting conditions, camera type and position. However, the videos collected contain acted non-spontaneous yawns, with participants performing yawning while they are not in a genuine drowsy state.

Therefore, in our work we collected a video dataset containing spontaneous expressions of sleep-deprived drivers in a driving simulator. Twelve participants were recruited and videos are collected in two phases. For phase 1, which is the drowsy state phase, each participant was deprived of sleep and caffeine for 24 hours before taking the driving task. For phase 2, which is the alert phase, the participants received enough rest before the driving task. For each phase, every participant completed a driving task of 90 minutes, consisting of a manual driving session of 45 minutes and an autonomous driving session of 45 minutes. The participants were 7 males and 5 females with ages ranging between 19 and 56.

A. Labelling

Labelling was carried out using the ELAN video annotation tool [16]. A researcher was asked to label the videos based on the following instructions:

- Mouth Action: coded as one label per frame. Labels are mutually exclusive, i.e. one label is permitted per frame. Labels are: 1) Yawn. 2) Speak.

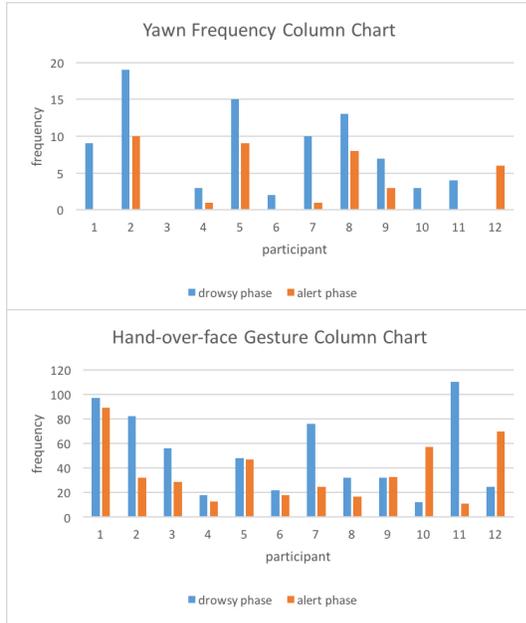


Figure 2. Analysing yawning and hand-over-face frequencies in drowsy versus alert states. The column chart shows that both can be used as cues for drowsiness detection algorithms.

- Hand Action: coded as one label per frame. Labels are mutually exclusive. Labels are: 1) Hand-inside-face gesture. 2) Hand-outside-face gesture.
- Noise State: coded as one label per frame. Labels are mutually exclusive. Labels are: 1) Away: when the driver’s face is outside of the camera view. 2) Asleep: if the driver actually falls asleep. This label is important in the subsequent analysis: we are interested in “drowsy” rather than “asleep” drivers, since we are looking for the early signs of fatigue.

B. Dataset analysis

After labelling, we performed some statistical analysis on the labelled data. By analysing the labels, we got a total of 123 yawns from these videos, with the mean duration of each yawn being approximately 6.28 seconds. We analysed the frequency of yawning and face touches in the drowsy versus the alert recording phases. As shown in Fig. 2, the yawning frequency in the drowsy phase (phase 1) is significantly higher than the frequency in the alert phase (phase 2). One-tail paired t-test showed the difference is significantly different with $p < 0.01$. This indicates that yawning detection is useful as a complementary cue in drowsiness detection. The column chart also indicates an interesting finding that hand-over-face gestures are generally also more frequent in the drowsy phase compared to the alert phase. To our knowledge, this is the first study that explored hand-over-face touches in drowsiness detection.

IV. FEATURE EXTRACTION

In order to build a yawning detection approach, we extracted a set of geometric and appearance features that can represent both hand-covered and uncovered spontaneous yawns, namely: mouth openness, Histograms of Oriented Gradients (HOGs) and Local Binary Patterns (LBP).

A. Degree of Mouth Openness

To avoid scaling of the mouth in different frames and considering the individual variances among participants, we use the aspect ratio of the mouth, by computing the ratio of the horizontal and vertical intensity projection of the mouth region, to represent the degree of mouth openness.

B. HOGs

The Histogram of Oriented Gradients (HOG) [18] is a popular feature for describing appearance in computer vision and image processing that has been successfully used in many detection tasks, including pedestrian detection [20] and facial landmark detection [21]. HOGs, which are the distribution of intensity gradients or edge directions, describe the local object appearance and shape within an image. A HOG descriptor counts the number of oriented gradient occurrences in a dense grid of uniformly spaced cells. These occurrences are represented as a histogram for each cell normalized in a larger block area. Since it operates on local cells, a HOG is invariant to geometric and photometric transformations, except for object orientation. Such changes would only appear in larger spatial regions.

