J. Mantas et al. (Eds.)

doi:10.3233/SHTI240731

Unsupervised Learning of Eye State Prototypes for Semantically Rich Blinking Detection

Yuxuan XIE[a,1], Tim BÜCHNER[a], Lukas SCHUHMANN[b], Orlando GUNTINAS-LICHIUS[b] and Joachim DENZLER[a]
[a] *Computer Vision Group, Friedrich Schiller University Jena, 07743 Jena, Germany*
[b] *Department of Otorhinolaryngology, Jena University Hospital, 07747 Jena, Germany*
ORCiD ID: Yuxuan Xie https://orcid.org/0009-0000-4117-8292, Tim Büchner https://orcid.org/0000-0002-6879-552X, Orlando Guntinas-Lichius https://orcid.org/0000-0001-9671-0784, Joachim Denzler https://orcid.org/0000-0002-3193-3300

Abstract. Blinking contributes to the health and protection of the eye and also holds potential in the context of muscle or nerve disorder diagnosis. Traditional methods of classifying eye blinking as *open* or *closed* are insufficient, as they do not capture medical-relevant aspects like closure speed, duration, or percentage. The issue could be solved by reliably detecting blinking intervals in high-temporal recordings. Our research demonstrates the reliable detection of blinking events through data-driven analysis of the eye aspect ratio. In an unsupervised manner, we establish an *eye state prototype* to identify blink intervals and measure inter-eye synchronicity between moments of peak closure. Additionally, our research shows that *manually defined prototypes* yield comparable results. Our results demonstrate inter-eye synchronicity up to 4.16 *ms.* We anticipate that medical professionals could utilize our methods to identify or define disease-specific prototypes as potential diagnostic tools.

Keywords. Blinking Detection, Eye-Aspect-Ratio, High-temporal Videos, Pattern Matching

1. Introduction

Blink analysis could be a novel diagnostic tool with the potential to detect muscular or neurological disorders such as facial palsy or Parkinson's disease by focusing on quick eyelid closure. However, current eye openness classification as either *opened* or *closed* disregards important diagnostic details such as blink speed, duration, and inter-eye synchronization. Critically, these require the whole blinking interval to be known. Further, the intricate nature of blinking is lost in low-temporal resolution (30 FPS). Annotations would demand considerable human labor, making supervised methods unfeasible. We propose an *eye state prototype blink matching* (ESPBM)[2] approach to overcome these limitations. Our experiments compare inter-eye synchronicity for

[1] Corresponding Author: Yuxuan Xie; E-mail: yuxuan.z.xie@gmail.com.
[2] The methods source code is available at https://github.com/cvjena/ESPBM

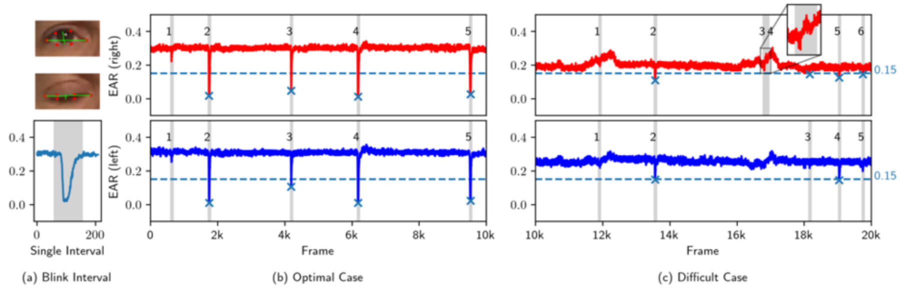


Figure 1. The subplots show a single blink interval based on the landmarks (**a**) and two eye aspect ratio (EAR) time series from different persons (**b**, **c**). The blinks are noted by numbers. We visualize the results of simple thresholding (dashed line), peak finding (crosses), and our prototype approach (gray areas). In the optimal case (**b**), only our method is amplitude invariant, including blink number one. The difficult case (**c**) contains more noise, likely due to head rotation, but our method can still detect the blink intervals.

unsupervisedly learned and manually defined prototypes. Both approaches can reliably extract the blink intervals for healthy probands, opening the possible research direction to identify or define disease-specific eye state prototypes as new diagnostic tools.

