YawDD: A Yawning Detection Dataset

Shabnam Abtahi, Mona Omidyeganeh, Shervin Shirmohammadi, and Behnoosh Hariri
Distributed and Collaborative Virtual Environment Research Laboratory (DISCOVER Lab)
University of Ottawa, Ottawa, Canada
{sabtahi | m_omid | shervin | bhariri}@discover.uottawa.ca

ABSTRACT

In this paper, we present two video datasets of drivers with various facial characteristics, to be used for designing and testing algorithms and models for yawning detection. For collecting these videos, male and female candidates were asked to sit in the driver's seat of a car. The videos are taken in real and varying illumination conditions. In the first dataset, the camera is installed under the front mirror of the car. Each participant has three or four videos and each video contains different mouth conditions such as normal, talking/singing, and yawning. In the second dataset, the camera is installed on the dash in front of the driver, and each participant has one video with the above-mentioned different mouth conditions all in the same video. The car was parked for both datasets to keep the environment safe for the participants. As a benchmark, we also present the results of our own yawning detection method, and show that we can achieve a much higher accuracy in the scenario with the camera installed on the dash in front of the driver.

Categories and Subject Descriptors

I.4.8 [**Image Processing And Computer Vision**]: Scene Analysis – Object Recognition

General Terms

Performance.

Keywords

Yawning dataset, driver yawing detection, operator fatigue

1. INTRODUCTION

Research shows fatigue and drowsiness of drivers are one of the most significant causes of car accidents. Based on a National Highway Traffic Safety Administration (NHTSA) report, 22 to 24 percent of car crashes happen when the driver is driving with drowsiness. This lack of alertness in driving causes a four- to six-times higher near-crash/crash risk compared to alert drivers [1]. Driver drowsiness negatively impacts the response time of the driver and, as a result, will increase the number of road accidents.

MMSys '14, March 19 - 21 2014, Singapore, Singapore

http://dx.doi.org/10.1145/2557642.2563678

Therefore, to increase the safety of transportation, there have been various research works conducted on how to detect a driver's drowsiness and fatigue while driving.

There are some warning signs associated with human fatigue, such as daydreaming while on the road, not driving between lanes, yawning, feeling impatient, feeling stiff, not looking at the road ahead, heavy eyes, and reacting slowly. Researchers have been investigated techniques to detect drowsiness of drivers based on the above signs. Particularly, many researchers have been focusing on yawning detection, since yawning is known to be associated with corresponding variation in sleepiness and sleep-related fatigue [17]. However, as we shall see in the next section, there is no freely-available driver yawning detection dataset which covers a variety of important conditions affecting the detection algorithm, and many of yawning detection studies use videos captured in a controlled lab environment, which does not represent reality. In this paper, we present two freely-available datasets which specifically target driver yawning. The first dataset contains 322 videos and the second dataset includes 29 videos, consisting of both male and female drivers, with and without glasses/sunglasses, from different ethnicities, and with 3 different mouth conditions: 1- normal driving with mouth closed (no talking), 2- talking or singing while driving, and 3- yawning while driving. It should be noted that while our dataset was developed particularly for driver yawning detection, it can also be used for yawning and fatigue detection in other applications where operator alertness is important, such as airplane pilots, operators of heavy machinery, airport radar operators, etc.

The rest of this paper is organized as follows. The next section gives a background about face datasets, yawning detection techniques, and the necessary requirements of a yawning detection dataset. Section 3 explains the data collection methodology including the information about camera, environment, participants and videos. Section 4 describes the basic statistics of our dataset, including the number of male and female participants and different facial characteristics. Section 5 describes an implementation and the results and limitations of a previously successful algorithm tested with the collected dataset. Section 6 explores different possibility of usage for our dataset. Finally, the last section 7 gives the information about dataset availability to be used by other researchers and its copyrights.

2. YAWNING DETECTION

Most existing techniques for yawning detection start by finding the location of the face, the latter an active research area for many years [2][3][4]. There are many different datasets providing static digital images and video clips of faces of people which can be used for face detection, such as the dataset in [5]. The situations that these pictures and videos are taken include but not limited to, close range, under controlled lightning condition, under variable illumination, and at moderate and varying distances. One of the popularly used face recognition datasets is the dataset used for Facial Recognition Technology (FERET), which contains a set of images taken in a semi controlled environment with different camera and different lighting [6]. The participants in the dataset were asked to display a different facial expression for the image. The face dataset which was used in the face recognition process of [1] is provided by Caltech face dataset [7] which has 450 images of frontal faces with different background and lighting conditions. While these datasets are popular, they were not intended for the specific situation of detecting driver yawning, and do not represent the conditions that exists during driving, as we will see in section 3.

