mEBAL: A Multimodal Database for Eye Blink Detection and Attention Level Estimation

Roberto Daza, Aythami Morales, Julian Fierrez, Ruben Tolosana
Biometrics and Data Pattern Analytics, BiDA-Lab, Universidad Autonoma de Madrid, Spain
{roberto.daza,aythami.morales,julian.fierrez,ruben.tolosana}@uam.es

ABSTRACT
This work presents mEBAL\(^1\), a multimodal database for eye blink detection and attention level estimation. The eye blink frequency is related to the cognitive activity and automatic detectors of eye blinks have been proposed for many tasks including attention level estimation, analysis of neuro-degenerative diseases, deception recognition, drive fatigue detection, or face anti-spoofing. However, most existing databases and algorithms in this area are limited to experiments involving only a few hundred samples and individual sensors like face cameras. The proposed mEBAL improves previous databases in terms of acquisition sensors and samples. In particular, three different sensors are simultaneously considered: Near Infrared (NIR) and RGB cameras to capture the face gestures and an Electroencephalography (EEG) band to capture the cognitive activity of the user and blinking events. Regarding the size of mEBAL, it comprises 6,000 samples and the corresponding attention level from 38 different students while conducting a number of e-learning tasks of varying difficulty. In addition to presenting mEBAL, we also include preliminary experiments on: i) eye blink detection using Convolutional Neural Networks (CNN) with the facial images, and ii) attention level estimation of the students based on their eye blink frequency.

\(^1\)https://github.com/BiDAlab/mEBAL

CCS CONCEPTS
- Applied computing → E-learning
- Computing methodologies → Activity recognition and understanding

KEYWORDS
Eye Blink, Attention level, E-learning, Cognitive modelling

1 INTRODUCTION
The importance of virtual education platforms has increased significantly in the last 10 years \cite{6} and the COVID-19 outbreak in 2020 has strengthened that importance. With a large percentage of the academic institutions around the world now in lockdown, virtual education has temporally replaced traditional education to a very large extent.

E-learning is not only useful in exceptional cases with an imposed social distancing, but also in many other scenarios in which traditional education is limited, as it can provide worldwide access, flexible schedule, and personalized learning strategies. However, e-learning also presents some challenges compared to the traditional face-to-face education, e.g.: the absence of a direct contact between teachers and students and the difficulties to certify the authorship during online evaluations.

E-learning platforms allow to capture student information to create personalized environments whose contents and methodologies can be adapted dynamically to the different needs of each student. Information such as the performance on questions, the time necessary to perform the different tasks, the emotional state \cite{37}, or the heart rate \cite{16} can be used to understand the student behavior and conditions \cite{15}.

Undoubtedly, e-learning platforms will benefit significantly by exploiting the attention level of the student \cite{32}. This could be used to: i) adapt dynamically the environment and content \cite{11,29} based on the attention level, and ii) improve the educational materials and resources with a posterior analysis of the e-learning sessions
Table 1: Existing databases for Eye Blink detection.

<table>
<thead>
<tr>
<th>Ref.</th>
<th>Year</th>
<th>Blinks</th>
<th>Users</th>
<th>Resol.</th>
<th>Att.</th>
<th>Sensors</th>
</tr>
</thead>
<tbody>
<tr>
<td>[41]</td>
<td>NA</td>
<td>61</td>
<td>1</td>
<td>720 × 576</td>
<td>No</td>
<td>1</td>
</tr>
<tr>
<td>[31]</td>
<td>2007</td>
<td>255</td>
<td>20</td>
<td>320 × 240</td>
<td>No</td>
<td>1</td>
</tr>
<tr>
<td>[9]</td>
<td>2014</td>
<td>353</td>
<td>4</td>
<td>640 × 480</td>
<td>No</td>
<td>1</td>
</tr>
<tr>
<td>[34]</td>
<td>2015</td>
<td>300</td>
<td>5</td>
<td>640 × 480</td>
<td>No</td>
<td>1</td>
</tr>
<tr>
<td>[20]</td>
<td>2019</td>
<td>381</td>
<td>172</td>
<td>1280 × 720</td>
<td>No</td>
<td>1</td>
</tr>
<tr>
<td><strong>Ours</strong></td>
<td><strong>2020</strong></td>
<td><strong>3000</strong></td>
<td><strong>38</strong></td>
<td><strong>1280 × 720</strong></td>
<td><strong>Yes</strong></td>
<td><strong>3</strong></td>
</tr>
</tbody>
</table>

(e.g. detecting the type of contents more appropriate for a specific student).