C. LBPs

The Local Binary Patterns (LBPs) [19] descriptor is another popular visual descriptor used in computer vision which is powerful for texture matching and image classification. Recent research [22] has shown that in some detection tasks a HOG + LBP descriptor considerably improves performance over either HOG or LBP alone.

To calculate the LBP value for a pixel in the grayscale image, after dividing the examined window into cells, we compare the central pixel value with the neighboring pixel values. Then after computing the histogram over the cell and optionally normalizing the histogram, concatenate histograms of cells. This description captures very fine grained details in images, and is simple and fast to implement.

V. EXPERIMENTAL EVALUATION

This section describes the methodology, implementation, results and discussion of our automatic detection approach of spontaneous yawns. The main stages of our yawning detection approach include: face region detection, data pre-processing, feature extraction and classification. To evaluate our method of yawning detection, we used the labelled spontaneous dataset described in section III.

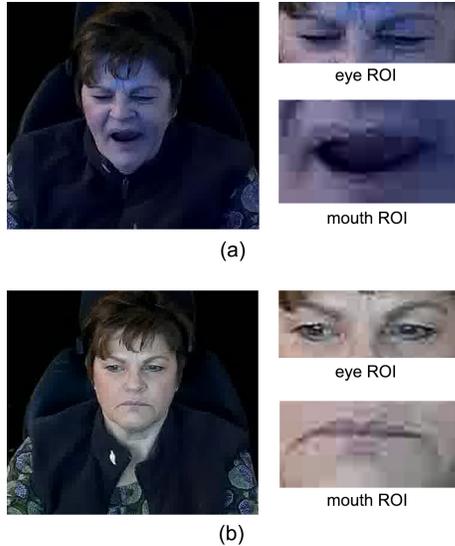


Figure 3. (a) Eye Regions of Interest (EROI) and Mouth Regions of Interest (MROI) in yawning state; (b) EROI and MROI in normal state.

A. Data pre-processing

First, we converted our videos into frames using FFmpeg tool [17]. We used OpenFace [23] to detect facial landmarks in the frames. Frames that have low face detection confidence score or when the covered areas of the face are so large that the face tracker failed were discarded. After filtering the frames, we got 15903 yawning images as positive samples. To balance our dataset, we randomly selected an equal set of negative samples including: silent faces (that is, participants are not yawning, speaking, or gesturing), speaking and frames having hand gestures without yawning. In total, we had 16330 negative samples.

B. Regions of interest

First, the face region is detected. The face tracker we use works well under partial occlusion. After landmark detection, we are able to identify the face regions of interest in all the valid frames. Two regions of interest were identified: Mouth Region of Interest (MROI) and Eye Region of Interest (EROI), described in Fig. 3. The MROI region is used to detect mouth features, which is of primary importance in yawning detection. The EROI is used to detect eye openness and possible wrinkles within eye area, which are also important features to detect yawning even when mouth is covered. Taking these two regions of interest into consideration has the following advantages: 1. We can capture distortion features around eyes and mouth even when the mouth is covered. 2. We can distinguish between yawning and speaking states by considering features that include mouth openness, mouth appearance, wrinkles between eyebrows, and eye appearance.

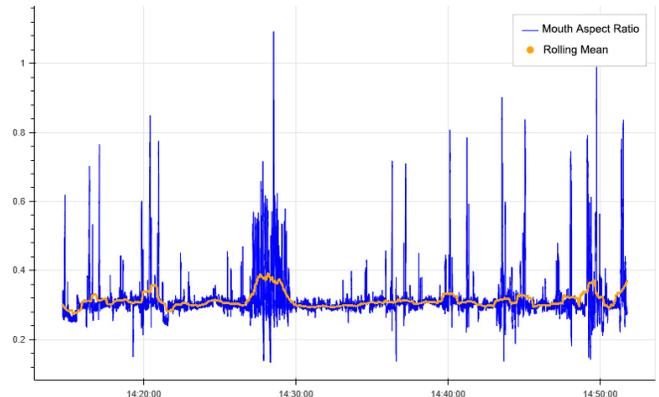


Figure 4. Sample plot of mouth openness during a period of time.