Specifically for yawning detection, [8] uses the difference image between two images and finding the yawning condition through calculating the vertical distance between the midpoint of nostrils and chin. But the dataset used to support this method is based on several videos taken in laboratory under controlled lighting conditions by a webcam. Using a webcam under laboratory lighting conditions does not match with how a driver is really filmed in a car, as we shall see in section 3. [9] proposes to detect driver yawning using a CCD camera to track a driver's mouth. The dataset contains 400 images (200 yawning images and 200 normal images) but was limited to only twenty videos. [10] tests their multiple visual cues based algorithm to decide whether the driver is fatigue or not by using real-time videos. In [1], we propose a method of yawning detection based on changes in the geometric features of the mouth. After testing the face detection algorithm with the Caltech dataset, 250 images of people in yawning and normal positions were used in order to increase the reliability of the driver yawning detection; however, stand-alone pictures are not as reliable as videos, simply because a single picture of an open mouth could be due to talking or singing, not necessarily yawning. Finally, the yawning detection method in [12] is based on face detection and mouth extraction within the face region. These methods were tested with over ten videos, which is a small size, taken in a car and under natural conditions, but still using a webcam.

As shown above, a number of attempts have been made to create datasets for yawning detection studies. However, each dataset has its own limitations. In summary, there is no freely-available dataset that has all of the following important characteristics:

1- The camera must be of a type that is typically used in actual driver monitoring systems. A webcam is usually not of this type.
2- The camera's position and its angle of capturing the driver must be the same as how it is actually setup in real driver monitoring systems. Most datasets and their corresponding algorithms work based on fully-frontal face views, which is impractical in a real car because the camera cannot be installed directly in front of the face of the driver (it would block the driver's view).
3- The driver must be filmed in various situations such as normal (no talking), talking or singing, and yawning. This is important to test the algorithm against false-positives.
4- Facial obstructions such as prescription glasses, sunglasses, mustache, and bread must be available to test the robustness of the algorithm.
5- Different ethnicities must be available to make sure the algorithm works with various skin colors, so errors caused by different skin colours such as the one reported in [15] are avoided.
6- Various lighting and illumination conditions (cloudy, sunny, etc.) should be available in the dataset.
7- The size of the dataset must be large enough to lead to statistically meaningful results. Considering gender, race, the facial situations explained in requirement 4-, and false positive situations explained in requirement 3-, a total of a dozen or so videos currently used in existing datasets is not enough.
8- The dataset must provide videos, not images, because a still image that appears to be yawning could actually be talking or singing. A video will give more information to better detect a yawning pattern.

Due to the lack of a yawning dataset that fulfills the above requirements, we have created a dataset by filming drivers' yawning. As explained earlier, In addition to yawning detection, these datasets can also be used for face recognition, eye detection, mouth detection and tracking in different illumination and lighting. However, its main purpose is for driver's yawning detection.

3. DATA COLLECTION METHODOLOGY

As noted in [14], face is a very non-rigid and complex structure; therefore, the appearance of a face is affected by a large number of different factors including "identity, face pose, illumination, facial expression, age, occlusion, and facial hair". As a result, to create reliable algorithms sensitive to these facial variations, a large enough dataset is required that includes "carefully controlled variations of these factors". To collect our dataset, we carefully configured the following components:

1- Camera
2- Environment
3- Participants
4- Videos

Each component will be described in details in the following sections.

3.1 Camera

To ensure a realistic setup, we worked with our industry partner CogniVue Corporation, whose APEX™ line of products specialize in smart in-car camera systems. Two different locations for installing the camera were considered in the video dataset collection. In the first scenario, the camera was installed under the front mirror, and in the second scenario the camera was installed on the driver's dash. The videos were collected using a Canon A720Is digital camera with the resolution set at 640x480 pixels and 24-bit true color (RGB) at 30 frames per second, resulting in a video that matches with the video produced by real driver monitoring systems.

3.2 Environment

In our datasets, the collected videos were taken in the car during day time. However, we have recorded videos in various lighting situations in order to provide a more complete dataset. The videos

Figure 1. Some Participants and their angle of capture; top: camera under the mirror; bottom: camera on the dash.

were recorded from early morning till sunset. Also, the weather varies from sunny to rainy, causing different lighting conditions. In some of the videos, there are also some background movements due to other passengers moving across the frame.

