SVM Classification for Drowsiness Detection Using Eye Aspect Ratio

Caio B. Souto Maior, Márcio C. Moura, João M. M. de Santana, Lucas M. do Nascimento, July B. Macedo, Isis D. Lins.
Center for Risk Analysis and Environmental Modeling, Department of Production Engineering, Federal University of Pernambuco, Recife, Brazil

Enrique L. Droguett
Mechanical Engineering Department, University of Chile, Santiago, Chile

1 Introduction

In recent times, human reliability has been extensively studied due to the risks imposed by human errors in industrial scenarios and the great potential of loss involved. Specifically, there is a global growing concern with the consequences of performance breakdown by operators in safety-critical task environments, since a distraction or inattention could damage the entire system, including the integrity of the people in it.

Despite the increasing automation, the human operator has a critical central role in the execution of these complex tasks. Additionally, high demands in life outside of work – as sleep disturbances and stress – present a cumulative challenge that allows humans to both respond effectively to new challenges and to reduce engagement with demands whenever stability is threatened by impending fatigue [1]. Operator monitoring can be employed as an attempt to ensure the correct performing of tasks and can be done with the aid computer vision. This computer vision technique aims to extract visual fatigue related characteristics in real-time using image/video processing.

Robust object detection is one of the most developed research domains in the field of computer vision and pattern recognition. It has been used in different useful applications in all the possible domains, like security and surveillance, industrial control, robotics, smart objects, semantic search, video-stream analysis, among others [2]. The same concept is also used to track human body (e.g. eye, mouth, arms), recognizing its shape and movement, creating diagnosis and also inferring future human behavior.

Specifically, metrics such as gaze, head movement, and eye blink rate provide interesting characteristics of a human state. Moreover, detecting eye blinks has important roles in systems that monitor a human operator's vigilance [3,4], in systems with warnings to avoid dry eye and computer vision syndromes [5, 6], in human computer interfaces that ease communication for disabled people [7], or for anti-spoofing protection in face recognition systems [8].

Blink analysis has been extensively applied to detect driver drowsiness, but almost no work extends the analysis to safety-critical tasks of an operator in process plants – such as nuclear and oil and gas industries.

2 General Objectives

This paper aims to create an efficient model to classify the operator drowsiness in industrial plants in real-time based on eye detection and blink analysis. From the eye performance, it is possible to determine if one is asleep or awake. Even more, it is expected to provide a contribution to the field of Operational Functional State (OFS), defined as "the variable capacity of the operator for effective task performance in response to task and environmental demands, and under the constraints imposed by cognitive and physiological processes that control and energize behavior" [1].

3 Related Works

As explained by [2], there are two major ways to detect blinking and, specifically, detect drowsiness:

1. By biological approaches such as electrooculogram (EOG) (as in ref. [9]), electroencephalogram (EEG) or electrocardiogram (ECG) readings;
2. By image processing and computer vision methods.

In this section, a few scientific papers will be presented, mainly related with the second category.

Ref. [3] depicts an interesting review of the role of computer vision technology applied to the development of monitoring systems to detect distraction. Authors explain the main methods for face detection, face tracking and detection of facial landmarks, as well as the main algorithms for biomechanical, visual and cognitive distraction. Additionally, there are some algorithms detecting mixed types of distraction and the relationship between facial expressions, key points for the development and implementation of sensors and test and training to driving monitoring systems are summarized.

A real-time algorithm to detect eye blinks in a video sequence from a standard camera is proposed in ref. [10]. They show that recent landmark detectors, trained on in-the-wild datasets, exhibit excellent robustness against a head orientation with respect to a camera, varying illumination and facial expressions, being sufficiently precise to reliably estimate the level of the eye opening. Also, they use a SVM classifier to detect eye blinks as a pattern of EAR values in a short temporal window.

In Ref. [11], a blinking detection method is proposed, based on Gabor filters by measuring the distance between the two arcs of the eye. First, the eye is detected by Viola-Jones' method [12]. Next, the Gabor filter extracts the pattern of the eye based on orientation angle. They apply a connected labeling method to detect the two arcs and measure the distance between them compared to a threshold.

A system that measures eye blinking rate and eye closure duration is proposed by [13]. The system consists of skin-color segmentation, facial features segmentation, iris positioning and blink detection. The proposed skin-segmentation procedure is based on a neural network approximation of a RGB skin-color histogram. The segmented eye regions are analyzed with the Circular Hough transform with the purpose of finding iris candidates.