2. Materials and Methods

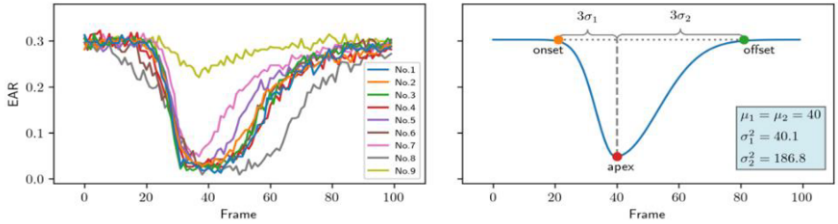
To better understand spontaneous blinking beyond the *open* or *closed* state, we model a blink pattern as an *eye state prototype* and use it to match similar subsequences. Additionally, we focus on inter-eye synchronicity to validate our approach. We recorded 30 healthy participants (age range: 22-82 years, 10 men, 20 women) using a frontal-facing iPhone 8 camera (Apple Inc, Cupertino, USA) at 240 FPS. The camera was angled slightly from below to ensure invariance regarding head tilting. Each participant watched a 20-minute animal documentary without further instructions. A single recording contains roughly 300,000 frames without any eye-closing state annotations.

2.1. Eye-Opening State Description via the Eye Aspect Ratio

The eye-opening state was first modeled by Ekman and Friesen's standardization through the Facial Action Coding System [1]. Initially, this state was binary: *open* or *closed*. However, this simple attribution is insufficient to measure medically relevant characteristics like eyelid closing speed, blink duration, or inter-eye synchronicity. A notion based on the actual physiological state of the eye is needed.

Utilizing facial landmarks for eyes enables an automatic, objective description of openness. People may not maintain the same position during the recording, so a descriptor unaffected by camera distance is necessary. Options include MRD1, MRD2, ocular surface area exposure, and eye aspect ratio (EAR)[2–5]. Notably, only EAR, by definition, is scale-invariant and does not require normalization relative to the pupil diameter. The EAR value reflects the ratio of eye height to width, remaining constant when the eye is open and approaching zero when closed, as depicted in **Figure 1a**. We employ JeFaPaTo[2] to extract the EAR time series from the video recordings. This software features automatic face detection, landmark extraction via MediaPipe [6], and computation of EAR values per eye using six eye landmarks. Frames are processed independently without any temporal smoothing, which is visible in **Figure 1b** and

Figure 1c. We propose a fully automatic data-driven method to extract blink intervals in EAR time series.



(a) Top 9 most similar subsequences from an EAR time series (b) Manually defined prototype via normal distributions

Figure 2. We present *unsupervised extracted* (**a**) prototypes and a *manually defined* (**b**) prototype used for blink detection using the eye aspect ratio (EAR). The chosen *learned prototype* (No. 1) most closely resembles all subsequences. Furthermore, we design the *manual prototype* by linking two univariate normal distributions to model a blink interval's shape. These distribution parameters could be adapted for disease-

2.2. Unsupervised Learning and Manual Design of Eye State Prototypes

There are multiple ways to describe the eye openness. The binary classification approach by [3] necessitates annotations, requiring substantial human labor, and could introduce unwanted biases. Simple rule-based methods fail to identify eyelid closure reliably. **Figure 1b** and **Figure 1c** show the results of the peak finding [2,7] (marked with crosses and selection threshold as dashed lines). Even in the optimal case, blinks with low amplitude (number one) are undetected. In the complex case, where the time series is noisy, delicate parameter tuning would be needed to detect all blinks. Critically, such extensive work is unfeasible for long videos and large amounts of participants.