It should also be noted that night-time videos were not collected, because for yawning detection at night, either the lighting must be controlled in the car, in which case normal day time detection approaches would work too, or infrared cameras must be used to overcome the extreme darkness of the image, which leads to completely different approaches of image processing and computer vision. However, in the future, our dataset can be extended by including night-time videos.

3.3 Participants

The participants were asked to sit in the driver's seat and fasten their seat belt to make the scenario more realistic. The dataset contains videos of 57 male and 50 female volunteers from different ages, ethnicities, and facial characteristics. The statistics about the participants is given in section 4. A high variety of appearances existed among these 107 volunteers. People participated with and without glasses, men with and without beard, men with and without moustache, women with and without scarf, different hairstyles and different clothing. Figure 1 shows a random selection of 12 of the volunteers. All of the participants signed an agreement allowing their videos to be used for non-commercial and research purposes, although only a subset of the participants agreed for their pictures to be published in research papers. Researchers who use our dataset and publish papers from it should pay special attention to the given tables in the dataset (see section 7) and NOT use in their papers the pictures of participants who have not agreed for their pictures to appear in papers.

3.4 Videos

In the first dataset (camera under the front mirror), for the majority of participants, three videos were taken. In the first video, the person was asked to sit in the driver's seat, fasten seat belt, and act driving. In the second video of the same person, s/he was asked to talk or sing while driving. This second video can be used to distinguish between talking/singing compared to yawning, where both scenarios might lead to an open mouth but only yawning should be detected. In the third video of the same person, s/he was asked to wait for a few seconds and yawn afterwards. The videos last between 15-40 seconds. A few participants were asked to have a fourth video of their yawning while they are talking. These videos are more challenging for researchers since the yawning happens right in the middle of talking, and so it might be more difficult to detect the former. Finally, some participants acted with and without their prescription glasses or sunglasses. In the second dataset (camera on the driver's dash), the participants were asked to sit on the front seat and fasten their seat belt. Then, a single video was taken where participants were asked to pretend that they were driving in normal, talking and yawning positions. The combination of these three phases is useful to create a more realistic scenario and to distinguish between these cases at the same time and detect the yawning position accurately. After collecting 342 videos, the audio was removed to reduce size.

4. BASIC STATISTICS

As mentioned earlier, volunteers with different genders, ages and ethnicity participated in our video dataset collection. Having a wide variety in participants gives a better and more reliable dataset in order to increase the accuracy of yawning detection systems. The age distribution of the participants is shown in figures 2 and 3.

We also included different facial characteristics. Statistically, we have videos for:

1- **Female participants:**

With prescription eyeglasses : 14 people

With sunglasses : 11 people

Without glasses: 40 people

With scarf: 3 people

Without scarf: 53 people

2- **Male participant:**

With prescription eye glasses : 35 people

With sunglasses : 5 people

Without glasses : 33 people

With moustache : 4 people

Without moustache: 59 people

With Beard: 7 people

Without beard: 56 people

In terms of ethnicity, we have people with various skin, hair, and eye colour, including blondes, brunettes, Caucasian, African, Middle-eastern, and Asian.