Since the 70s there are studies relating the eye blink rate with cognitive activity like attention [2, 19]. The studies suggest that lower eye blink rates can be associated to high attention periods while higher eye blink rates are related to low attention levels. Therefore, in this context, automatic eye blink detection can be a tool for estimating the attention level of the students and improving e-learning platforms.

Similar rationale motivates also the usage of eye blink detection in many other problems where knowing the cognitive activity is useful, e.g.: driver fatigue detection [4, 18], lie detection [23, 26], detection of mild cognitive impairment [25], face anti-spoofing [12, 17, 31], dry eye syndrome recovery [35], human-computer interfaces [1] that ease communication for disabled people [30], fake news and DeepFakes detection [24, 27, 42], and activity recognition [21]. Even though our work has been developed with an e-learning application in mind, our research and resources (in particular mEBAL) can be very useful in all these related problems.

In the present work, two different approaches are used to detect eye blinks: i) EEG-based detection, using electroencephalography signals acquired with a head band. This detection is used to create a candidate list of eye blinks that will be then manually validated to create a groundtruth. And ii) Image-based detection, using the face images acquired using RGB and NIR cameras. This detection comprises face detection, landmark detection to locate the eye region, and eye blink detection. Fig. 1 shows the acquisition framework of mEBAL and the proposed eye blink detectors. The image-based detector could be used in applications where only a webcam is available. The main contributions of this study are:

- A new multimodal database for Eye Blink detection and Attention Level estimation: mEBAL². The database comprises data from 38 students and it includes: 3,000 blink samples (378,000 frames in total), 3,000 no-blink samples (378,000 frames in total), and the cognitive activity of the students. This database is eight times larger than existing eye blink databases and it is unique in its multimodal nature.
- An image-based eye blink detector trained over the proposed mEBAL and evaluated over another public eye blink database [20], considering in-the-wild scenarios. This detector consists of Convolutional Neural Networks (CNN).
- A preliminary experiment evaluating the eye blink detector as an estimator of the level of attention from RGB face images.

³https://github.com/BitDLab/mEBAL

The rest of the paper is organized as follows: Section 2 summarizes works related to eye blink detection. Section 3 presents our database. Section 4 describes the baseline eye blink detector developed for our experiments. Section 5 analyzes the results obtained on eye blink detection and attention level estimation. Finally, remarks and future work are drawn in Section 6.

2 RELATED WORKS

One of the challenges to train robust eye blink detectors is the absence of large-scale public databases. Table 1 summarizes the most popular eye blink detection databases in the literature. As can be seen, all available databases comprise only a few hundred samples. The main differences lie in the number of users and acquisition conditions. While databases such as [9, 31, 41] present controlled environments, recent works explore the detection of eye blinks in unconstrained scenarios [20].

Regarding the detection methods, they can be divided into: i) motion-based [9, 28], which exploit the dynamic information around eye blink events; ii) appearance-based [7, 39], which process the images and extract features related to the texture and shape of the eyes; and iii) mixed approaches based on the combination of image and temporal information [20].

To the best of our knowledge, our contributed mEBAL database is the first one that provides both eye blink images and cognitive activity information.

3 MEBAL: MULTIMODAL FRAMEWORK FOR EYE BLINK AND ATTENTION LEVEL

3.1 Acquisition Sensors

We designed a multimodal acquisition framework to monitor cognitive and eye blink activity during the execution of online tasks (see Fig. 1) based on the edBB platform for remote education assessment [14]. The following sensors are considered in the acquisition setup (see Table 2):

- An EEG headset by NeuroSky³ that captures 5 channels of electroencephalographic information (α, β, γ, δ, θ). These signals provide temporal information related to the cognitive activity of the student. The sensor also provides a temporal sequence with the eye blink strength. The sampling rate of the band is 1 Hz. The EEG band is used to capture the cognitive activity of the student and the eye blink candidates that will be used as groundtruth (as we will see later, some eye blinks must be discarded as false positives).
- An Intel RealSense (model D435i), which comprises 1 RGB and 2 NIR cameras. An average blink takes 100ms-400ms

Table 2: Sensors included in the mEBAL framework.