4 Methodology

This paper methodology is based on three main parts: (1) real-time eye detection, (2) SVM Classification and (3) Drowsiness detection and warnings.

4.1 Real-time Eye Detection

To construct our real-time drowsiness classification model, it is necessary to identify the operator's eye on the image. Based on it, we build a blink detector to recognize every blink and distinguish it from a nap or drowse.

Unlike traditional image processing methods for computing blinks, which typically involve some combination of: (1) eye localization, (2) thresholding to find the whites of the eyes, (3) determining if the "white" region of the eyes disappears for a period (indicating a blink), we use a metric called eye aspect ratio (EAR), introduced by [10] . Furthermore, we use a weighted average eye aspect ratio from a neutral face and a smiling face. This procedure is done to capture one's natural EAR and subtle changes when facial expression varies; this will be discussed in section 3.2.

The eye detection is performed with facial landmarks. Our goal is to detect important facial structures on the face using shape prediction methods. We use the dlib library [14] for facial landmark detection based on [15], which uses Histograms of Oriented Gradients and Linear Support Vector Machines in the procedure. The library provides landmarks for the entire face, displayed as light green dots in Fig. 1. The landmarks are adaptive to recognize the shape of distinct human faces.

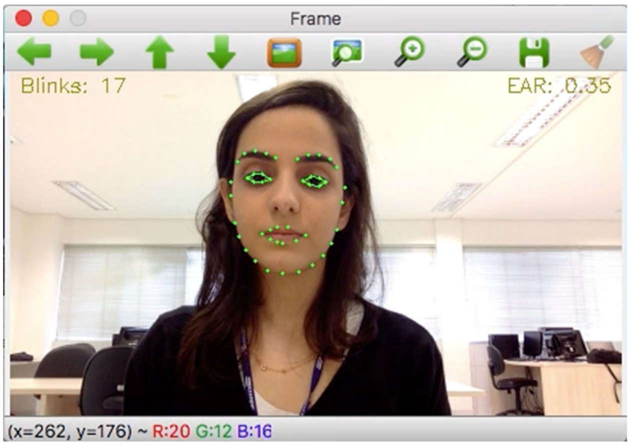

Figure 1: Landmarks provided for the entire face.

4.2 *SVM Classification*

In this part, only the landmarks related with the eyes are used. First, the proportion between width and height of the eye, based on its landmarks, is calculated. This is called EAR. Fig. 2 depicts the landmarks using in the procedure.

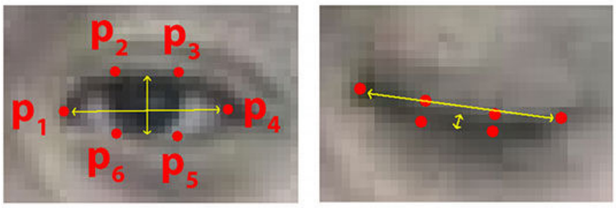

Figure 2: Eyes landmarks. Source: Ref. [10]

The EAR is calculated based on the following equation, proposed by ref. [10]:

$$EAR = \frac{\|p_2 - p_6\| + \|p_3 - p_5\|}{2\|p_1 - p_4\|} \qquad (1)$$

One's EAR is calculated for each frame of the real-time video. Thus, a decrease in the EAR is expected when the user closes his eyes going back to a normal level when eyes are open again. This approach can be used to detect blinks as well as eye openness.

In order to detect drowsiness, two methods of classification were applied, based on the approach proposed by ref. [10]. The first approach is here called threshold model. A threshold was considered so that if the user's EAR becomes lesser than this threshold for a number of frames (e.g. three consecutives frames), a blink is then detected. However, a single threshold is not adequate for every user, since each person has his own 'natural EAR' due to his face features. Thus, it was not possible to set a universal threshold, being necessary the calibration for every user. Also, even after calibration, this approach leads to many false positive alerts, once natural expressions such as talking or smiling tended to decrease the EAR.

A second approach was considered, where the EARs from 15 consecutive frames were concatenated to create a user's state feature. This feature could be classified in three categories: open eye, blink or closed eye (0,1 and 2, respectively). Specifically, the state feature is considered a blink only when the touch of eyelids occurs in the seven central frames of the state feature (i.e. from the 5^{th} to the 11^{th} frame). This procedure was adopted to avoid counting the same blink twice or more when running the detection model. Therefore, the model uses the fifteen-dimension input (15 consecutives EARs) and returns a scalar (state feature).