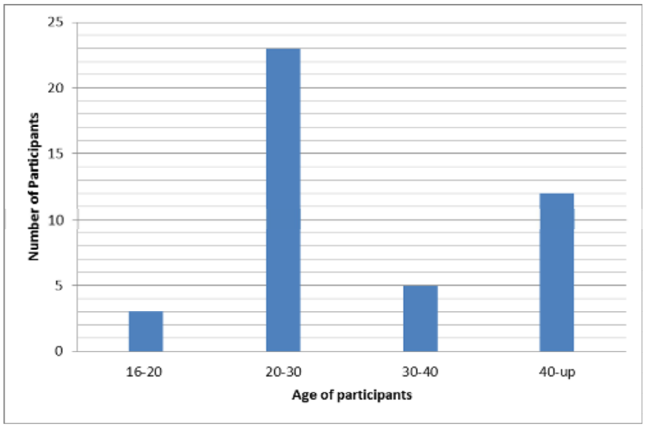

Figure 2. Female Participants

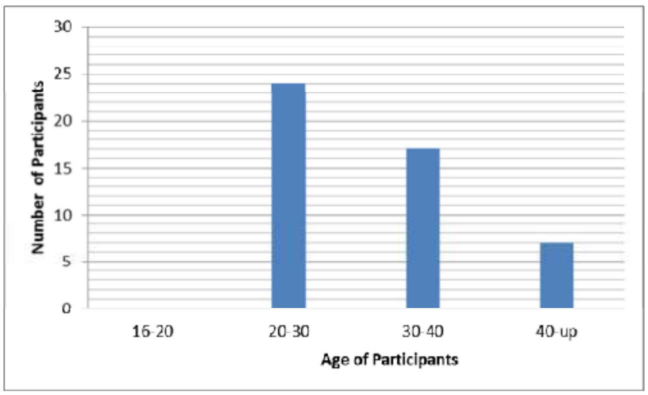

Figure 3. Male Participants

5. BENCHMARK IMPLEMENTATION

In our previous study, we developed and implemented a yawning detection algorithm tested by a dataset consisting of 280 frontal face images [1]. With that dataset, we were able to successfully detect the face and the mouth in a standby and in a yawning position for each image with 100% accuracy. However, that method was too computationally complex to be implemented in an actual embedded smart camera platform. So in [16] we proposed a computationally lightweight method based on the Viola-Jones theory for face and mouth detection, and back projection measurement technique for yawning detection. The details of the method can be read in [16]. Due to its lower complexity, this method was implementable on actual in-car smart camera systems such as CogniVue's APEX^{TM}, with results for both datasets shown in Table I. As can be seen from the table, the results for the first dataset are not very good, since the Viola Jones face detector is trained to distinguish faces that are tilted up to about ±45 degrees out of plane (towards a profile view) and to about ±15 degrees in plane. This does not match our camera under the mirror scenario where the angles are different, but it does match our second scenario where the camera is installed on the drover's dash, where the results are much better as can be seen from Table I.

We are now working on designing higher-accuracy detection techniques that are computationally not too expensive to run on in-car smart camera systems. It is quite easy to come up with algorithms that have 100% detection accuracy using a workstation or a personal computer, as we have also done in [1]. The challenge is to design detection methods that can run on computationally-limited embedded camera systems. We hope that by making our dataset freely available, other researchers will also be able to use it as a dataset for driving, and design algorithms and methods to detect yawning that are more accurate and still computationally executable on in-car embedded smart cameras.

Table I. Detection results

	Face detection	Mouth detection	Yawning detection
Camera under the mirror	85%	50%	22%
Camera on the dash	93%	78%	60%

6. OTHER APPLICATIONS

While primarily developed for driver yawning detection, our dataset can also be used by researches who are developing algorithms related to yawning detection for pilots, radar operators, and heavy machinery operators. These are all situations where the alertness of the operator is crucial, and its lack could be fatal. Our dataset can also be used for face recognition and tracking, as well as mouth/eye detection and tracking.

7. AVALABILITY and FORMAT

Both datasets are available for free as long as they are used for non-commercial and research purposes. Instructions on how to obtain them are posted at our website[1]. The detailed description of each individual video is also posted in that address. As mentioned before, in the first dataset a camera is installed under the front mirror of the car and each participant has three/four videos and each video contains different mouth conditions such as normal, talking/singing, and yawning. It provides 322 videos consisting of both male and female drivers, with and without glasses/sunglasses, from different ethnicities, and in 3 different situations: 1- normal driving (no talking), 2- talking or singing while driving, and 3- yawning while driving. In the second dataset, the camera is installed on the dash of the car in front of the driver and one video of each participant is taken while the participant is driving normally, talking, and yawning.

The videos are in 640x480 24-bit true color (RGB) 30 frames per second AVI format without audio. All participants have signed agreements that their video can be used for non-commercial research purposes. In addition, a small number of participants have additionally granted permission for their faces to appear in research publications. These are clearly indicated in the tables on

[1] http://www.eecs.uottawa.ca/~shervin/yawning

the aforementioned website. Researchers who use our dataset and publish papers from it should pay special attention to the descriptions in the tables and NOT use in their papers the pictures of participants who have not agreed for their pictures to appear in papers. At the time of writing this paper, the size of both datasets together is about 5 Gigabytes. This might change in the future as other researchers contribute more videos to the dataset. To contribute videos to this dataset, please follow the instructions on the aforementioned website.