<table>
<thead>
<tr>
<th>Sensors</th>
<th>Sampling Rate</th>
<th>Features</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 RGB Camera</td>
<td>30 Hz</td>
<td>MP4 with codec H.264</td>
</tr>
<tr>
<td>2 Infrared Cameras</td>
<td>30 Hz</td>
<td>MP4 with codec H.264</td>
</tr>
<tr>
<td>EEG Band</td>
<td>1 Hz</td>
<td>α, β, γ, δ, θ</td>
</tr>
</tbody>
</table>

²https://github.com/BitDLab/mEBAL

³https://store.neurosky.com/pages/mindwave
were validated, we stored the 10 frames previous and posterior to we made a manual refinement of the eye blink samples detected by was labelled following a semi-supervised method. First, eye blink candidates were selected using the EEG band signals (eye blink detection are equal.

\begin{table}[h]
\centering
\caption{Eye Blink detection results on HUST-LEBW dataset.}
\begin{tabular}{|c|c|c|c|c|}
\hline
Method & Eye & Recall & Precision & F1  \\
\hline
[28] & Left & 0.0164 & 0.6667 & 0.0320  \\
& Right & 0.0159 & \textbf{1.0000} & 0.0313  \\
\hline
[5] & Left & 0.0164 & \textbf{1.0000} & 0.0323  \\
& Right & 0.0000 & 0.0000 & 0.0000  \\
\hline
[40] & 2 eyes & 0.0714 & 0.4500 & 0.1233  \\
\hline
[9] & Left & 0.0574 & 0.4118 & 0.1007  \\
& Right & 0.0317 & 0.3077 & 0.0576  \\
\hline
[39] & Left & 0.3607 & 0.6471 & 0.4632  \\
& Right & 0.3016 & 0.5758 & 0.3958  \\
\hline
[20] & Left & 0.5410 & 0.8919 & 0.6735  \\
& Right & 0.4444 & 0.7671 & 0.5628  \\
\hline
Ours & Left & \textbf{0.9603} & 0.6080 & \textbf{0.7446}  \\
& Right & \textbf{0.7950} & 0.7348 & \textbf{0.7637}  \\
\hline
\end{tabular}
\end{table}

3.2 Description of the mEBAL Database

Students performed 8 different tasks categorized in the following three groups: i) enrollment form: name and surname, ID number, nationality, e-mail address, etc. (low level of attention is expected); ii) writing questions: these questions are oriented to measure the students’ cognitive abilities under different situations such as solving logical problems, describing images, crosswords, finding differences, etc. (increasing level of attention is expected); and iii) multiple choice questions: aimed to detect the students’ attention and focus levels (high level of attention is expected).

mEBAL comprises a total of 6,000 samples divided in two halves (blink and no-blink) from both eyes acquired with 1 RGB and 2 NIR cameras. Each sample comprises 21 frames (around 600 ms.) for a total number of images of 756,000 (6,000 × 21 × 2 × 3). Aspects such as the user position and changes in the illumination were considered during the acquisition in order to simulate realistic e-learning scenarios. 11 out of the 38 students used glasses.

The mEBAL dataset was obtained from the raw data provided in the edBBbd [14]. The eye blink and attention level information was labelled following a semi-supervised method. First, eye blink candidates were selected using the EEG band signals (eye blink strength is an attribute provided by the EEG band SDK). Second, we made a manual refinement of the eye blink samples detected by the band to eliminate false positives. Once the eye blink samples were validated, we stored the 10 frames previous and posterior to the eye blink event (21 frames in total for each eye blink). These frames can be used to exploit the temporal information proposed in some approaches of the literature [39].

Figure 2: Landmark detection and ROI extraction. Example images included in mEBAL. Note that in some no-blink cases the eyes seem to be closed due to the gaze orientation.

landmark detection [8] to track the eye position (see Fig. 2). Entire face images, eye bounding boxes, and the cropped eyes are provided in our contributed mEBAL database. Additionally, we include the cognitive temporal signals \(a, \beta, \gamma, \delta, \theta\) provided by the EEG band. As the number of no-blink images is much larger than the blink images, we subsampled the no-blink images to obtain the same number of samples per class: Blink and no-Blink (see Fig. 1).

4 EYE BLINK DETECTION ALGORITHM

Inspired in the popular VGG16 architecture [38], we propose an eye blink detector based on a CNN trained from scratch. The proposed network comprises an input layer of 50 × 50 size, followed by 3 convolutional layers with ReLU activation (32/32/64 filters of size 3 × 3), with 3 max pooling layers between them, a dense layer of 64 units with ReLU activations, and a final output layer with one unit (sigmoid activation). Also, we use dropout (0.5) to reduce overfitting. The batch size is set up to 50. Adam optimizer is considered with default parameters (0.001 learning rate). The network is trained as a binary classifier (eyes open or closed), using the mEBAL subset of RGB cropped eyes (see Fig. 2).