As a pre-processing step, we filtered out the frames in which the MROI and EROI areas fell out of the image boundary. We also left out frames where the detection confidence scores are low, which is usually because of extreme distortion of the head. We then resized the two regions detected in order to get feature vectors of the same dimensions for subsequent steps.

C. Implementation

One of the most popular features of yawning is the wide openness of people’s mouth. Therefore, we extracted mouth openness as our first feature. Using the OpenFace facial landmark data, we get the position of 68 points in the face, with 20 outlining the lips. After filtering the frames when the driver is absent or asleep, and where the confidence scores of the facial detection are low, we plotted how mouth openness changes over time for the sake of visualising the change (Fig. 4). Places where openness remains high over a period may indicate a yawn.

Besides mouth openness geometric features, we extract two appearance features: HOGs and LBPs, which are invariant to illumination, geometric and photometric transformations. We extracted HOGs and LBPs of the MROI and EROI regions from each 500×500 pixel image of the participant’s face. For HOGs, we used 8×8 pixel cells with 18 gradient orientations and a block size of 2×2 cells, leading to a 5040-dimensional HOG descriptor. We then used Principal Component Analysis (PCA) for dimensionality reduction, keeping 90% of explained variance, leading to a 1091 long dimension vector per frame. For the LBP feature, we consider a radius of 8 and a number of pixels in the neighborhood of 24. We then aggregated the appearance features in a temporal manner by taking the mean value in a window $w = 10$ frames.

For classification, we implemented a linear SVM classifier using scikit-learn library [24]. We employed a leave-one-participant-out cross validation approach in our experiments, by holding all videos of one participant out for testing

at each iteration. This ensures the generalisability of the proposed approach.

D. Results and Discussion

We trained a binary classifier using linear SVMs on our spontaneous dataset using the extracted geometric and appearance features described above. Using leave-one-participant-out cross validation, our approach managed to detect spontaneous yawns with an accuracy of 94.63% and a F1 score of 92.77% .

Our proposed fusion method provides an accurate classification of yawning based on both geometric and appearance features of two face regions. As the negative samples included speaking and silent drivers, normal mouth opening is efficiently differentiated from yawning mouth opening. Moreover, because the positive samples included a wide range of spontaneous yawns, both hand-covered yawns and uncovered yawns can be efficiently detected in our approach. The features selected proved invariant to different lighting conditions. As shown in Fig 3, lighting varied according to the simulated scenes in the driving simulator, which indicates that our method remains robust in changing lighting conditions and generalisable to real road scenarios. We did not include a baseline comparison here, because to the best of our knowledge, there is not available spontaneous dataset of natural yawns that is publicly available for comparison.

VI. CONCLUSION AND FUTURE WORK

In this paper we introduced a new method for real time driver yawning detection and proved that yawning is an important cue that should not be neglected in drowsiness detection systems. Moreover, we analysed the presence of face touches in drowsy versus alert videos of natural drivers' expressions and proposed the use of face touches as a novel indicator of the drowsiness state. We also presented a method for automatic detection of spontaneous yawns based on fusing geometric and appearance features extracted from both the mouth and eye regions. Our approach is invariant to different lighting conditions, skin color, and individual facial appearance differences. The proposed algorithm is also computationally simple and efficient.

Further work can be done to possibly improve the system. First, we may consider fusing more visual cues, such as Percent Eye Closure (PERCLOS) and face touches. We may also evaluate a non-linear kernel SVM or temporal machine learning algorithms to check if this will lead to any improvement in performance. In addition, our dataset contains videos for both manual and autonomous driving sessions, and the drivers tend to have more hand touches in autonomous sessions. More difference between the two sessions can be explored.

Ultimately, we can use our approach as a building block in a multimodal drowsiness detection system fusing other

features such as eye status (e.g. PERCLOS, eye closure duration, blink frequency), head orientation, and hand-over-face gestures frequency. This can help in designing a personalized and natural warning system for drivers in the future.

ACKNOWLEDGMENT

The work presented in this paper was funded and supported by Jaguar Land Rover, Coventry, UK.