We propose an *eye state prototype blink matching* (ESPBM) approach for eyelid closure detection independent of peak amplitude. Our method includes *unsupervisedly learned* and *manually defined prototypes* to model various eye openness states. We utilize Stumpy for the prototype extraction by finding repeated subsequences along the matrix profile [8,9]. We enforce z-normalization on the time series to ensure detection regardless of the blink amplitude. The extraction window size can be chosen regarding the temporal resolution. Please note that large windows could hinder the detection of subsequent blinks. Hence, we select a window of *m=100* frames for our data (equal to 400 *ms* at 240 FPS), enabling complete inclusion of blink *onset, apex,* and *offset*, as shown in **Figure 2b**. For our experiments, we chose the most similar subsequence among all extracted candidates as the learned prototype, with Figure 2a displaying the top nine.

Additionally, we observe that most similar subsequences resemble a skew-normal distribution. Therefore, we design the *manual prototype* by combining two normal distributions. We learn these parameters for our experiments based on the observed patterns, yielding $\mu_1 = \mu_2 = 40.1$, $\sigma_1^2 = 41.1$, and $\sigma_2^2 = 186.8$ within a window size *m=100*. Onset and offset locations are $3 \cdot \sigma_{1,2}$ according to direction. This modeling could be used to detect specific eyelid speed patterns to create disease-specific prototypes.

We leverage the fast pattern-matching algorithm of Stumpy [8] to extract the intervals resembling the prototypes. No further preprocessing of the time series is required, further simplifying the applicability of our approach. Our current experiments focus on inter-eye synchronicity using the apex location, see **Figure 2b**. Future research could aim to estimate the onset and offset points of the extracted intervals to allow a more detailed analysis of the blinks.

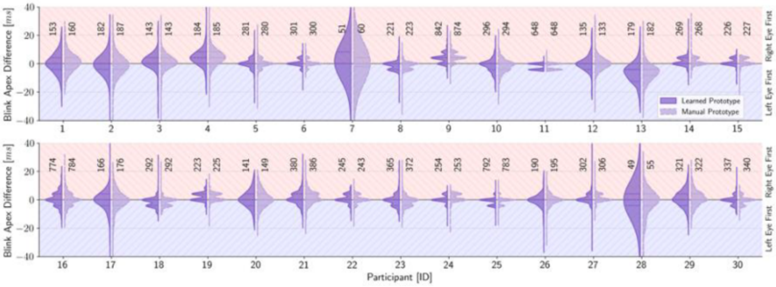


Figure 3. The plot illustrates the inter-eye synchronicity (apex) distributions, in milliseconds, among 30 healthy individuals. The blink count per person is annotated above each violin. Observable patterns in the data align between the *unsupervised learned* and *manually defined eye state prototypes*. These patterns, if further investigated, may provide person-specific insights. Nevertheless, the results show 75% of inter-eye synchronicity is within ±4.16 *ms*.

3. Results and Discussion

Eye state prototypes allow for the extraction of blinking intervals, offering more details than binary classification such as *open* or *closed*. We compare both methods based on inter-eye synchronicity. **Figure 3** displays our measured inter-eye synchronicity distributions of blink apex differences across all participants. We computed the apex location of a blink interval as the point of lowest EAR value. These locations were then used to match left and right eye blinks within 200 *ms*. **Table 1** suggests that both methods perform comparably with minor differences. Most blinks fall within a ±4.16 *ms* window (equating to one frame at 240 FPS). From the data, most blinks are synchronous among our healthy participants. We observe slight asynchronicity for participants no. 4 and no. 13. We hypothesize that a lower temporal resolution may not reveal this asynchronicity. Yet, an even higher temporal resolution could disclose more detailed results, even in individuals without health issues. Finally, the individual blink count significantly fluctuates throughout the 20-minute video, ranging from a minimum of 51 to a maximum of 842 blinks. Considering these figures, investigating potential unknown correlations with factors such as eye dryness or fatigue may be interesting.

Additionally, specific disparity distributions are present. Most individuals exhibit an unimodal distribution, with most blinks at a 0.00 *ms* difference. However, some distributions have several peaks at differences of -4.16 *ms*, 0.00 *ms*, and 4.16 *ms*, which might indicate either constraints due to temporal resolution or minimal asynchronicity. A more significant number of participants could reveal the actual cause.