Support Vector Machine (SVM) was used as the classification learning algorithm, once it presents the advantage that previous knowledge about neither the function behavior nor the relation between input and output are required. The problem could be seen as follows: there exists a mapping function $y = f(x)$, unknown, of real values and, possibly, non-linear between an input vector $\boldsymbol{x}$ and an output scalar y and the only available information is the data set D, called training data. This corresponds to solving a convex and quadratic optimization problem with the Karush-Kuhn-Tucker (KKT) conditions as necessary and sufficient to guarantee a global optimum. The goal is not to look for the perfect alignment between the function $f(x)$ and D, but the best representation for the mapping. This mapping will separate each region in the classification problem, i.e., delimit whether an input *x* (state feature) is open eye, blink or closed eye.

As aforementioned, SVM needs a set of training data to infer about state feature. Thus, a data base was created containing 170 feature state samples. From the total of 170 samples, 73 contained open eyes EARs, 57 contained blinking EARs, and 40 contained closed eyes EARs. This data base contained data from each state's feature (i.e. 0, 1, 2) and includes different gender users, natural and abnormal facial expressions, glass wearing users and distinct head positions. Data were collected from several users, once EAR usually differs for each person. This was done to teach the algorithm using ordinary situations, when all these cases could be possible.

Therefore, SVM was trained and the database mentioned before was used as training set. Also, SVM requires a set of parameters in its formulation, and providing it with adequate values in the training phase is a considerable challenge since in principle, those are defined a priori. An optimization function was used to determine a suitable choice of parameters that best represent the mapping function. The presented methodology and SVM model were developed in Python language, using Scikit-learn package [16].

Then, a new data set, called test set, was created to measure the accuracy of the learning model. This is used to verify whether the model has strong performance when predicting unseen data. The test data set contained a total of 89 samples, with 17 open eyes, 30 blinks, and 42 closed eyes. The model demonstrated to be reliable when predicting the different states from testing data. No false positives were observed, i.e., detecting blinks or closed eyes when eyes were in fact open; two blinks (6.67% of blink test data) were missed, i.e., classified as open eyes; all occurrences of closed eyes were predicted successfully.

4.3 Drowsiness Detection and Warnings

Once predictions provided by SVM were considered satisfactory, real-time monitoring could be achieved. After every 7 new frames received from the video feed (approximately 0.28 seconds), a new state feature was created using the EARs from the latest set of 15 frames available, and SVM classifies this feature in one of the three classes (open eyes, blink, closed eyes).

The interesting here is dealing with drowsiness detection. Thus, SVM outputs (i.e. sequence of 0, 1 and 2) are used to determine whether the user is drowsiness or not. Two different situations were considered as drowsiness states:

1. If within a period of 30 seconds, 5 or more outputs are 2 (closed eyes). This situation occurs when the user's eyes close for brief moments, i.e., a slow blink;
2. If 4 or more consecutives outputs (i.e. predictions of SVM) are 2 (closed eyes). This occurs when the user's eyes remain closed for more than a second.

In practice, the first situation deals with many long blinks in a short lapse of time. Consecutives long blinks are highly correlated with first stages of drowsiness. The second situation deals with continuously closed eyes, i.e., the user keeps the eyes closed for more than one second. This situation could be seen as a later drowsiness stage, where the attention to the task is virtually null.

When either of the two situations above mentioned are detected, a notification pops up in the screen, warning the possible state of drowsiness, as depicted in Fig. 4. This pop up aims to alert the user about his own state and avoid inattention.

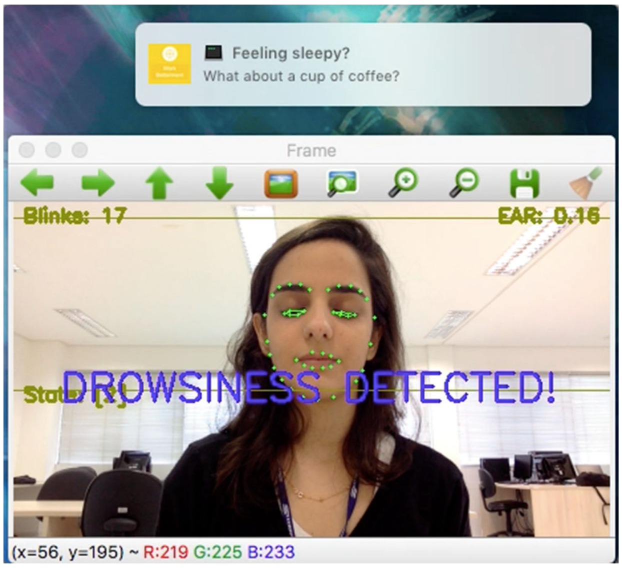

Figure 3: Drowsiness alert screen and notification.