8. CONCLUSIONS

To design and develop successful driver yawning detection methods and systems, two important issues must be considered: 1- the dataset used to design and test the algorithm must contain realistic driving videos using a variety of faces and under various conditions; and 2- the designed system must be implementable in an actual in-car smart camera system. Unfortunately, the great majority of yawning detection research use images or videos taken under controlled laboratory conditions, and/or develop algorithms that are too computationally complex to work in an actual car embedded system. By publishing our dataset in this paper, we contribute to the first issue above, and we also provide benchmark results of our algorithm in [16] that satisfies the second issue above. We hope researchers will be able to use both to improve driver yawning detection systems.

REFERENCES

[1] S. Abtahi, B. Hariri, S. Shirmohammadi, "Driver Drowsiness Monitoring Based on Yawning Detection", *Proc. IEEE International Instrumentation and Measurement Technology Conference*, Binjiang, Hangzhou, China, May 10-12, 2011.

[2] J.F. Cohn, A.J. Zlochower, J. Lien, and T. Kanade, "Automated Face Analysis by Feature Point Tracking has High Concurrent Validity with Manual FACS Coding," *Psychophysiology*, vol. 36, 1999, pp. 35-43.

[3] S. Dubuisson, F. Davoine, and M.A. Masson, "Solution for Facial Expression Representation and Recognition," *Signal Processing-Image Comm.*, vol. 17, pp. 657-673, 2002.

[4] N. A. A.Rahman, K.C. Wei and J. See. "RGB-H-CbCr Skin Colour Model for Human Face Detection." *Proc. MMU International Symposium on Information & Communications Technologies*, 2006

[5] A.J.O'Toole, J. Harms, S.L. Snow, D.R. Hurst, M.R.Pappas, J.H.Ayyad,H.Abdi, " A video Database of Moving Faces and People," *IEEE Transaction on Pattern Analysis and Machine Intelligence*, vol.27, pp.812-816, May 2005

[6] In P. J. Phillips, H. Moon, S. Rizvi, and P. J. Rauss. "The FERET evaluation methodology for face-recognition algorithms", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(10):10901104, 2000. http://www.nist.gov/humanid/feret/.

[7] http://www.vision.caltech.edu/html-files/archive.html

[8] U. Yufeng, W. Zengcai, "Detecting driver yawning in successive images." *Proc. International Conf. on Bioinformatics and Biomedical Engineering*, 2007 , pp. 581-583.

[9] X. Fan, B. Yin, Y. Fun. "Yawning Detection For Monitoring Driver Fatigue." *Proc. Sixth International Conf. on Machine Learning and Cybernetics*, Hong Kong, 2007, pp. 664-668.

[10] T. Azim, M.A. Jaffar, A.M. Mirza. "Automatic Fatigue Detection of Drivers through Pupil Detection and Yawning Analysis." *Proc. International Conf. on Innovative Computing, Information and Control*, 2009, pp. 441-445.

[11] L. Li, Y. Chen , Z. Li. "Yawning Detection for Monitoring Driver Fatigue Based on Two Cameras." *Proc. International IEEE Conf. on Intelligent Transportation Systems*, St. Louis, MO, USA, 2009, pp. 12-17.

[12] T.Wang, P. Shi. "Yawning Detection For Determining Driver Drowsiness." *Proc. IEEE International Workshop VLSI Design & Video Tech.*, Suzhou,China, 2005, pp. 373-376.

[13] B. Martinkauppi,M. Soriano, S. Huovinen, M. Laaksonen,"Face Video Database", *Proc. Conference on Colour in Graphics, Image, and Vision*, 2002, Amsterdam, Netherlands, May 2002, pp. 380-383

[14] R. Gross, "Face Databases", Handbook of Face Recognition. Springer-Verlag, 2005

[15] T. Staedter , "Face Detection Tech Isn't Racist; It's Just Stupid", Discovery News, Tue Dec 29, 2009 http://news.discovery.com/tech/face-detection-tech-isnt-racist-its-just-stupid.html

[16] S. Abtahi, S. Shirmohammadi, B. Hariri, D. Laroche, and L. Martel, "A Yawning Measurement Method Using Embedded Smart Cameras", *Proc. IEEE International Instrumentation and Measurement Technology Conference*, Minneapolis, USA, May 6-9 2013, pp. 1605-1608.

[17] A.C. Gallup and O.T. Eldakar, "The Thermoregulatory Theory of Yawning: What We Know from over 5 Years of Research", *Frontiers in Neuroscience*, January 2 2013, doi: 10.3389/fnins.2012.00188.