This study does not focus on the potential relations between blink patterns and medical implications. Instead, we demonstrate that blinking can be identified in a data-driven manner without requiring human annotations or intervention.

Table 1. The inter-eye synchronicity statistics for the *unsupervised learned* and *manually defined* prototypes among all 30 participants. Both prototypes perform similarly, and synchronicity is within a ±4.16 *ms window* (one frame at 240 FPS).

Prototype	Count	Mean	Std.	25%	50%	75%
Learned	9045	0.319 ms	5.972 ms	0.000 ms	0.000 ms	±4.166 ms
Manual	8942	0.332 ms	5.633 ms	0.000 ms	0.000 ms	±4.166 ms

4. Conclusions and Outlook

We propose a non-human-involved eye-blink detection method. Both *unsupervisedly learned* and *manually defined prototypes* yield similar detection results in healthy individuals. With minimal hyperparameter selection, our approach simplifies integration into clinical procedures or medical studies. Additionally, we establish a healthy individual baseline for future disease-specific pattern identification regarding inter-eye synchronicity. This method could pave the way for new research on the correlation between blinking behavior and neurological disorders. Substantial head rotation impacts landmark placement, affecting the eye-width calculation and causing non-eyelid-based changes in the EAR score. Even under these instances, we can still detect blink intervals using the general pattern, as seen in **Figure 1c**. Yet, making our method invariant to head rotations would enhance its applicability in routine clinical use for healthcare professionals and patients. As we provide the apex location with our method and the whole blink interval, further studies could investigate blink duration, opening and closing speed, acceleration, or classification into partial or complete eyelid closure [2]. This could be accomplished by estimating the normal distribution parameters for each extracted interval. We hope our method enables medical experts to study the behavior of spontaneous eyelid closure.

Ethics Approval: Written informed consent was obtained from all participants. The ethics committee of the Jena University Hospital approved the study (No. 2019-1539).

References

[1] Ekman P, Friesen WV. Facial Action Coding System. Consulting Psychologists Press. 1978, doi: 10.1037/t27734-000.

[2] Büchner T, Mothes O, Guntinas-Lichius O, Denzler J. JeFaPaTo - A joint toolbox for blinking analysis and facial features extraction. J Open Source Softw. 2024;9(97):6425, doi: 10.21105/joss.06425

[3] Soukupova T, Cech J. Real-Time Eye Blink Detection using Facial Landmarks. 21st computer vision winter workshop; 2016 Feb;2.

[4] Schulz CB, Clarke H, Makuloluwe S, Thomas PB, Kang S. Automated extraction of clinical measures from videos of oculofacial disorders using machine learning: feasibility, validity and reliability. Eye (Lond). 2023 Sep;37(13):2810-16, doi: 10.1038/s41433-023-02424-z.

[5] Chen H-C, Tzeng S-S, Hsiao Y-C, et al. Smartphone-Based Artificial Intelligence-Assisted Prediction for Eyelid Measurements: Algorithm Development and Observational Validation Study. JMIR MHealth UHealth. 2021;9:e32444, doi: 10.2196/32444.

[6] Lugaresi C, Tang J, Nash H, et al. MediaPipe: A Framework for Building Perception Pipelines [Internet]. arXiv; 2019, doi: 10.48550/arXiv.1906.08172.

[7] Virtanen P, Gommers R, Oliphant TE, et al. SciPy 1.0: fundamental algorithms for scientific computing in Python. Nat Methods. 2020;17:261–72, doi: 10.1038/s41592-019-0686-2.

[8] Yeh C-CM, Zhu Y, Ulanova L, et al. Matrix Profile I: All Pairs Similarity Joins for Time Series: A Unifying View That Includes Motifs, Discords and Shapelets. 2016 IEEE 16th Int Conf Data Min ICDM [Internet]. 2016. p. 1317–22, doi: 10.1109/ICDM.2016.0179.

[9] Law SM. STUMPY: A Powerful and Scalable Python Library for Time Series Data Mining. J Open Source Softw. 2019;4:1504, doi: 10.21105/joss.01504.