5 CONCLUSION

A machine learning-based algorithm to automatically classify the operator state feature was presented. The method used face landmarks to estimate the user's EAR, subsequently applying an optimized-SVM to classify the state feature. Then, a decision rule was adopted to determine whether the user/operator was drowsy or not.

The application is physically non-intrusive and could be run alongside other programs in a consumer-class computer without noticeable practical performance impact.

6 FUTURE WORK

Currently, the authors are working with an independent drowsiness basis [17] and applying the proposed methodology to in a real time experiment.As a future work, we expect to include monitoring of other facial landmarks (e.g. mouth landmarks) to improve the detection method and avoid false positive drowsiness state when the user yawns, for example.

Moreover, machine learning with state features containing EARs from longer sets of frames and different input data (other than the EAR itself), extending the drowsiness detection. Our goal is to couple physics and biological signals (e.g. EEG & ECG) alongside with computer vision to enhance the performance of the detection.

REFERENCES

[1] Hockey G.R.J., Gaillard A.W.., Burov O. *Operator Functional State: The assessment and Prediction of Human Performance Degradation in Complex Tasks*. (2003)

[2] Lefkovits S., Lefkovits L., Emerich S. "Detecting the Eye and its Openness with Gabor Filters". *2017 5th International Symposium on Digital Forensic and Security (ISDFS)*. (2017)

[3] Fernandez A., Usamentiaga R., Carus J.L., Casado R. "Driver distraction using visual-based sensors and algorithms". *Sensors (Switzerland)*. (2016)

[4] Salehian S., Far B. "Embedded Real Time Blink Detection System for Driver Fatigue Monitoring". *The 27th International Conference on Software Engineering and Knowledge Engineering*. (2015)

[5] Yan Z., Hu L., Chen H., Lu F. "Computer Vision Syndrome: A widely spreading but largely unknown epidemic among computer users". *Computers in Human Behavior*, vol. 24, n. 5, 2026–2042. (2008)

[6] Divjak M., Bischof H. "Eye Blink Based Fatigue Detection for Prevention of Computer Vision Syndrome.". *Conference on Machine Vision Applications*, 350–353. (2009)

[7] Torricelli D., Goffredo M., Conforto S., Schmid M. "An adaptive blink detector to initialize and update a view-basedremote eye gaze tracking system in a natural scenario". *Pattern Recognition Letters*, vol. 30, n. 12, 1144–1150. '2009)

[8] Pan G., Sun L., Wu Z., Lao S. "Eyeblink-based anti-spoofing in face recognition from a generic webcamera". In Proceedings of the IEEE International Conference on Computer Vision. (2007)

[9] Zhang C., Wu X., Zhang L., He X., Lv Z. "Simultaneous detection of blink and heart rate using multi-channel ICA from smart phone videos". *Biomedical Signal Processing and Control*, vol. 33, 189–200. (2017)

[10] Soukupová T., Cech J. "Real-Time Eye Blink Detection using Facial Landmarks". *21st Computer Vision Winter Workshop*. (2016)

[11] Arai K., Mardiyanto R. "Real Time Blinking Detection Based on Gabor Filter". *International Journal of Recent Trends in Human Computer Interaction (IJHCI)*, vol. 1, n. 3, 33–45. (n.d.-k)

[12] Viola P., Jones M. "Robust real-time object detection". *International Journal of Computer Vision*, vol. 57, n. 2, 137–154. (2001)

[13] Lenskiy A. a., Lee J.-S. "Driver's eye blinking detection using novel color and texture segmentation algorithms". *International Journal of Control, Automation and Systems*, vol. 10, 317–327. (2012)

[14] King D. "Dlib". (2003)

[15] Kazemi V., Sullivan J. "One millisecond face alignment with an ensemble of regression trees". In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (pp. 1867–1874). (2014)

[16] Pedregosa F., Varoquaux G. *Scikit-learn: Machine learning in Python*. Journal of Machine Learning Research (Vol. 12). (2011)

[17] Massoz Q., Langohr T., Francois C., Verly J.G. "The ULg multimodality drowsiness database (called DROZY) and examples of use". *2016 IEEE Winter Conference on Applications of Computer Vision,*

WACV 2016. (2016)