REFERENCES

- [1] Rau, Paul Stephen, *Drowsy driver detection and warning system for commercial vehicle drivers: field operational test design, data analyses, and progress*. 19th International Conference on Enhanced Safety of Vehicles, 2005.
- [2] Jackson, P., et al. *Fatigue and road safety: a critical analysis of recent evidence*. London: Department for Transport 21, 2011.
- [3] Roman, Bittner, et al., *Fatigue indicators of drowsy drivers based on analysis of physiological signals*. Medical Data Analysis (2001): 62-68.
- [4] Wang, Tiesheng, and Pengfei Shi, *Yawning detection for determining driver drowsiness*. VLSI Design and Video Technology, 2005. Proceedings of 2005 IEEE International Workshop on. IEEE, 2005.
- [5] Ji, Qiang, Zhiwei Zhu, and Peilin Lan, *Real-time nonintrusive monitoring and prediction of driver fatigue*. IEEE transactions on vehicular technology 53.4 (2004): 1052-1068.
- [6] Wang, Tiesheng, and Pengfei Shi, *Yawning detection for determining driver drowsiness*. VLSI Design and Video Technology, 2005. Proceedings of 2005 IEEE International Workshop on. IEEE, 2005.
- [7] Abtahi, Shabnam, Behnoosh Hariri, and Shervin Shirmohammadi, *Driver drowsiness monitoring based on yawning detection*. Instrumentation and Measurement Technology Conference (I2MTC), 2011 IEEE. IEEE, 2011.
- [8] Rongben, Wang, et al. *Monitoring mouth movement for driver fatigue or distraction with one camera*. Intelligent Transportation Systems, 2004. Proceedings. The 7th International IEEE Conference on. IEEE, 2004.
- [9] Lu, Yufeng, and Zengcai Wang, *Detecting driver yawning in successive images*. Bioinformatics and Biomedical Engineering, 2007. ICBBE 2007. The 1st International Conference on. IEEE, 2007.
- [10] Saradadevi, Mandalapu, and Preeti Bajaj, *Driver fatigue detection using mouth and yawning analysis*. International journal of Computer science and network security 8.6 (2008): 183-188.
- [11] Fan, Xiao, Bao-Cai Yin, and Yan-Feng Sun, *Yawning detection for monitoring driver fatigue*. Machine Learning and Cybernetics, 2007 International Conference on. Vol. 2. IEEE, 2007.

- [12] Azim, Tayyaba, M. Arfan Jaffar, and Anwar Majid Mirza, *Automatic fatigue detection of drivers through pupil detection and yawning analysis*. Innovative Computing, Information and Control (ICICIC), 2009 Fourth International Conference on. IEEE, 2009.
- [13] Ibrahim, Masrullizam Mat, et al. *Yawn analysis with mouth occlusion detection*. Biomedical Signal Processing and Control 18 (2015): 360-369.
- [14] Fan, Xiao, Bao-Cai Yin, and Yan-Feng Sun, *Yawning detection for monitoring driver fatigue*. Machine Learning and Cybernetics, 2007 International Conference on. Vol. 2. IEEE, 2007.
- [15] Abtahi, Shabnam, et al., *YawDD: A yawning detection dataset*. Proceedings of the 5th ACM Multimedia Systems Conference. ACM, 2014.
- [16] Brugman, Hennie, Albert Russel, and Xd Nijmegen, *Annotating Multi-media/Multi-modal Resources with ELAN*. LREC. 2004.
- [17] FFmpeg Developers, *ffmpeg tool (Version be1d324) [Software]*. Available from <http://ffmpeg.org/> 2016.
- [18] Dalal, Navneet, and Bill Triggs, *Histograms of oriented gradients for human detection*. Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on. Vol. 1. IEEE, 2005.
- [19] Ojala, Timo, Matti Pietikainen, and Topi Maenpaa, *Multiresolution gray-scale and rotation invariant texture classification with local binary patterns*. IEEE Transactions on pattern analysis and machine intelligence 24.7 (2002): 971-987.
- [20] Dalal, Navneet, and Bill Triggs, *Histograms of oriented gradients for human detection*. Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on. Vol. 1. IEEE, 2005.
- [21] Zhu, Xiangxin, and Deva Ramanan, *Face detection, pose estimation, and landmark localization in the wild*. Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on. IEEE, 2012.
- [22] Wang, Xiaoyu, Tony X. Han, and Shuicheng Yan, *An HOG-LBP human detector with partial occlusion handling*. Computer Vision, 2009 IEEE 12th International Conference on. IEEE, 2009.
- [23] Tadas Baltrusaitis, Peter Robinson, and Louis-Philippe Morency, *OpenFace: an open source facial behavior analysis toolkit*. in IEEE Winter Conference on Applications of Computer Vision, 2016.
- [24] Pedregosa, Fabian, et al., *Scikit-learn: Machine Learning in Python*. Journal of Machine Learning Research 12.Oct (2011): 2825-2830.