5 EXPERIMENTS AND RESULTS

5.1 Experimental Protocol

The evaluation is performed over the public HUST-LEBW benchmark for eye blink detection presented in [20]. Note that the mEBAL database (used for training our blink detector) was obtained in a controlled environment, while the HUST-LEBW dataset (used for testing our blink detector) was obtained in the wild. This will allow to measure the generalization ability of the proposed eye blink detector to unseen scenarios [13, 33].

The HUST-LEBW dataset includes 381 eye blink and 292 no-blink samples. Each sample comprises 13 frames. All 13 frames are processed with the CNN proposed in Section 4, which generates for each input image an eye blink strength score. Among the 13 scores obtained for a sample (one per frame), the maximum is selected to represent the sample score. The decision threshold is fixed to the point in which the False Positive and False Negative rates in blink detection are equal.
5.2 Eye Blink Detection Results

Table 3 presents the results and comparison with previous approaches evaluated over the same HUST-LEBW Dataset [5, 9, 20, 28, 39, 40]. The results show how our method outperforms state-of-the-art eye blink detection algorithms for Recall and F1 metrics. There is an important difference between left and right eyes in terms of performance. This difference is caused by the characteristics of the HUST-LEBW dataset (e.g., head orientation in the samples).

Note that the approach proposed in [20] was developed using the training set provided with HUST-LEBW. This training set comprises images with similar characteristics as those used in the evaluation set. Our results were obtained training with mEBAL, acquired under controlled conditions. Thus, the high performance obtained with our method demonstrates the good generalization capacity of our approach to unseen scenarios. These results demonstrate the potential of mEBAL to train a new generation of eye blink detectors.

5.3 Attention Level Estimation Results

In this section, we analyze the relationship between the eye blink rate and the attention level estimated by the algorithm provided with the EEG band. We present a preliminary experiment focused on sudden changes in the level of attention of students. According to the literature [2, 19], we should observe an inverse relationship between eye blink rate and level of attention. Fig. 3 shows the attention level and eye blink estimation of 4 different students during a period of 4 minutes. We have averaged the level of attention for each 20 seconds using a sliding window of 5 seconds. The estimated and groundtruth eye blinks are calculated as bpm (blinks per minutes) using a sliding window of 5 seconds again. All three signals were normalized using the min-max technique [10, 22].

The results show a small difference between the eye blink detector and the groundtruth provided by the EEG band. These results demonstrate again the high accuracy of the trained image-based blink detector. Regarding the level of attention and the eye blink rate, in most instances, the level of attention and eye blink signals have a negative correlation, which is coherent with the literature. High peaks in the attention level are usually correlated with low eye blink rates and vice versa. However, other factors can affect the eye blink rates and more context is necessary to improve the estimation of the level of attention based on the eye blink rate.

6 CONCLUSION

This work has presented mEBAL, a new multimodal database for eye blink detection and attention level estimation, the largest one in the literature for research in these problems. This database improves previous databases in sensors (EEG band, NIR and RGB cameras) and samples: 6,000 samples in two halves (blink and no-blink) of both eyes for a total of 756,000 images.

Also, we have performed experiments to: i) detect the eye blinks with Convolutional Neural Networks trained from scratch using RGB images, and ii) predict the level of attention of students conducting various e-learning tasks based on their eye blink frequency.

The results achieved have proved that mEBAL can be used to train accurate eye blink detectors under realistic acquisition conditions. In fact, our results have outperformed the state of the art with a simple yet powerful CNN learning architecture, thanks mainly to the utility of our contributed database, which is well suited for data-driven eye blink detection approaches.

Future work should consider other recognition architectures better adapted to the eye blink detection problem (e.g., combining CNNs and Recurrent Neural Networks [3, 10] to incorporate time information [43]). The preliminary experiment carried out to measure the correlation between the attention level and the eye blink frequency has shown encouraging results and should be also analyzed in more depth.

Even though our work has been developed with e-learning in mind [14, 15], the contributed resources and methods for eye blink detection can be very useful for other problems as well, e.g.: driver fatigue detection [18], lie detection, DeepFakes detection [24], face anti-spoofing [17], human-computer interfaces [1], and others.

ACKNOWLEDGMENTS

This work has been supported by projects: PRIMA (ITN-2019-860315), TRESPASS-ETN (ITN-2019-860813), IDEA-FAST (IMI2-2018-15-two-stage-853981), BBLEC (RTI2018-101248-B-I00 MINECO-FEDER), and edBB (Universidad Autonoma de Madrid). Ruben Tolosana and postdoc support from CAM/FEDER. Roberto Daza is supported by a PhD FPI fellowship from MINECO-FEDER.
REFERENCES


